Proceedings of the 20th Amsterdam Colloquium

Brochhagen, T.; Roelofsen, F.; Theiler, N.

Proceedings of the
20th Amsterdam Colloquium

20th Amsterdam Colloquium 2015

Edited by Thomas Brochhagen, Floris Roelofsen & Nadine Theiler

Foreword

This is a collection of papers presented at the 20th Amsterdam Colloquium, organized by the Institute for Logic, Language, and Computation (ILLC) at the University of Amsterdam, December 16–18, 2015. The bi-annual Amsterdam Colloquia aim at bringing together linguists, philosophers, logicians, cognitive scientists and computer scientists who share an interest in the formal study of the semantics and pragmatics of natural and formal languages.

Besides the regular programme, the 2015 edition featured two workshops on Negation and Reasoning in natural language, respectively, and one evening lecture, jointly organized with the E.W. Beth Foundation. The programme included eight invited talks and 46 contributed talks.

We would like to thank the members of the programme committee and all the reviewers, listed below, for their efforts in selecting the contributed talks. We would also like to thank Maria Aloni, Tamara Dobler, Luca Incurvati, Fenneke Kortenbach and Peter van Ormondt for their help in organizing the colloquium.

Lastly, we would like to thank the ILLC, the E.W. Beth Foundation, the Netherlands Organization for Scientific Research (NWO), and the European Research Council (ERC) for financial support.

<div align="right">

Thomas Brochhagen
Floris Roelofsen
Nadine Theiler

</div>

Programme Committee

General programme

Robert van Rooij (chair)	ILLC, University of Amsterdam
Franz Berto	ILLC, University of Amsterdam
Paul Dekker	ILLC, University of Amsterdam
Henriette de Swart	Institute for Linguistics, Utrecht University

Workshop "Negation: logical, linguistic and philosophical perspectives"

Franz Berto	ILLC, University of Amsterdam
Luca Incurvati	ILLC, University of Amsterdam
Julian Schlöder	ILLC, University of Amsterdam

Workshop "Reasoning in natural language: symbolic and sub-symbolic approaches"

Jakub Szymanik	ILLC, University of Amsterdam
Willem Zuidema	ILLC, University of Amsterdam

Reviewers

Dora Achourioti	University of Amsterdam
Luis Alonso-Ovalle	McGill University
Daniel Altshuler	Hampshire College
Scott Anderbois	Brown University
Chris Barker	New York University
David Beaver	University of Texas at Austin
Oliver Bott	University of Tübingen
Adrian Brasoveanu	UC Santa Cruz
Lisa Bylinina	Meertens Institute, Amsterdam
Lucas Champollion	New York University
Emmanuel Chemla	Ecole Normale Supérieure, Paris
Roberto Ciuni	University of Amsterdam
Ariel Cohen	Ben-Gurion University of the Negev
Cleo Condoravdi	Stanford University
Elizabeth Coppock	University of Gothenburg
Chris Cummins	University of Edinburgh
Michael De	University of Utrecht
Judith Degen	Stanford University
Jakub Dotlacil	University of Groningen
Catarina Dutilh Novaes	University of Groningen
Regine Eckardt	University of Konstanz
Paul Egre	Institut Jean-Nicod, Paris
Donka Farkas	University of California, Santa Cruz
Tim Fernando	Trinity College Dublin
Raquel Fernández	University of Amsterdam

Michael Franke University of Tübingen
Jon Gajewski University of Connecticut
Valentine Hacquard University of Maryland
Andreas Haida Humboldt University, Berlin
Wesley Holliday University of California, Berkeley
Laurence Horn Yale University
Thomas Icard Stanford University
Gerhard Jäger University of Tübingen
Andreas Kapsner Ludwig-Maximilians-Universität München
Magdalena Kaufmann University of Connecticut
Stefan Kaufmann University of Connecticut
Chris Kennedy University of Chicago
Nathan Klinedinst University College London
Angelika Kratzer University of Massachusetts at Amherst
Fred Landman Tel Aviv University
Staffan Larsson University of Gothenburg
Emar Maier University of Groningen
Edwin Mares Victoria University of Wellington
Yaron McNabb University of Groningen
Louise McNally Universitat Pompeu Fabra
Paula Menendez-Benito Universitat Pompeu Fabra
Sarah Murray Cornell University
Reinhard Muskens University of Tilburg
Rick Nouwen University of Groningen
Hitoshi Omori Kyoto University
Edgar Onea University of Göttingen
Francesco Paoli University of Cagliari
Doris Penka University of Konstanz
Christoper Potts Stanford University
Nausicaa Pouscoulous University College London
Jessica Rett University of California, Los Angeles
Craige Roberts Ohio State University
Maribel Romero University of Konstanz
Susan Rothstein Bar-Ilan University
Uli Sauerland ZAS, Berlin
Philippe Schlenker Ecole Normale Supérieure, Paris / New York University
Bernhard Schwarz McGill University
Roger Schwarzschild Rutgers University
Yael Sharvit University of California, Los Angeles
Martin Stokhof University of Amsterdam
Matthew Stone Rutgers University
Yasutada Sudo University College London
Eric Swanson University of Michigan
Kristen Syrett Rutgers University
Anna Szabolcsi New York University
Kjell Johan Sæbø University of Oslo
Judith Tonhauser The Ohio State University
Wataru Uegaki Keio University/Institut Jean-Nicod, Paris

Bob van Tiel	Université Libre de Bruxelles
Achille Varzi	Columbia University
Kai von Fintel	Massachusetts Institute of Technology
Klaus von Heusinger	University of Köln
Galit W. Sassoon	Bar Ilan University
Heinrich Wansing	Ruhr-Universität Bochum
Yoad Winter	University of Utrecht
Henk Zeevat	University of Düsseldorf
Hedde Zeijlstra	University of Göttingen
Ede Zimmermann	University of Potsdam
Malte Zimmermann	Goethe Universität Frankfurt
Sarah Zobel	University of Tübingen

Contents

Invited

Contributed

Learning in the Rational Speech Acts Model

Will Monroe and Christopher Potts

Stanford University, California, U.S.A.
`wmonroe4@cs.stanford.edu`, `cgpotts@stanford.edu`

Abstract

The Rational Speech Acts (RSA) model treats language use as a recursive process in which probabilistic speaker and listener agents reason about each other's intentions to enrich the literal semantics of their language along broadly Gricean lines. RSA has been shown to capture many kinds of conversational implicature, but it has been criticized as an unrealistic model of speakers, and it has so far required the manual specification of a semantic lexicon, preventing its use in natural language processing applications that learn lexical knowledge from data. We address these concerns by showing how to define and optimize a trained statistical classifier that uses the intermediate agents of RSA as hidden layers of representation forming a non-linear activation function. This treatment opens up new application domains and new possibilities for learning effectively from data. We validate the model on a referential expression generation task, showing that the best performance is achieved by incorporating features approximating well-established insights about natural language generation into RSA.

1 Pragmatic language use

In the Gricean view of language use [18], people are rational agents who are able to communicate efficiently and effectively by reasoning in terms of shared communicative goals, the costs of production, prior expectations, and others' belief states. The Rational Speech Acts (RSA) model [11] is a recent Bayesian reconstruction of these core Gricean ideas. RSA and its extensions have been shown to capture many kinds of conversational implicature and to closely model psycholinguistic data from children and adults [7, 2, 23, 30, 33].

Both Grice's theories and RSA have been criticized for predicting that people are more rational than they actually are. These criticisms have been especially forceful in the context of language production. It seems that speakers often fall short: their utterances are longer than they need to be, underinformative, unintentionally ambiguous, obscure, and so forth [1, 10, 16, 24, 28, 29]. RSA can incorporate notions of bounded rationality [4, 13, 20], but it still sharply contrasts with views in the tradition of [6], in which speaker agents rely on heuristics and shortcuts to try to accurately describe the world while managing the cognitive demands of language production.

In this paper, we offer a substantially different perspective on RSA by showing how to define it as a trained statistical classifier, which we call *learned RSA*. At the heart of learned RSA is the back-and-forth reasoning between speakers and listeners that characterizes RSA. However, whereas standard RSA requires a hand-built lexicon, learned RSA infers a lexicon from data. And whereas standard RSA makes predictions according to a fixed calculation, learned RSA seeks to optimize the likelihood of whatever examples it is trained on. Agents trained in this way exhibit the pragmatic behavior characteristic of RSA, but their behavior is governed by their training data and hence is only as rational as that experience supports. To the extent that the speakers who produced the data are pragmatic, learned RSA discovers that; to the extent that their behavior is governed by other factors, learned RSA picks up on that too. We validate the model on the task of *attribute selection for referring expression generation* with a widely-used corpus of referential descriptions (the TUNA corpus; [34, 15]), showing that it improves on heuristic-driven models and pure RSA by synthesizing the best aspects of both.

(a) Simple reference game.

	beard	*glasses*	*tie*
r_1	.5	.5	0
r_2	0	.5	.5
r_3	0	0	1

(b) s_0

	r_1	r_2	r_3
beard	1	0	0
glasses	.5	.5	0
tie	0	.33	.67

(c) l_1

	beard	*glasses*	*tie*
r_1	.67	.33	0
r_2	0	.6	.4
r_3	0	0	1

(d) s_1
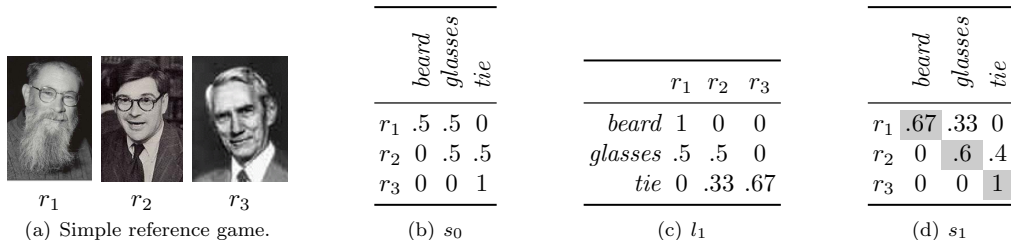
Figure 1: Ambiguity avoidance in RSA.

2 RSA as a speaker model

RSA is a descendent of the signaling systems of [25] and draws on ideas from iterated best response (IBR) models [13, 20], iterated cautious response (ICR) models [21], and cognitive hierarchies [4] (see also [17, 31]). RSA models language use as a recursive process in which speakers and listeners reason about each other to enrich the literal semantics of their language. This increases the efficiency and reliability of their communication compared to what more purely literal agents can achieve.

For instance, suppose a speaker and listener are playing a reference game in the context of the images in Figure 1(a). The speaker S has been privately assigned referent r_1 and must send a message that conveys this to the listener. A literal speaker would make a random choice between *beard* and *glasses*. However, if S places itself in the role of a listener L receiving these messages, then S will see that *glasses* creates uncertainty about the referent whereas *beard* does not, and so S will favor *beard*. In short, the pragmatic speaker chooses *beard* because it's unambiguous for the listener.

RSA formalizes this reasoning in probabilistic Bayesian terms. It assumes a set of messages M, a set of states T, a prior probability distribution P over states T, and a cost function C mapping messages to real numbers. The semantics of messages is defined by a lexicon $\mathcal{L}$, where $\mathcal{L}(m, t) = 1$ if m is true of t and 0 otherwise. The agents are then defined as follows:

$$s_0(m \mid t, \mathcal{L}) \propto \exp\left(\lambda\left(\log \mathcal{L}(m, t) - C(m)\right)\right) \qquad (1)$$

$$l_1(t \mid m, \mathcal{L}) \propto s_0(m \mid t, \mathcal{L})P(t) \qquad (2)$$

$$s_1(m \mid t, \mathcal{L}) \propto \exp\left(\lambda\left(\log l_1(t \mid m, \mathcal{L}) - C(m)\right)\right) \qquad (3)$$

The model that is the starting point for our contribution in this paper is the pragmatic speaker s_1. It reasons not about the semantics directly but rather about a pragmatic listener l_1 reasoning about a literal speaker s_0. The strength of this pragmatic reasoning is partly governed by the temperature parameter λ, with higher values leading to more aggressive pragmatic reasoning.

Figure 1 tracks the RSA computations for the reference game in Figure 1(a). Here, the message costs C are all 0, the prior over referents is flat, and $\lambda = 1$. The chances of success for the literal speaker s_0 are low, since it chooses true messages at random. In contrast, the chances of success for s_1 are high, since it derives the unambiguous system highlighted in gray.

The task we seek to model is a language generation task, so we present RSA from a speaker-centric perspective. It has been explored more fully from a listener perspective. In that formulation, the model begins with a literal listener reasoning only in terms of the lexicon $\mathcal{L}$ and state priors. Models of this general form have been shown to capture a wide range of pragmatic behaviors [2, 12, 22, 23, 30] and to increase success in task-oriented dialogues [35, 36].

| COLOUR:GREEN ORIENTATION:LEFT SIZE:SMALL TYPE:FAN X-DIMENSION:1 Y-DIMENSION:1 | COLOUR:GREEN ORIENTATION:LEFT SIZE:SMALL TYPE:SOFA X-DIMENSION:1 Y-DIMENSION:2 | COLOUR:RED ORIENTATION:BACK SIZE:LARGE TYPE:FAN X-DIMENSION:1 Y-DIMENSION:3 |
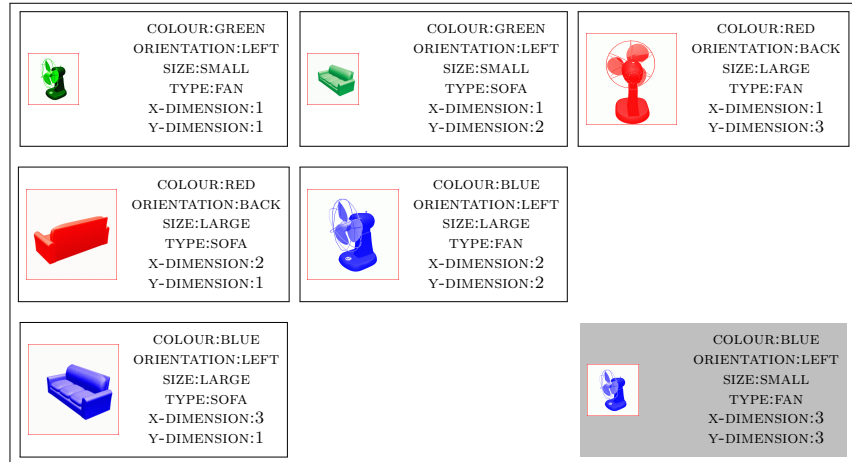
Utterance: "blue fan small"
Utterance attributes: [*colour:blue*]; [*size:small*]; [*type:fan*]

Figure 2: Example item from the TUNA corpus. Target is in gray.

RSA has been criticized on the grounds that it predicts unrealistic speaker behavior [16]. For instance, in Figure 1, we confined our agents to a simple message space. If permitted to use natural language, they will often produce utterances expressing predicates that are redundant from an RSA perspective—for example, by describing r_1 as *the man with the long beard and sweater*, even though *man* has no power to discriminate, and *beard* and *sweater* each uniquely identify the intended referent. This tendency has several explanations, including a preference for including certain kinds of descriptors, a desire to hedge against the possibility that the listener is not pragmatic, and cognitive pressures that make optimal descriptions impossible. One of our central objectives is to allow these factors to guide the core RSA calculation.

3   The TUNA corpus

In Section 6, we evaluate RSA and learned RSA in the TUNA corpus [34, 15], a widely used resource for developing and testing models of natural language generation. We introduce the corpus now because doing so helps clarify the learning task faced by our model, which we define in the next section.

In the TUNA corpus, participants were assigned a target referent or referents in the context of seven other distractors and asked to describe the target(s). Trials were performed in two domains, *furniture* and *people*, each with a *singular* condition (describe a single entity) and a *plural* condition (describe two). Figure 2 provides a (slightly simplified) example from the singular furniture section, with the target item identified by shading. In this case, the participant wrote the message "blue fan small". All entities and messages are annotated with their semantic attributes, as given in simplified form here. (Participants saw just the images; we include the attributes in Figure 2 for reference.)

The task we address is *attribute selection*: reproducing the multiset of attributes in the message produced in each context. Thus, for Figure 2, we would aim to produce {[*size:small*],

[*colour:blue*], [*type:fan*]}. This is less demanding than full natural language generation, since it factors out all morphosyntactic phenomena. Section 6 provides additional details on the nature of this evaluation.

4 Learned RSA

We now formulate RSA as a machine learning model that can incorporate the quirks and limitations that characterize natural descriptions while still presenting a unified model of pragmatic reasoning. This approach builds on the two-layer speaker-centric classifier of [17], but differs from theirs in that we directly optimize the performance of the pragmatic speaker in training, whereas [17] apply a recursive reasoning model on top of a pre-trained classifier. Like RSA, the model can be generalized to allow for additional intermediate agents, and it can easily be reformulated to begin with a literal listener.

Feature representations.   To build an agent that learns effectively from data, we must represent the items in our dataset in a way that accurately captures their important distinguishing properties and permits robust generalization to new items [8, 26]. We define our feature representation function ϕ very generally as a map from state–utterance–context triples $\langle t, m, c \rangle$ to vectors of real numbers. This gives us the freedom to design the feature function to encode as much relevant information as necessary.

As noted above, in learned RSA, we do not presuppose a semantic lexicon, but rather induce one from the data as part of learning. The feature representation function determines a large, messy hypothesis space of potential lexica that is refined during optimization. For instance, as a starting point, we might define the feature space in terms of the cross-product of all possible entity attributes and all possible utterance meaning attributes. For m entity attributes and n utterance attributes, this defines each $\phi(t, m, c)$ as an mn-dimensional vector. Each dimension of this vector records the number of times that its corresponding pair of attributes co-occurs in t and m. Thus, the representation of the target entity in Figure 2 would include a 1 in the dimension for clearly good pairs like COLOUR:BLUE $\wedge$ [*colour:blue*] as well as for intuitively incorrect pairs like SIZE:SMALL $\wedge$ [*colour:blue*].

Because ϕ is defined very generally, we can also include information that is not clearly lexical. For instance, in our experiments, we add dimensions that count the color attributes in the utterance in various ways, ignoring the specific color values. We can also define features that intuitively involve negation, for instance, those that capture entity attributes that go unmentioned. This freedom is crucial to bringing generation-specific insights into the RSA reasoning.

Literal speaker.   Learned RSA is built on top of a *log-linear model*, standard in the machine learning literature and widely applied to classification tasks [19, 27].

$$S_0(m \mid t, c; \theta) \propto \exp(\theta^T \phi(t, m, c)) \tag{4}$$

This model serves as our literal speaker, analogous to s_0 in (1). The lexicon of this model is embedded in the *parameters* (or *weights*) θ. Intuitively, θ is the direction in feature representation space that the literal speaker believes is most positively correlated with the probability that the message will be correct. We train the model by searching for a θ to maximize the conditional likelihood the model assigns to the messages in the training examples. Assuming the training is

effective, this increases the weight for correct pairings between utterance attributes and entity attributes and decreases the weight for incorrect pairings.

To find the optimal θ, we seek to maximize the conditional likelihood of the training examples using first-order optimization methods (described in more detail in **Training**, below). This requires the gradient of the likelihood with respect to θ. To simplify the gradient derivation and improve numerical stability, we maximize the log of the conditional likelihood:

$$J_{S_0}(t, m, c, \theta) = \log S_0(m \mid t, c; \theta) \tag{5}$$

The gradient of this log-likelihood is

$$
\begin{aligned}
\frac{\partial J_{S_0}}{\partial \theta} &= \phi(t, m, c) - \frac{1}{\sum_{m'} \exp(\theta^T \phi(t, m', c))} \sum_{m'} \exp(\theta^T \phi(t, m', c)) \phi(t, m', c) \\
&= \phi(t, m, c) - \sum_{m'} S_0(m' \mid t, c; \theta) \phi(t, m', c) \\
&= \phi(t, m, c) - \mathbb{E}_{m' \sim S_0(\cdot \mid t, c; \theta)} \left[\phi(t, m', c) \right]
\end{aligned}
\tag{6}
$$

where the first two equations can be derived by expanding the proportionality constant in the definition of S_0.

Pragmatic speaker. We now define a pragmatic listener L_1 and a pragmatic speaker S_1. We will show experimentally (Section 6) that the learned pragmatic speaker S_1 agrees better with human speakers on a referential expression generation task than either the literal speaker S_0 or the pure RSA speaker s_1.

The parameters for L_1 and S_1 are still the parameters of the literal speaker S_0; we wish to update them to maximize the performance of S_1, the agent that acts according to $S_1(m \mid t, c; \theta)$, where

$$S_1(m \mid t, c; \theta) \propto L_1(t \mid m, c; \theta) \tag{7}$$

$$L_1(t \mid m, c; \theta) \propto S_0(m \mid t, c; \theta) \tag{8}$$

This corresponds to the simplest case of RSA in which $\lambda = 1$ and message costs and state priors are uniform: $s_1(m \mid t, \mathcal{L}) \propto l_1(t \mid m, \mathcal{L}) \propto s_0(m \mid t, \mathcal{L})$.

In optimizing the performance of the pragmatic speaker S_1 by adjusting the parameters to the simpler classifier S_0, the RSA back-and-forth reasoning can be thought of as a non-linear function through which errors are propagated in training, similar to the activation functions in neural network models [32]. However, unlike neural network activation functions, the RSA reasoning applies a different non-linear transformation depending on the pragmatic context (sets of available referents and utterances).

For convenience, we define symbols for the log-likelihood of each of these probability distributions:

$$J_{S_1}(t, m, c, \theta) = \log S_1(m \mid t, c; \theta) \tag{9}$$

$$J_{L_1}(t, m, c, \theta) = \log L_1(t \mid m, c; \theta) \tag{10}$$

The log-likelihood of each agent has the same form as the log-likelihood of the literal speaker, but with the value of the distribution from the lower-level agent substituted for the score $\theta^T \phi$. By a derivation similar to the one in (6) above, the gradient of these log-likelihoods can thus

be shown to have the same form as the gradient of the literal speaker, but with the gradient of the next lower agent substituted for the feature values:

$$\frac{\partial J_{S_1}}{\partial \theta} = \frac{\partial J_{L_1}}{\partial \theta}(t, m, c, \theta) - \mathbb{E}_{m' \sim S_1(\cdot | t, c; \theta)}\left[\frac{\partial J_{L_1}}{\partial \theta}(t, m', c, \theta)\right] \tag{11}$$

$$\frac{\partial J_{L_1}}{\partial \theta} = \frac{\partial J_{S_0}}{\partial \theta}(t, m, c, \theta) - \mathbb{E}_{t' \sim L_1(\cdot | m, c; \theta)}\left[\frac{\partial J_{S_0}}{\partial \theta}(t', m, c, \theta)\right] \tag{12}$$

The value J_{S_0} in (12) is as defined in (5).

Training. As mentioned above, our primary objective in training is to maximize the (log) conditional likelihood of the messages in the training examples given their respective states and contexts. We add to this an ℓ_2 regularization term, which expresses a Gaussian prior distribution over the parameters θ. Imposing this prior helps prevent overfitting to the training data and thereby damaging our ability to generalize well to new examples [5]. With this modification, we instead maximize the log of the posterior probability of the parameters and the training examples jointly. For a dataset of M training examples $\langle t_i, m_i, c_i \rangle$, this log posterior is:

$$J(\theta) = -\frac{M}{2}\ell||\theta||^2 + \sum_{i=1}^{M} \log S_1(m_i \mid t_i, c_i; \theta) \tag{13}$$

The stochastic gradient descent (SGD) family of first-order optimization techniques [3] can be used to approximately maximize $J(\theta)$ by obtaining noisy estimates of its gradient and "hill-climbing" in the direction of the estimates. (Strictly speaking, we are employing stochastic gradient *ascent* to maximize the objective rather than minimize it; however, SGD is the much more commonly seen term for the technique.)

The exact gradient of this objective function is

$$\frac{\partial J}{\partial \theta} = -M\ell\theta + \sum_{i=1}^{M} \frac{\partial J_{S_1}}{\partial \theta}(t_i, m_i, c_i, \theta) \tag{14}$$

using the per-example gradient $\frac{dJ_{S_1}}{d\theta}$ given in (11). SGD uses the per-example gradients (and a simple scaling of the ℓ_2 regularization penalty) as its noisy estimates, thus relying on each example to guide the model in roughly the correct direction towards the optimal parameter setting. Formally, for each example (t, m, c), the parameters are updated according to the formula

$$\theta \quad := \quad \theta + \alpha\left(-\ell\theta + \frac{\partial J_{S_1}}{\partial \theta}(t, m, c, \theta)\right) \tag{15}$$

The learning rate α determines how "aggressively" the parameters are adjusted in the direction of the gradient. Small values of α lead to slower learning, but a value of α that is too large can result in the parameters overshooting the optimal value and diverging. To find a good learning rate, we use AdaGrad [9], which sets the learning rate adaptively for each example based on an initial step size η and gradient history. The effect of AdaGrad is to reduce the learning rate over time such that the parameters can settle down to a local optimum despite the noisy gradient estimates, while continuing to allow high-magnitude updates along certain dimensions if those dimensions have exhibited less noisy behavior in previous updates.

5   Example

In Figure 3, we illustrate crucial aspects of how our model is optimized, fleshing out the concepts from the previous section. The example also shows the ability of the trained S_1 model to make a specificity implicature without having observed one in its data, while preserving the ability to produce uninformative attributes if encouraged to do so by experience.

As in our main experiments, we frame the learning task in terms of attribute selection with TUNA-like data. In this toy experiment, the agent is trained on two example contexts, consisting of a target referent, a distractor referent, and a human-produced utterance. It is evaluated on a third test example. This small dataset is given in the top two rows of Figure 3. The utterance on the test example is shown for comparison; it is not provided to the agent.

Our feature representations of the data are in the third row. Attributes of the referents are in SMALL CAPS; semantic attributes of the utterances are in [*square brackets*]. These representations employ the cross-product features described in Section 4; in TUNA data, properties that the target entities do not possess (e.g., ¬GLASSES) are also included among their "attributes."
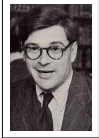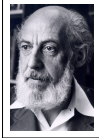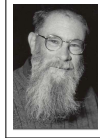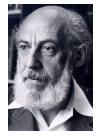
Below the feature representations, we summarize the gradient of the log likelihood ($\frac{\partial J_{S_1}}{\partial \theta}$) for each example, as an $m \times n$ table representing the weight update for each of the mn cross-product features. (We leave out the ℓ_2 regularization and AdaGrad learning rate for simplicity.) Tracing the formula for this gradient (11) back through the RSA layers to the literal listener (5), one can see that the gradient consists of the feature representation of the triple $\langle t, m, c \rangle$ containing the correct (human-produced) message, minus adjustments that penalize the other messages according to how much the model was "fooled" into expecting them.

The RSA reasoning yields gradients that express both lexical and contextual knowledge. From the first training example, the model learns the lexical information that [*person*] and [*glasses*] should be used to describe the target. However, this knowledge receives higher weight in the association with GLASSES, because that attribute is disambiguating in this context. As one would hope, the overall result is that intuitively good pairings generally have higher weights, though the training set is too small to fully distinguish good features from bad ones. For example, after seeing both training examples and failing to observe both a beard and glasses on the same individual, the model incorrectly infers that [*beard*] can be used to indicate a lack of glasses and vice versa. Additional training examples could easily correct this.

Figure 3(b) shows the distribution over utterances given target referent as predicted by the learned pragmatic speaker S_1 after one pass through the data with a fixed learning rate $\alpha = 1$ and no regularization ($\ell = 0$). We compare this distribution with the distribution predicted by the learned literal speaker S_0 and the pure RSA speaker s_1. We wish to determine whether each model can (i) minimize ambiguity; and (ii) learn a prior preference for producing certain descriptors even if they are redundant.

The distributions in Figure 3(b) show that the linear classifier correctly learns that human-produced utterances in the training data tend to mention the attribute [*person*] even though it is uninformative. However, for the referent that was not seen in the training data, the model cannot decide among mentioning [*beard*], [*glasses*], both, or neither, even though the messages that don't mention [*glasses*] are ambiguous in context. The pure RSA model, meanwhile, chooses messages that are unambiguous, but because it has no mechanism for learning from the examples, it does not prefer to produce [*person*] without a manually-specified prior.

Our pragmatic speaker S_1 gives us the best of both models: the parameters θ in learned RSA show the tendency exhibited in the training data to produce [*person*] in all cases, while the RSA recursive reasoning mechanism guides the model to produce unambiguous messages by including the attribute [*glasses*].

Training examples   Test example

	Training examples		Test example

Context

r_2 r_3 r_3 r_4 r_1 r_4

Utterance [*person*] *with* [*glasses*]   [*person*] *with* [*beard*]   [*person*] *with* [*glasses*]

Features
for true
utterance

PERSON $\land$ [*person*]	PERSON $\land$ [*person*]	PERSON $\land$ [*person*]
PERSON $\land$ [*glasses*]	PERSON $\land$ [*beard*]	PERSON $\land$ [*glasses*]
GLASSES $\land$ [*person*]	$\neg$GLASSES $\land$ [*person*]	GLASSES $\land$ [*person*]
GLASSES $\land$ [*glasses*]	$\neg$GLASSES $\land$ [*beard*]	GLASSES $\land$ [*glasses*]
$\neg$BEARD $\land$ [*person*]	BEARD $\land$ [*person*]	BEARD $\land$ [*person*]
$\neg$BEARD $\land$ [*glasses*]	BEARD $\land$ [*beard*]	BEARD $\land$ [*glasses*]

Gradient

	[*person*]	[*glasses*]	[*beard*]		[*person*]	[*glasses*]	[*beard*]
PERSON	**1**	1	-1	PERSON	**1**	-1	1
GLASSES	**2**	**2**	**-2**	GLASSES	0	0	0
BEARD	0	0	0	BEARD	**2**	**-2**	**2**
$\neg$GLASSES	-1	**-1**	**1**	$\neg$GLASSES	1	**-1**	1
$\neg$BEARD	1	**1**	**-1**	$\neg$BEARD	-1	**1**	**-1**

(unused)

(a) Learned S_1 model training. Gradient values given are $6\frac{\partial J_{S_1}}{\partial \theta}$, evaluated at $\theta = \vec{0}$.

s_1		S_0		S_1		
r_1	r_4	r_1	r_4	r_1	r_4	
.08	.25	.03	.00	.10	.11	$\emptyset$
.08	.25	.22	.10	.16	.13	[*person*]
.17	0	.03	.00	.11	.07	[*glasses*]
.08	.25	.03	.04	.08	.17	[*beard*]
.17	0	.22	.01	.18	.08	[*person*], [*glasses*]
.08	.25	.22	.74	.12	.19	[*person*], [*beard*]
.17	0	.03	.00	.10	.11	[*glasses*], [*beard*]
.17	0	.22	.10	.16	.11	[*person*], [*glasses*], [*beard*]

(b) Pure RSA (s_1), linear classifier (S_0), and learned RSA (S_1) utterance distributions. RSA alone minimizes ambiguity but can't learn overgeneration from the examples. The linear classifier learns to produce [*person*] but fails to minimize ambiguity. The weights in learned RSA retain the tendency to produce [*person*] in all cases, while the recursive reasoning yields a preference for the unambiguous descriptor [*glasses*].

Figure 3: Specificity implicature and overgeneration in learned RSA.

6   Experiments

Data.   We report experiments on the TUNA corpus (Section 3 above). We focus on the *singular* portion of the corpus, which was used in the 2008 and 2009 Referring Expression Generation Challenges. We do not have access to the train/dev/test splits from those challenges, so we report five-fold cross-validation numbers. The singular portion consists of 420 *furniture* trials involving 176 distinct referents and 360 *people* trials involving 228 distinct referents.

Evaluation metrics.   The primary evaluation metric used in the attribute selection task with TUNA data is *multiset Dice* calculated on the attributes of the generated messages:

$$\frac{2\sum_{x \in D} \min\left[\mathbb{Z}_{a(m_i)}(x), \mathbb{Z}_{a(m_j)}(x)\right]}{|a(m_i)| + |a(m_j)|} \tag{16}$$

Here, $a(m)$ is the multiset of attributes of message m, D is the non-multiset union of $a(m_i)$ and $a(m_j)$, $\mathbb{Z}_X(x)$ is the number of occurrences of x in the multiset X, and $|a(m_i)|$ is the cardinality of multiset $a(m)$. *Accuracy* is the fraction of examples for which the subset of attributes is predicted perfectly (equivalent to achieving multiset Dice 1).

Experimental set-up.   We evaluate all our agents in the same pragmatic contexts: for each trial in the *singular* corpus, we define the messages M to be the powerset of the attributes used in the referential description and the states T to be the set of entities in the trial, including the target. The message predicted by a speaker agent is the one with the highest probability given the target entity; if more than one message has the highest probability, we allow the agent to choose randomly from the highest probability ones.

In learning, we use initial step size $\eta = 0.01$ and regularization constant $\ell = 0.01$. RSA agents are not trained, but we cross-validate to optimize λ and the function defining message costs, choosing from (i) $C(m) = 0$; (ii) $C(m) = |a(m)|$; and (iii) $C(m) = -|a(m)|$.

Features.   We use *indicator features* as our feature representation; that is, the dimensions of the feature representation take the values 0 and 1, with 1 representing the truth of some predicate $P(t, m, c)$ and 0 representing its negation. Thus, each vector of real numbers that is the value of $\phi(t, m, c)$ can be represented compactly as a set of predicates.

The baseline feature set consists of indicator features over all conjunctions of an attribute of the referent and an attribute in the candidate message (e.g., $P(t, m, c) = \text{RED}(t) \wedge [blue] \in m$). We compare this to a version of the model with additional *generation features* that seek to capture the preferences identified in prior work on generation. These consist of indicators over the following features of the message:

(i)   attribute type (e.g., $P(t, m, c) = $ "m contains a color");

(ii)   pair-wise attribute type co-occurrences, where one can be negated (e.g., "m contains a color and a size", "m contains an object type but not a color"); and

(iii)   message size in number of attributes (e.g., "m consists of 3 attributes").

For comparison, we also separately train literal speakers S_0 as in (4) (the log-linear model) with each of these feature sets using the same optimization procedure.

Table 1: Experimental results: mean accuracy and multiset Dice (five-fold cross-validation). **Bold**: best result; ***bold italic***: not significantly different from best ($p > 0.05$, Wilcoxon signed-rank test).

Model	Furniture		People		All	
	Acc.	Dice	Acc.	Dice	Acc.	Dice
RSA s_0 (random true message)	1.0%	.475	0.6%	.125	1.7%	.314
RSA s_1	1.9%	.522	2.5%	.254	2.2%	.386
Learned S_0, basic feats.	16.0%	.779	9.4%	.697	12.9%	.741
Learned S_0, gen. feats. only	5.0%	.788	7.8%	.681	6.3%	.738
Learned S_0, basic + gen. feats.	**28.1%**	**.812**	17.8%	.730	*23.3%*	*.774*
Learned S_1, basic feats.	23.1%	.789	11.9%	.740	17.9%	.766
Learned S_1, gen. feats. only	17.4%	.740	1.9%	.712	10.3%	.727
Learned S_1, basic + gen. feats.	*27.6%*	.788	**22.5%**	**.764**	**25.3%**	**.777**

Results. The results (Table 1) show that training a speaker agent with learned RSA generally improves generation over the ordinary classifier and RSA models. On the more complex *people* dataset, the pragmatic S_1 model significantly outperforms all other models. The value of the model's flexibility in allowing a variety of feature designs can be seen in the comparison of the different feature sets: we observe consistent gains from adding generation features to the basic cross-product feature set.   Moreover, the two types of features complement each other: neither the cross-product features nor the generation features in isolation achieve the same performance as the combination of the two.

Of the models in Table 1, all but the last exhibit systematic errors. Pure RSA performs poorly for reasons predicted by [16]—for example, it under-produces color terms and head nouns like *desk*, *chair*, and *person*. This problem is also observed in the trained S_1 model, but is corrected by the generation features. On the *people* dataset, the S_0 models under-produce *beard* and *hair*, which are highly informative in certain contexts. This type of communicative failure is eliminated in the S_1 speakers.

The performance of the learned RSA model on the *people* trials also compares favorably to the best dev set performance numbers from the 2008 Challenge [14], namely, .762 multiset Dice, although this comparison must be informal since the test sets are different. (In particular, the Accuracy values given in [14] are unfortunately not comparable with the values we present, as they reflect "perfect match with *at least one* of the two reference outputs" [emphasis in original].) Together, these results show the value of being able to train a single model that synthesizes RSA with prior work on generation.

7   Conclusion

Our initial experiments demonstrate the utility of RSA as a trained classifier in generating referential expressions. The primary advantages of this version of RSA stem from the flexible ways in which it can learn from available data. This not only removes the need to specify a complex semantic lexicon by hand, but it also provides the analytic freedom to create models that are sensitive to factors guiding natural language production that are not naturally expressed in standard RSA.

This basic presentation suggests a range of potential next steps. For instance, it would be

natural to apply the model to pragmatic interpretation (the listener's perspective); this requires no substantive formal changes to the model as defined in Section 4, and it opens up new avenues in terms of evaluating pragmatic models in standard classification tasks like sentiment analysis, topic prediction, and natural language reasoning. In addition, for all versions of the model, one could consider including additional hidden speaker and listener layers, incorporating message costs and priors into learning, to capture a wider range of pragmatic phenomena.

References

[1] Peter Baumann, Brady Clark, and Stefan Kaufmann. Overspecification and the cost of pragmatic reasoning about referring expressions. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, 2014.

[2] Leon Bergen, Roger Levy, and Noah D. Goodman. Pragmatic reasoning through semantic inference. Ms., MIT, UCSD, and Stanford, 2014.

[3] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In Yves Lechevallier and Gilbert Saporta, editors, *Proceedings of the 19th International Conference on Computational Statistics*, pages 177–186. Springer, Berlin, 2010.

[4] Colin F. Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898, 2004.

[5] Stanley F. Chen and Ronald Rosenfeld. A Gaussian prior for smoothing maximum entropy models. Technical Report CMU-CS-99-108, Carnegie Mellon University, 1999.

[6] Robert Dale and Ehud Reiter. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263, 1995.

[7] Judith Degen and Michael Franke. Optimal reasoning about referential expressions. In Sarah Brown-Schmidt, Jonathan Ginzburg, and Staffan Larsson, editors, *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue*, pages 2–11, 2012.

[8] Pedro Domingos. A few useful things to know about machine learning. *Communications of ACM*, 55(10):78–87, 2012.

[9] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, pages 2121–2159, 2011.

[10] Paul E. Engelhardt, Karl G.D. Bailey, and Fernanda Ferreira. Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language*, 54(4):554–573, 2006.

[11] Michael C. Frank and Noah D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998, 2012.

[12] Michael C. Frank and Noah D. Goodman. Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, 75(1):80–96, 2014.

[13] Michael Franke. *Signal to Act: Game Theory in Pragmatics*. ILLC Dissertation Series. Institute for Logic, Language and Computation, University of Amsterdam, 2009.

[14] Albert Gatt, Anja Belz, and Eric Kow. The TUNA Challenge 2008: Overview and evaluation results. In *Proceedings of the Fifth International Natural Language Generation Conference*, INLG '08, pages 198–206. Association for Computational Linguistics, 2008.

[15] Albert Gatt, Ielka van der Sluis, and Kees van Deemter. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 49–56. DFKI GmbH, 2007.

[16] Albert Gatt, Roger P. G. van Gompel, Kees van Deemter, and Emiel Krahmer. Are we Bayesian referring expression generators? In *Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society*, 2013.

[17] Dave Golland, Percy Liang, and Dan Klein. A game-theoretic approach to generating spatial descriptions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 410–419. Association for Computational Linguistics, 2010.

[18] H. Paul Grice. Logic and conversation. In Peter Cole and Jerry Morgan, editors, *Syntax and Semantics*, volume 3: Speech Acts, pages 43–58. Academic Press, 1975.

[19] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, Berlin, 2 edition, 2009.

[20] Gerhard Jäger. Game theory in semantics and pragmatics. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: An International Handbook of Natural Language Meaning*, volume 3, pages 2487–2425. Mouton de Gruyter, 2012.

[21] Gerhard Jäger. Rationalizable signaling. *Erkenntnis*, 79(4):673–706, 2014.

[22] Justine T. Kao, Leon Bergen, and Noah D. Goodman. Formalizing the pragmatics of metaphor understanding. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 719–724. Cognitive Science Society, 2014.

[23] Justine T. Kao, Jean Y. Wu, Leon Bergen, and Noah D. Goodman. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33):12002–12007, 2014.

[24] Willem JM Levelt. *Speaking: From Intention to Articulation*, volume 1. MIT press, 1993.

[25] David Lewis. *Convention*. Harvard University Press, 1969. Reprinted 2002 by Blackwell.

[26] Percy Liang and Christopher Potts. Bringing machine learning and compositional semantics together. *Annual Review of Linguistics*, 1(1):355–376, 2015.

[27] Peter McCullagh and John A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1989.

[28] Brian McMahan and Matthew Stone. A Bayesian model of grounded color semantics. *Transactions of the Association for Computational Linguistics*, 3:103–115, 2015.

[29] Thomas Pechmann. Incremental speech production and referential overspecification. *Linguistics*, 27(1):89–110, 1989.

[30] Christopher Potts, Daniel Lassiter, Roger Levy, and Michael C. Frank. Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. To appear in *Journal of Semantics*, 2015.

[31] Seymour Rosenberg and Bertram D. Cohen. Speakers' and listeners' processes in a word communication task. *Science*, 145:1201–1203, 1964.

[32] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning internal representations by error propagation. In David E. Rumelhart and James L. McClelland, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1: Foundations, pages 318–362. MIT Press, Cambridge, MA, 1986.

[33] Alex Stiller, Noah D. Goodman, and Michael C. Frank. Ad-hoc scalar implicature in adults and children. In Laura Carlson, Christoph Hoelscher, and Thomas F. Shipley, editors, *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, pages 2134–2139. Cognitive Science Society, 2011.

[34] Kees van Deemter, Ielka van der Sluis, and Albert Gatt. Building a semantically transparent corpus for the generation of referring expressions. In *Proceedings of the Fourth International Natural Language Generation Conference*, pages 130–132. Association for Computational Linguistics, 2006.

[35] Adam Vogel, Max Bodoia, Christopher Potts, and Dan Jurafsky. Emergence of Gricean maxims from multi-agent decision theory. In *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1072–1081. Association for Computational Linguistics, 2013.

[36] Adam Vogel, Andrés Gómez Emilsson, Michael C. Frank, Dan Jurafsky, and Christopher Potts. Learning to reason pragmatically with cognitive limitations. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 3055–3060. Cognitive Science Society, 2014.

Some New Thoughts on Conditionals

Graham Priest

Philosophy Programmes, CUNY Graduate Center, and the University of Melbourne
priest.graham@gmail.com

1 Introduction

A conditional expresses a connection of some kind between two propositions or states of affairs. The relationship is some kind of dependence; but what, exactly, that is, is, of course, the 64 thousand dollar question. The canonical expression of a conditional in English is of the form: If X then Y.[1] But conditionals can be expressed without using 'if' (were I younger, I would go out rocking every night); and not everything which uses the word 'if' is a conditional (if I may say so, you are looking stunning today). The canonical construction suggests that there is only one relation of conditionality. This may be the natural default assumption but, of course, it may be wrong, as many have supposed.

Indeed, nearly everything about the nature of conditionals is philosophically contentious. The consensus of the 1960s concerning the simple-minded theory of the material conditional has blown apart, leaving no present consensus.

This paper is hardly an attempt to solve all of the many issues concerning conditionals. I doubt that anyone is able to do this. Rather, what I wish to do is put conditionals in a new perspective—one which seems to be relatively simple, natural, and provides a straightforward solution to some standard tangles.

2 Conditionals and Imported Information

For a start, some have argued that conditionals are not truth-apt. This, however, cannot be right. Conditional can occur embedded in contexts which require the embedded sentence to be truth-apt, such as: 'Mary believes that if she goes to the party she will have fun' and 'It is possible that if she goes she will have fun'. Conditionals must, then, have truth conditions. A natural thought is that we evaluate a conditional, 'if A then B' by considering situations in which A is true, and seeing if B is true in these. But which situations? Not all of them. Certain information carries over from the actual world, and must hold in them. Thus, consider the conditional 'If global warming continues at its present pace, sea levels will rise by at least two metres before the end of the century'. We are assuming that in this hypothetical—and hopefully (but increasingly unlikely) counter-factual—situation, the laws of physics, and notably those concerning geo-meteorology, are the same as those of the actual world.

Let us call the information that is carried over *the imported information*. It is to be noted that information is imported in a quite different context: determining what holds in a work of fiction. Given a work of fiction, many of the things that hold, hold because of the explicitly say-so of the author. Thus, in the worlds that realise the Holmes novels, Holmes lives in Baker St, because Doyle tell us so. But it is also true that one can't travel from from London to Edinburgh in an hour, that large doses of arsenic kill people, and so on. Doyle never says

---

[1] The verb of X may be in indicative or subjunctive moods; that of Y can be in other moods, such as interrogative or imperative. How this is possible has to be part of a full story of conditionals, but I will ignore these other moods in what follows.

these things. They are just imported from the facts about the world—or at least, the world of Britain circa 1900. Now, though conditionals and truth in fiction are different issues, it appears to me that the phenomenon of importation is exactly the same in both of them. If, therefore, one could solve the problem of what information, exactly, is imported into a work of fiction, one would have solved the problem of what is imported into an antecedent conditional situation—and vice versa.

Call this the *importation problem*. If we had a solution to it, we would have gone a long way towards answering the 64 thousand dollar question. I'm afraid that I don't (at least presently—one can always hope!). But even without a precise answer, a couple of things are evident.

First, there would appear to be no reason to suppose that irrelevant matters get imported. Thus, it is true that Graham Priest was born in London in 1948. Yet both of the following would appear to be false: In the Holmes novels, Graham Priest would be born in London in 1948; if Emile Zola had written the Holmes novels, Graham Priest would have been born in London in 1948.

Secondly, and most importantly, what information is imported is context-dependent. Thus, suppose that we are driving on a freeway, and the topic of discussion is high-speed transport. You might say 'if this car were a photon, then some cars would travel at about 3×10^8 m/sec'. What is being imported here is the fact that photons travel with the speed of light. But if the topic of discussion were, instead, a hypothetical physics, you might say 'If this car were a photon, then some photons would travel at about 3 m/sec'. Here, what is being imported is the fact that the car is travelling at 100 km/h. I note that the information that is imported may depend on what those who find themselves in the context concerned know. It is not imported simply because they know it, however—much less believe it to be true. It is imported because it *is* true, and bears on the hypothetical scenario envisaged.

3   A Semantics

One way to make these thoughts formally precise is fairly standard.[2] A propositional language contains the connectives $\wedge$, $\neg$, and $>$. $>$ is the conditional. $\vee$ and $\supset$ may be defined in the usual way. The set of formulas is F. An interpretation is a structure $\langle W, \{R_A : A \in F\}, \nu \rangle$. W is a set of worlds, or situations (hypothetical or otherwise). For every $A \in F$, R_A is a binary relation on W; wR_Aw' may be thought to express the fact that w' is a world at which A is true, and at which all the information imported from w holds. ν is a function which maps every world, w, and every propositional parameter, p, to either 1 or 0; we write this $\nu_w(p) = 1$ (or 0). As I noted, what information imports, and so R_A, depends on the context, c. So the R's may be thought of as dependent on a context parameter, c. However, this plays no role in the formal semantics, so I omit mention of it.

Truth at a world, $\Vdash$, is now defined as follows:

- $w \Vdash p$ iff $\nu_w(p) = 1$

- $w \Vdash \neg A$ iff it is not the case that $w \Vdash A$

- $w \Vdash A \wedge B$ iff $w \Vdash A$ and $w \Vdash B$

- $w \Vdash A > B$ iff for all w' such that wR_Aw', $w' \Vdash B$

---

[2]See Priest (2008), ch. 4. To handle the semantics of counter-logicals properly, the semantics need to be expanded to include impossible worlds, as in ch. 10 (10.7). However, I ignore this point here.

An inference from premises, Σ, to conclusion, A, is valid, $\Sigma \models A$ iff:

- for any interpretation, and $w \in W$: if $w \Vdash B$ for every $B \in \Sigma$, $w \Vdash A$.

These semantics give the basic conditional logic, C. No constraints are put on the R_As. The intuitive interpretation motives some constraints, however. The first is that:

- if $wR_A w'$ then $w' \Vdash A$

for w' is one of the worlds where A holds. This verifies $\models A > A$. The second is:

- if $w \Vdash A$ then $wR_A w$

for if A is true at w, then whatever information is imported from w, it is true at w; hence, w is one of the worlds that w accesses under R_A. This constraint validates: $A, A > B \models B$.

Thus the logic generated by the intuitive understanding explained is at least as strong as C^+. Whether the understanding motivates other constraints, I leave as an open question.

4   East Gate, West Gate

I want to turn, instead, to the issue of how some of the high-profile examples in the literature are accommodated by the understanding I have described.

The first concerns "Gibbardian standoffs", formulated originally by Gibbard (1981). I take the example as cleaned up by Bennett, who explains the scenario as follows.[3]

> Top Gate holds back water in a lake behind a dam; a channel running down from it splits into two distributaries, one (blocked by East Gate) running eastwards and the other (blocked by West Gate) running westwards. The gates are connected as follows: if east lever is down, opening Top Gate will open East Gate so that the water runs eastwards; and if west lever is down, opening Top Gate will open West Gate so that the water will run westwards. On the rare occasions when both levers are down, Top Gate cannot be opened because the machinery cannot move three gates at once.
>
> Just after the lever-pulling specialist has stopped work, Wesla knows that west lever is down, and thinks 'If Top Gate is open, all the water will run westward'; Esther knows that east lever is down, and thinks 'If Top Gate is open, all the water will run east'.

Both Esther and Wesla seem to speak truth, though they appear to disagree with each other. How is this possible? Moreover Southie, who also knows the connections between the gates and the levers, knows nothing of the current settings. Southie can, however, hear what Esther and Wesla say, and knows them to be reliable. Southie concludes that Top Gate is closed. How so?

Take Esther first. In the context in which she finds herself, the information available to her is that the east lever is down. So this information may import into any hypothetical situation she considers. She considers a scenario in which Top Gate is open, and imports the information that east lever is down. In such situations, the water will flow east. Hence she says: If Top Gate is open, all the water will flow east. The situation with Wesla is exactly the same, except that in the context in which he finds himself, the information available to him is that the west

---
[3]Bennett (2003), p. 85.

lever is down. Both Esther and Wesla speak truly. Their different contexts deliver different importing information.

Next, consider Southie. One might suppose that Southie reason as follows:

- We know, by testimony, that if Top Gate is open the water will flow east, and if Top Gate is open the water will flow west. It cannot flow both east and west, so Top Gate must be closed.

Such reasoning is incorrect, since the two conditionals are true in different contexts, and cannot be conjoined. In the same way, if Alfie, in New York, says 'It's 04.00h', and Beth in London says 'It's 09.00h', Costa in Melbourne, who hears them both, cannot conclude 'Its 04.00h and 09.00h'. What is going on is this. Southie knows that Esther speaks what she knows to be true. The only way for her to say what she says is that she knows that east lever is down. If that were not the case, she could not import it into her hypothetical situation to conclude that, in it, the water is flowing east. Symmetrically, Southie knows from what Wesla says, that west lever is down. On the basis of this, and knowing the facts about the connections between levers and gates, Southie concludes that Top Gate is closed.

5   Interlude: the English Subjunctive

The second example I want to discuss in the notorious Oswald/Kennedy pair, put forward originally by Adams (1970). But first, an interlude on the English subjunctive. The English subjunctive mood is vestigial, and is also the linguistic analogue of an endangered species. However, to the extent that it is still extant, it works like this.

English has only two tenses: present, *I love*, and past (imperfect), *I loved*. Things which are expressed by grammatical tenses in many other languages are expressed in English by using auxiliary verbs, notably *have* and *will*. So we have future, *I will love*, (past) perfect, *I have loved,* pluperfect, *I had loved*, future perfect, I *will have loved.*

Each of the two tenses has an indicative mood and a subjunctive mood. Take the present tense first. For regular verbs, the present subjunctive is the same as the infinitive, (*to*) *love.* But so is the indicative in all persons, except the third person singular, where one adds an *s*. So the only person in which one can tell the difference is the third person singular: *she loves you* (indicative); *I would that she love me* (subjunctive).

The most irregular verb in English is (*to*) *be.* Here, none of the persons in the indicative is the same as the infinitive (*am/are/is*, *are/are/are*). The subjunctive is, however, as in regular verbs: that is, the same as the infinitive. So the difference between indicative and subjunctive shows up in all persons. I *am*, *I be*; *she is*, *she be*; *they are*, *they be.*[4]

Turning to the past tense: in regular verbs, the past subjunctive is the same as the past indicative (and so the same in all persons).[5] However, again, the verb (*to*) *be* is irregular, and the indicative conjugates (was/were/was, were/were/were). The subjunctive in all persons is *were*. So the difference shows up in the first and third persons singular.[6]

6   Oswald and Kennedy

We now turn to the oft' cited pair:

---

[4]As in the subjunctives: If I /she be allowed to speak my/her mind, it will be a very interesting occasion.
[5]As in: I loved her. I would that she loved me.
[6]As in: He told me she was married. I wish that it were not so.

[1] If Oswald didn't shoot Kennedy, someone else did

[2] If Oswald hadn't shot Kennedy, someone else would have

It is usually claimed that the verb of the antecedent in [1] is indicative, and that that in [2] is subjunctive. It is also claimed that the two have the same antecedent. But the first is true, and the second is false (assuming the results of the Warren commission). Hence, there are two kinds of conditionals: indicative and subjunctive.

Is this so? Consider [1]. We may agree that the verb *did* is a past indicative. To evaluate the conditional, we consider a possible situation in which Oswald didn't shoot Kennedy. We import the information that someone shot Kennedy. So in that situation someone else shot Kennedy. So [1] is true.

What of [2]? For a start, it is not entirely obvious that the verb in the antecedent of [2] is in the subjunctive mood. The main verb is *to have*, which appears in the past tense, *had*. This could be indicative or subjunctive. However, [2] may naturally be rephrased as:

[3] If Oswald were not to have shot Kennedy, somone else would have

The main verb here is the past tense of *to be*; and we have a third person singular. So *were* is the subjunctive.

So far, so good. How do we evaluate [2]? Someone who says this, would appear to be saying exactly the same as someone in the past—just prior to the time of the shooting of Kennedy—who says:[7]

[4] If Oswald does not shoot Kennedy, someone else will.

It would appear, then, that the tense and mood of [2] conspire to take [4], and move its point of evaluation to a past point in time. That is, [2] is the tense of [4]. Generally, 'if *A* were to have been the case, *B* would have been the case' is the past tense of 'If *A* is the case, *B* will be the case'. Call this the *Backshift Thesis*.

Given the Backshift Thesis, we evaluate [2] as follows. We go back to a time just prior to the time at which Kennedy was shot, and evaluate [4] there. We consider a situation where Oswald does not shoot Kennedy. We import what we know from the Warren commission, that Oswald was acting alone. So in that situation, it is false that someone else will shoot Kennedy. So [4] is false of that time, and [2] is false of now.

The past subjunctive does not, then, deliver a different kind of conditional. The moods and tenses of the verbs in the conditional merely conspire to form the past tense of a conditional.[8] [1] and [2] differ in truth value, since the temporal shift makes it natural to import different information into their antecedents.

7   The Backshift Thesis

One might well doubt the Backshift Thesis. Here is a putative counter-example, put to me by Hartry Field.[9]

Professor *X* is doing an experiment to detect a mooted particle, the tachyon. He sets up an experimental device, which gives a positive result. He exclaims happily (and truly):

---

[7]Or with a present subjunctive: 'If Oswald shoot not Kennedy, someone else will.

[8]The *general* behaviour and import of tenses and moods in conditionals is a very tricky subject which, fortunately, we may avoid here.

[9]Hartry and I taught a course on Conditionals in New York in the Fall of 2014. Thanks go to him for many enjoyable and insightful conversations.

[5] If the apparatus is working correctly, we will be justified in believing that there are tachyons.

Later he discovers that the apparatus was not working correctly, and whether there are tachyons is still unknown. The Backshift Thesis says that what [5] expresses at the time, is expressed later by:

[6] If the apparatus had been working correctly, we would have been justified in believing that there are tachyons.

But this is false. Had the apparatus been working correctly, it might or might not have shown a positive result; so we might or might not have been justified in believing that there are tachyons.

However, let us pay careful attention to the information that is imported in each conditional. In its context, the natural understanding of [5] imports information including that the experiment has given positive results. To evaluate it, we consider a world where the apparatus is working correctly, add the information that it gives a positive result, and the existence of a justification follows. However, with the same importation, [6] is also true. Had the apparatus been working correctly, then, given that it had positive results, we would have been justified in believing there to be tachyons.

In its context, the natural understanding of [6] imports information including that it is not known whether or not there are tachyons. So, in some hypothetical situations where the apparatus was working correctly, the results are positive; and in some they are negative. It is not the case in all of them that we have good reason to believe that there are tachyons. But with the same importation, [5] is also false. If the apparatus is working correctly, and we do not know whether or not there are tachyons, it does not follow that we will have good reason to believe that they exist. We just do not know what the outcome of the experiment will be.

[5] and [6] therefore stand or fall together. If we import the information that the results were positive, both stand; if we import the information that the existence of tachyons is unknown, both fall. Granted, it is more *natural* to import different information in the two cases. Be that as it may, the apparent difference between [5] and [6] is not due to the falsity of the Backshift Thesis, but to the change in context which motivates different imported information.

8   Present Subjunctives

I have argued that in the Oswald/Kennedy example, the subjunctive antecedent does not betoken a different kind of conditional. It merely shifts the point of evaluation to the past.

If the mere fact that the verb of an antecedent is in the subjunctive mood delivered a different kind of conditional, one might expect to find this with present subjunctives, just as much as past subjunctives. We do not. There is no significant difference between: 'If Julie goes to the party, she will have fun' and 'if Julie go to the party, she will have fun'. To evaluate both conditionals, we consider situations where Julie goes to the party, we import what we know about what sorts of things will happen at the party, what sort of person Julie is, and see whether she will have fun there. The difference between the two conditionals, if there is one, is that with the subjunctive mood, the speaker expresses more hesitation about whether they expect the antecedent situation to be realised.

Some, however, have claimed to find a difference in conditionals even when the subjunctive is a present subjunctive. Edgington gives the following example:[10]

----

[10]Edgington (1995), p. 239. I have changed her numbering.

[T]here are two prisoners, Smith and Jones. We have powerful evidence that one of them will try to escape tonight. Smith is a docile, unadventurous chap, Jones just the opposite, and very persistent. We are inclined to think that it is Jones who will try to escape. We have no reason to accept:

[7] If Jones were not to try to escape tonight, Smith would.

However, we could be wrong in thinking that it is Jones who will escape:

[8] If Jones doesn't try to escape tonight, Smith will.

So [7] is false, but [8] is true. But what is making the difference here is not the subjunctive, but the information being imported. In [8] we import the information that one of Jones and Smith will try to escape tonight, so in a situation where Jones does not try to escape, Smith does. But if we import the same information into [7], the result is exactly the same. In [7] the natural imported information is that Smith is not the kind of person to try to escape. So in a scenario where Jones does not try to escape, no one does. But if one imports the same information into [8], it is false for exactly the same reason. Perhaps it is more natural to make different importations in the two cases, but one can hear each conditional in both ways.

9   Conclusion

I summarise the main points of this essay. The truth value of a conditional depends on the information which is imported from the actual situation, which is added to that in the antecedent. (And the information concerns what is true, not what is held to be true.) If in all situations that realise both, the consequent is true, so is the whole conditional. If in some it is false, so is the whole conditional. What information is imported is context-dependent, and may change depending on the interests, knowledge, etc. of those using the conditionals.

The idea explains naturally what is going on in some high-profile examples from the literature—perhaps most notably, where there appears to be a difference between conditionals whose antecedents are indicative and conditionals whose antecedents are subjunctive. Past subjunctives indicate a temporal backshift of the point of evaluation, and so affect the information imported. Present subjunctives have no such effect.

I am well aware that this essay is nothing more than the beginning of a discussion. I am sure, for example, that there are many other examples of conditionals that could profitably be examined, and much that could be learned from them. If I have done enough in this essay to make its central ideas worthy of further investigation, I am content. (That's a conditional.)

References

[1] Adams, E. (1970), 'Subjunctive and Indicative Conditionals', *Foundations of Language* 6: 89–94.

[2] Bennett, J. (2003), *A Philosophical Guide to Conditionals*, Oxford: Oxford University Press.

[3] Edgington, D. (1995), 'On Conditionals', *Mind* 104: 235-309.

[4] Gibbard, A. (1981), 'Two Recent Theories of Conditionals', pp. 211-47 of Harper, W. L., Stalnaker, R., and Pearce, G. (eds), *Ifs*, Dordrecht: D. Reidel Publishing Co.

[5] Priest, G. (2008), *Introduction to Non-Classical Logic*, 2nd edn, Cambridge: Cambridge University Press.

The conservativity of *many*[*]

Maribel Romero

University of Konstanz, Konstanz, Germany
`maribel.romero@uni-konstanz.de`

Abstract

Besides their conservative cardinal and proportional meanings, *many* and *few* have been argued to allow for a 'reverse' proportional reading that defies the convervativity universal (Westerståhl, 1985). We develop a compositional analysis that derives the correct truth conditions for this reading while maintaining conservativity. First, an amendment is proposed to Cohen's (2001) reverse proportional truth conditions. Second, mirroring the decomposition of other degree expressions like *tall*, *many* is decomposed into the parametrized determiner MANY and *POS*. *POS* is allowed to scope out of its host and scope sententially, and a comparison class C is retrieved via the (focus or contrastive topic) associate of *POS*. Keeping a unified conservative denotation for proportional MANY, the regular proportional reading obtains when *POS*' associate is external to the original host NP and the reverse proportional reading arises when it is internal to the host NP. The same applies to *few*.

1 Introduction

The study of generalized quantifiers has been a successful enterprise in semantic theory over several decades. One of its most important insights is that natural language determiners cannot denote just any function in $D_{<et,<et,t>>}$ but only those functions that satisfy certain constraints. Conservativity, defined in (1), is one of the constraints that have been argued for (Keenan & Stavi 1986; Barwise & Cooper 1981, U3; van der Does & van Eicjk 1996):

(1) A determiner denotation $f \in D_{<et,<et,t>>}$ is conservative iff, for any P and $Q \in \mathrm{D}_{<e,t>}$:
$f(P)(Q) = 1$ iff $f(P)(P \cap Q) = 1$

(2) Conservativity Universal:
Determiners in natural language are always interpreted as conservative functions.

An interesting case concerns the determiners *many* and *few*. Partee (1988) (and a long tradition thereafter) distinguishes two readings: the cardinal reading (3a)-(4a) and the proportional reading (3b)-(4b). To see these readings exemplified, consider sentences (6)-(7) in scenario (5). Sentence (6) is judged true in virtue of its cardinal reading and sentence (7) in virtue of its proportional reading. Once the context-dependent parameters n and p have been fixed for a given context, the functions denoted by $many_{\mathrm{card/prop}}$ and $few_{\mathrm{card/prop}}$ are conservative.

(3) Many *P*s are *Q*.
  a. CARDINAL reading: $|P \cap Q| > n$, where n is a large natural number.
  b. PROPORTIONAL reading: $|P \cap Q| : |P| > p$, where p is a large proportion.

(4) Few *P*s are *Q*.
  a. CARDINAL reading: $|P \cap Q| < n$, where n is a small natural number.

---

[*]Many thanks Doris Penka, Sven Lauer, Bernhard Schwarz and Lucas Champollion for their valuable questions and comments. Thanks to the audience of *NELS 46* for their useful input. Remaining errors are mine.

    b.   PROPORTIONAL reading: $|P \cap Q| : |P| < p$, where p is a small proportion.

(5)   Scenario: All the faculty children were at the 1980 picnic, but there were few faculty children back then. Almost all faculty children had a good time.

(6)   There were few faculty children at the 1980 picnic.

(7)   Many faculty children had a good time.

However, Westerståhl (1985) famously noted an additional reading of *many*, the so-called "reverse" proportional reading. Besides its regular proportional reading, which is false in scenario (8) (since, among all the Scandinavians, 14 does not count as many), sentence (9) has another proportional reading roughly paraphrasable as 'Many winners of the Nobel Prize in literature are Scandinavians' that makes it true in that scenario. The same point has been made for *few* (Herburger, 1997): Sentence (10) has a reading paraphrasable as 'Few applicants were cooks'. Formalizing these intuitive paraphrases gives us the truth conditions in (11)-(12). Crucially, these truth conditions render $many_{\text{rev-prop}}$ and $few_{\text{rev-prop}}$ non-conservative.

(8)   Scenario: Of a total of 81 Nobel Prize winners in literature, 14 come from Scandinavia.

(9)   Many Scandinavians have won the Nobel Prize in literature.

(10)   Few cooks applied.

(11)   Many Ps are Q.
      REVERSE PROP. reading: $|P \cap Q| : |Q| > p$, where p is a large proportion.

(12)   Few Ps are Q.
      REVERSE PROP. reading: $|P \cap Q| : |Q| < p$, where p is a small proportion.

Efforts have been made in the literature to derive the reverse proportional reading of *many* and *few* in a principled way (Cohen, 2001; Herburger, 1997; de Hoop & Solà, 1996, a.o.), the key issue being whether, in such a principled derivation, the determiners remain conservative or challenge the conservativity universal.

The goal of this paper is two-fold: (i) to clarify the exact truth conditions of the reverse proportional reading and, (ii) building on Romero (2015), to derive these truth conditions compositionally while maintaining conservativity.

For the first goal, we will propose an amendment to Cohen's (2001) truth conditions.

For the second goal, the point of departure is the observation in the literature that the reverse proportional reading is available only if (part of) the N' complement of the determiner is focused (F) (Herburger, 1997) or functions as contrastive topic (CT) (Cohen, 2001), as indicated in (13)-(14). In a nutshell, our proposal will be the following. Just like degree adjectives like *tall* decompose into the stem TALL and the positive degree operator *POS* (Heim (2006); von Stechow (2009)), the determiners *many* and *few* decompose into MANY+*POS* and FEW+*POS* respectively. The determiners MANY and FEW will be defined as conservative. *POS* in determiners does exactly what it does in adjectives: it scopes out of its host and combines with the appropriate comparison class C via an associate, which we will implement as a F- or CT-associate.[1] Crucially, the F/CT associate of *POS* may lie outside its original host –as noted in the literature– or inside it –as we will observe in this paper. We will show that whether we obtain the regular or the reverse proportional reading of *many* / *few* depends on whether the F/CT-associate of *POS* is external or internal to the original host NP.

---

[1] I will talk about the F/CT associate of *POS* loosely, without commitment as to whether *POS* is conventionally or non-conventionally F- (or CT-) sensitive (see Beaver & Clark (2008)).

(13)   Many Scandinavians$_{\mathbf{F/CT}}$ have won the Nobel Prize in literature.

(14)   Few cooks$_{\mathbf{F/CT}}$ applied.

The rest of the paper is organized as follows. Section 2 takes a closer look at the truth conditions corresponding to the reverse proportional reading. Section 3 provides some background on *POS* with adjectives and presents the novel observation that *POS*'s associate can be internal to the original host NP. Section 4 spells out the proposal. Section 5 examines some further predictions. Section 6 concludes.

2   Truth conditions of the reverse proportional reading

We start with the truth conditions suggested by Westerståhl's (1985) intuitive paraphrase:

(15)   Westerståhl (1985):
  a. Paraphrase: 'Many of the Nobel Prize winners are Scandinavians.'
  b. REVERSE PROPORTIONAL reading of *Many Ps are Q*:
      $|P \cap Q| : |Q| > p$, where p is a large proportion.

Cohen (2001) shows that this characterization of the reading is incorrect: the truth conditions in (15b) make no reference to the proportion $|P \cap Q| : |P|$, but this proportion matters. To see this, consider scenario (16). While two Andorrans having won the prize suffices to make sentence (17) true, it is doubtful that the same number renders sentence (9) true. Yet, the formalization in (15b) only asks us to consider the proportion of winners that are Andorrans/Scandinavians (i.e., $|P \cap Q| : |Q|$), which is 2/112 for either sentence. Thus, for a contextually given value of p, (15b) wrongly predicts both sentences to be have the same truth value. What these examples show is that the proportion of Andorrans/Scandinavians that have won the prize (namely 2/60,000 vs. 2/20,000,000) plays a decisive role.

(16)   Scenario: 112 Nobel Prize winners in literature. 2 out of a total of 60,000 Andorrans have won it. 2 out of a total of 20,000,000 Scandinavians have won it.

(17)   Many ANDORRANS have won the Nobel Prize in literature.

To solve this problem, Cohen (2001) factors $|P \cap Q|{:}|P|$ into the truth conditions of *many*. Furthermore, P is argued to function as a contrastive topic and to invoke a set of alternatives ALT(P), to which $\cup$ is applied to yield $\cup$ALT(P). In our examples, ALT(P) is the set {Scandinavian, Mediterranean, Middle Eastern, Andorran, . . . } and $\cup$ALT(P) is the set containing the world population. Cohen's intuitive paraphrase and proposed truth conditions are in (18). These truth conditions still render *many* non-conversative.

(18)   Cohen (2001):
  a. Paraphrase: 'The proportion of Scandinavians that have won the Nobel Prize in literature is large compared to the proportion of the world population that have won the Nobel Prize in literature.'
  b. REVERSE PROPORTIONAL reading of *Many Ps are Q*:
      $|P \cap Q| : |P| > |\cup \mathrm{ALT}(P) \cap Q| : |\cup \mathrm{ALT}(P)|$

We point out that Cohen's characterization of the reverse proportional reading is not fully correct either: (18) makes no use of the point-wise alternatives $|P' \cap Q|{:}|P'|$, $|P'' \cap Q|{:}|P''|$,

$|P''' \cap Q|:|P'''|$, etc., but these alternatives matter. To see this, consider sentence (19) on the two scenarios below. Under the regular proportional reading, the sentence is false in both scenarios (since, among the 1000 students in this school, 8 does not count as many). But, under the reverse proportional reading, the truth value intuitively differs in the two scenarios, even though the proportion of students of this school that got an A (namely, 8/1000) and the overall proportion of students in this town that got an A (namely, 140/24000) is the same in both scenarios. The truth values differ because the distribution of the alternative proportions matters. In scenario (20), the distribution of A-students per school peeks at the interval [5, 6]. This makes 8 A-students count as many and the sentence is judged true. In scenario (21), the distribution of A-students peeks at the interval [6, 7, 8, 9]. This makes 8 A-students hardly count as many and thus the sentence is intuitively judged false.

(19)   Many students in this school got an A on the final exam.

(20)   Scenario: 24 schools in this town, with 1000 students each. 140 out of the total 24000 students in this town got an A on the final exam. In the school we are referring to, 8 of the 1000 students got an A. For most schools, the number of students that got an A ranges between 5 and 6, e.g. as in Fig. 1.

(21)   Scenario: 24 schools in this town, with 1000 students each. 140 out of the total 24000 students in this town got an A on the final exam. In the school we are referring to, 8 of the 1000 students got an A. For most schools, the number of students that got an A ranges between 6 and 9, e.g. as in Fig 2.
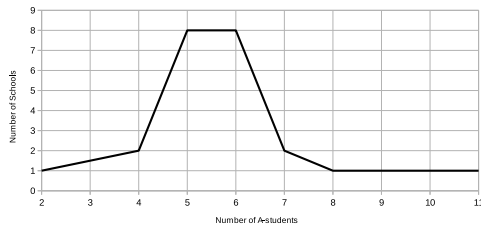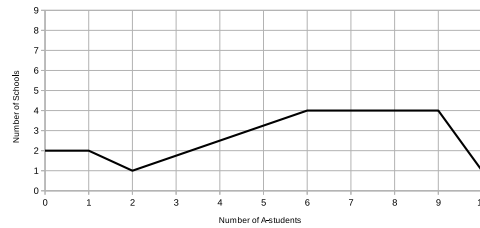
Figure 1   Figure 2

To solve this problem, the alternative proportions have to be factored in, as paraphrased and formalized in (22). The function θ combines with the set containing all these alternative proportions and yields a threshold value for that set, to which the original proportion $|P \cap Q| : |P|$ is compared.[2]

(22)   a.   Paraphrase: 'The proportion of Scandinavians that have won the Nobel Prize in literature is large compared to a threshold based on the proportions of inhabitants of other worlds regions that have won the Nobel Prize in literature.'

       b.   REVERSE PROPORTIONAL reading of *Many Ps are Q*:
            $$|P \cap Q|:|P| \quad > \quad \theta(\{|P' \cap Q|:|P'| : P' \in \mathrm{ALT}(P)\})$$

Note that these truth conditions still make reverse proportional *many* non-conservative. This takes us to our second goal: to arrive a these correct truth conditions compositionally while maintaing that all natural language determiners denote conservative functions.

---

[2]We leave open what mathematical operations θ applies to that set to obtain the threshold value. See Schöller & Franke (2015) for an algorithm compatible with (22), developing ideas from Fernando & Kamp (1996).

3 *POS* with adjectives

3.1 Background on *POS* with adjectives

Adjectives may appear in the comparative, superlative or positive form. This gives us the family of degree operators defined in (23)-(25), where $Q_{<d,t>}$ in *-er* corresponds to the second comparison term and $\boldsymbol{Q}_{<dt,t>}$ in *-est* and *POS* corresponds to the comparison class (Heim, 1999, 2006; von Stechow, 2009). For *POS*, L takes a set of sets of degrees on a given scale (e.g. the set containing, for each 8-year old x, the set of degrees of tallness x reaches) and returns the so-called neutral segment on that scale (the interval of degrees that make an 8-year old neither tall nor short plus the maximal degree lower than that interval and plus the minimum degree higher than that interval). This is depicted in (26).

(23) $[\![-er]\!] = \lambda Q_{<d,t>}.\lambda P_{<d,t>}.Q \subset P$

(24) $[\![-est]\!] = \lambda \boldsymbol{Q}_{<dt,t>}.\lambda P_{<d,t>}.\forall Q \in \boldsymbol{Q}[Q \neq P \rightarrow Q \subset P]$

(25) $[\![POS]\!] = \lambda \boldsymbol{Q}_{<dt,t>}.\lambda P_{<d,t>}.L_{<<dt,t>,<dt>>}(\boldsymbol{Q}) \subseteq P]$

(26) |- - - - - - - - - - - - - - - - - -[////////]- - - - - - - - - - - - - - - - ∞

Once the λQ-argument has been filled up (by the denotation of overt material or by a context-dependent variable C), we have a generalized quantifier over degrees, which must move out of its original host to gain scope, as in (27). The resulting truth conditions are in (28)-(30):[3]

(27) LF: [[-er/-est/POS C] 1 [Lucia is t_1-tall]]

(28) a. (Greta is 1,26m). Lucía is taller (than that).

 b. $\lambda d.\text{tall}(\text{greta}, d) \subset \lambda d.\text{tall}(\text{lucia}, d)$

(29) a. Lucía is tallest (among the girls in her class).

 b. $\forall Q \in \{\lambda d.\text{tall}(\text{greta}, d), \lambda d.\text{tall}(\text{sarah}, d), \lambda d.\text{tall}(\text{lucia}, d), \lambda d.\text{tall}(\text{liv}, d), \ldots\}$
   $[Q \neq \lambda d.\text{tall}(\text{lucia}, d) \rightarrow Q \subset \lambda d.\text{tall}(\text{lucia}, d)]$

(30) a. Lucía is tall (for an 8-year old).

 b. $L(\{\lambda d.\text{tall}(\text{valentin}, d), \lambda d.\text{tall}(\text{jonah}, d), \lambda d.\text{tall}(\text{lucia}, d), \ldots\}) \subseteq \lambda d.\text{tall}(\text{lucia}, d)$

It has been noted that the superlative morpheme *-est* with adjectives allows for an absolute and a relative reading, as in (31), and that the exact relative reading depends (at least partly) on the information structure of the sentence, as illustrated in (32) (Heim, 1999; Szabolcsi, 1986). Here we are interested in the relative reading. As sketched in (33), under this reading *-est* scopes out of its NP host and the comparison class C is retrieved (partly) from the focus value of the LF sister of *[-est C]* via the squiggle operator (Heim, 1999):

(31) John climbed the highest mountain.

 a. Absolute: "John climbed a mountain higher than any other (relevant) mountain".

 b. Relative: "John climbed a higher mountain than anybody else (relevant) climbed".

(32) a. John wrote the longest letter to Mary$_{\mathbf{F}}$. $\mapsto$ compares *recipients* of John's letters

 b. John$_{\mathbf{F}}$ wrote the longest letter to Mary. $\mapsto$ compares *senders* of letters to Mary

---

[3]For simplicity we treat degree operators extensionality. The intensional treatment is illustrated in (i).

(i) $[\![-est]\!] = \lambda \boldsymbol{Q}_{<<s,dt>,t>}.\lambda P_{<s,dt>}.\lambda w.\forall Q \in \boldsymbol{Q}[Q \neq P \rightarrow Q(w) \subset P(w)]$

(33) Relative reading of *-est*:

    a. LF: [[**-est C**][1[John$_\mathbf{F}$ climbed A t_1-high mountain]] $\sim$ C]

    b. $[\![1[John\ climbed\ a\ t_1\text{-}high\ mountain]]\!] = \lambda d'$.John climbed a d'-high mountain

    c. $[\![C]\!] \subseteq \{\lambda d'.$ John climbed a d'-high mountain, Bill climbed a d'-high mountain, $\lambda d'.$ Paul climbed a d'-high mountain, $\ldots\}$

    d. $[\![(31)]\!] = 1$ iff $\forall Q \in [\![C]\!]$ $[Q \neq \lambda d.\exists x[\text{climb}(j,x) \wedge \text{mount}(x) \wedge \text{high}(x,d)] \rightarrow Q \subset \lambda d.\exists x[\text{climb}(j,x) \wedge \text{mount}(x) \wedge \text{high}(x,d)]]$

A parallel absolute/relative ambiguity has been detected for *POS* with adjectives, where, again, the exact relative reading depends on what element *POS* associates with (Schwarz, 2010). This is shown in (34) and (35). Schwarz (2010) extends Heim's (1999) analysis of *-est* to *POS*. Again, here we are interested in the relative reading, adapted from Schwarz (2010) in (36):[4]

(34) Mia has an expensive hat.

    a. Absolute: 'Mia has a hat that is expensive for a hat'

    b. Relative: 'Mia has a hat that is expensive for somebody like Mia to have (e.g., for a 3-year old)'.

(35) Paul gave Mia an expensive hat.
    $\mapsto$ a hat that is expensive for somebody like Paul (e.g. unemployed people) to give
    $\mapsto$ a hat that is expensive for somebody like Mia (e.g. a 3-year old) to get

(36) Relative reading of *POS*:

    a. LF: [[**POS C**] [1 [Mia$_{\mathbf{F/CT}}$ has a t_1-expensive hat]] $\sim$ C]

    b. $[\![C]\!] \subseteq \{\lambda d'.$ Mia has a d'-expensive hat, $\lambda d'.$ Katie has a d'-expensive hat, $\ldots\}$

    c. $[\![(34)]\!] = 1$ iff $L([\![C]\!]) \subseteq \lambda d.\exists x[\text{have}(m,x) \wedge \text{hat}(x) \wedge \text{expensive}(x,d)]$

3.2 A novel observation on *POS* with adjectives

In the relative readings in (35) above, the associate of *POS* (namely, *Mia* or *Paul*) is external to the original host NP *[an expensive hat]*. We note that the associate may be internal to the host NP as well: Sentence (37) has a reading that makes it true in scenario (38) for the comparison class (39). This comparison class corresponds to having *car* as the associate of *POS*.

(37) (For what he has been giving her, now) Rockefeller gave Kate an inexpensive car$_{\mathbf{F/CT}}$.

(38) Scenario: Rockefeller just gave Kate a very expensive car. Still, this present compares poorly to his previous astronomically expensive presents (e.g. apartment in Manhattan, island in Pacific, etc.)

(39) $[\![C]\!] \subseteq \{\lambda d'.R$ gave K a d'-inexpensive car, $\lambda d'.R$ gave K a d'-inexpensive apartment in Manhattan, $\lambda d'.R$ gave K a d'-inexpensive island in the Pacific, $\ldots\}$

In sum, we have seen that adjectives decompose into STEM+*-er*/*-est*/*POS* and we have witnessed how *POS* operates in the relevant readings: it scopes out of its host NP to gain sentential scope and it retrieves its comparison class *C* (partly) from the LF sister of *[POS C]* by cycling in different alternatives to *POS*' associate. Furthermore, this associate may be

---

[4]The use of focus/topic alternatives is not from Schwarz (2010). Schwarz uses a 3-place lexical entry for *POS* and thus does not need to generate alternatives from the information structure of the sentence. We assume the 2-place entry and need to generate alternatives somehow. To this end, we will assume that the associate of *POS* (e.g. *Paul* or *Mia* in (35)) bears focal stress or functions as contrastive topic.

external or internal to the original host NP. A similar decomposition has been proposed for certain determiners: *more* decomposes as MANY+*-er* (Hackl, 2000) and *most* as MANY+*-est* (Hackl, 2009). In the next section, we will propose that *many* decomposes as MANY+POS and we will parsimoniously assume that the behaviour of POS witnessed in adjectives is paralleled in determiners.

4 Proposal

The ingredients of the proposal are the following:

i. *Many* is decomposed into the parametrized determiner MANY and the degree operator POS. Similarly, *few* is decomposed into the parametrized determiner FEW and POS.[5]

ii. There is only one proportional determiner MANY$_{prop}$ and only one proportional determiner FEW$_{prop}$, both of which are conservative.

iii. Just as we saw with adjectives, POS in determiners *many* and *few* scopes sententially and retrieves a comparison class C from its syntactic scope based on its F-/CT-associate. The exact reading obtained depends on the associate.

In the following, we show that the regular proportional reading arises when POS' associate is external to the NP host and the reverse proportional reading obtains when the associate is internal to the NP host.

4.1 Proportional readings of *many*

Once we severe POS from *many*, we are left with two determiner morphemes MANY: the cardinal one in (40) and the proportional one in (41)[6]. Since we are interested in the proportional reading in this talk, we will concentrate on (41).

(40) $[\![\text{MANY}_{\text{card}}]\!] = \lambda d_d.\lambda P_{<e,t>}.\lambda Q_{<e,t>}.|P \cap Q| \geq d$

(41) $[\![\text{MANY}_{\text{prop}}]\!] = \lambda d_d.\lambda P_{<e,t>}.\lambda Q_{<e,t>}.(|P \cap Q| : |P|) \geq d$

When we use MANY$_{\text{prop}}$ and POS is associated with an element <u>external</u> to the host NP, the regular proportional reading arises:

(42)   Many (of the few) faculty children had a good$_{\textbf{F/CT}}$ time.

(43)   (Regular) proportional reading:

     a. LF: [[**POS C**] [1[[t_1-MANY$_{\text{prop}}$ faculty children] has a good$_{\textbf{F/CT}}$ time]] $\sim$ C]

     b. $[\![C]\!] \subseteq \{\lambda d'.(|\{x : \text{fac-child}(x)\} \cap \{x : \text{have-good-time}(x)\}| : |\{x : \text{fac-child}(x)\}|) \geq d',$
                $\lambda d'.(|\{x : \text{fac-child}(x)\} \cap \{x : \text{have-bad-time}(x)\}| : |\{x : \text{fac-child}(x)\}|) \geq d',$
                $\lambda d'.(|\{x : \text{fac-child}(x)\} \cap \{x : \text{have-regular-time}(x)\}| : |\{x : \text{fac-child}(x)\}|) \geq d', \ldots\}$

     c. $L([\![C]\!]) \subseteq \lambda d.(|\{x : \text{fac-child}(x)\} \cap \{x : \text{have-good-time}(x)\}| : |\{x : \text{fac-child}(x)\}|) \geq d$

---

[5]The parametrized determiner FEW will be further decomposed into LITTLE+MANY in Section 5 (Heim, 2006). For the examples in the present section, the simpler decomposition suffices.

[6]Adjectival uses of cardinal *many/few*, as exemplified in (i), suggest that MANY$_{\text{card}}$ / FEW$_{\text{card}}$ may be adjectives rather than a determiners. See also Hackl (2009)'s analysis of the absolute reading of *most* based on an adjectival version of MANY$_{\text{card}}$. We leave a potential extension in this direction for future research.

(i)   The many/few students of the University of Konstanz protested.

When we use $\text{MANY}_{\text{prop}}$ but *POS* is associated with an element <u>internal</u> to the host NP, we obtain the reverse proportional reading. The truth conditions derived in (45b)-(45c) correspond precisely to the characterization of the reverse proportional reading argued for in Section 2.

(44)   Many Scandinavians$_{\mathbf{F/CT}}$ have won the Nobel Prize in literature.

(45)   Reverse proportional reading:

     a.   LF: [[**POS C**] [1[[t_1-$\text{MANY}_{\text{prop}}$ Scandinavians$_{\mathbf{F/CT}}$] have won NP]] $\sim$ C]

     b. $[\![C]\!] \subseteq \{\lambda d'.(|\{x : \text{Scandinavian}(x)\} \cap \{x : \text{NP-winner}(x)\}| : |\{x : \text{Scandinavian}(x)\}|) \geq d'$,
           $\lambda d'.(|\{x : \text{Mediterranean}(x)\} \cap \{x : \text{NP-winner}(x)\}| : |\{x : \text{Mediterr.}(x)\}|) \geq d'$,
           $\lambda d'.(|\{x : \text{M.Eastern}(x)\} \cap \{x : \text{NP-winner}(x)\}| : |\{x : \text{M.Eastern}(x)\}|) \geq d', \dots\}$

     c. $L([\![C]\!]) \subseteq \lambda d.(|\{x : \text{Scandinavian}(x)\} \cap \{x : \text{NP-winner}(x)\}| : |\{x : \text{Scandinavian}(x)\}|) \geq d$

4.2   Proportional readings of *few*

Once we separate *POS* from *few*, we are left with the parametrized determiners FEW_{card} and FEW_{prop} below. Again, we will concentrate on the latter:

(46) $[\![\text{FEW}_{\text{card}}]\!] = \lambda d_d.\lambda P_{<e,t>}.\lambda Q_{<e,t>}.|P \cap Q| < d$ (TO BE REVISED)

(47) $[\![\text{FEW}_{\text{prop}}]\!] = \lambda d_d.\lambda P_{<e,t>}.\lambda Q_{<e,t>}.(|P \cap Q| : |P|) < d$ (TO BE REVISED)

When we use FEW_{prop} and *POS* is associated with an element in the sentence <u>external</u> to the host NP, the regular proportional reading obtains:

(48)   Few (of the many) demonstrators had a good$_{\mathbf{F/CT}}$ time.

(49)   (Regular) proportional reading:

     a.   LF: [[**POS C**] [1[[t_1-FEW_{prop} demonstrators] has a good$_{\mathbf{F/CT}}$ time]] $\sim$ C]

     b. $[\![C]\!] \subseteq \{\lambda d'.(|\{x : \text{demonstr}(x)\} \cap \{x : \text{have-good-time}(x)\}| : |\{x : \text{demonstr}(x)\}|) < d'$,
           $\lambda d'.(|\{x : \text{demonstr}(x)\} \cap \{x : \text{have-bad-time}(x)\}| : |\{x : \text{demonstr}(x)\}|) < d'$,
           $\lambda d'.(|\{x : \text{demonstr}(x)\} \cap \{x : \text{have-regular-time}(x)\}| : |\{x : \text{demonstr}(x)\}|) < d'$,
           $\dots\}$

     c. $L([\![C]\!]) \subseteq \lambda d.(|\{x : \text{demonstr}(x)\} \cap \{x : \text{have-good-time}(x)\}| : |\{x : \text{demonstr}(x)\}|) < d$

When we use FEW_{prop} but *POS* is associated with an element in the sentence <u>internal</u> to the host NP, the reverse proportional reading results, with the truth conditions we argued for:

(50)   Few cooks$_{\mathbf{F/CT}}$ applied.

(51)   Reverse proportional reading:

     a.   LF: [[**POS C**] [1[[t_1-FEW_{prop} cooks$_{\mathbf{F/CT}}$] applied]] $\sim$ C]

     b. $[\![C]\!] \subseteq \{\lambda d'.(|\{x : \text{cooks}(x)\} \cap \{x : \text{apply}(x)\}| : |\{x : \text{cooks}(x)\}|) < d'$,
           $\lambda d'.(|\{x : \text{someliers}(x)\} \cap \{x : \text{apply}(x)\}| : |\{x : \text{someliers}(x)\}|) < d'$,
           $\lambda d'.(|\{x : \text{waiters}(x)\} \cap \{x : \text{apply}(x)\}| : |\{x : \text{waiters}(x)\}|) < d', \dots\}$

     c. $L([\![C]\!]) \subseteq \lambda d.(|\{x : \text{cooks}(x)\} \cap \{x : \text{apply}(x)\}| : |\{x : \text{cooks}(x)\}|) < d$

In sum, we have proposed a compositional analysis that derives the correct truth conditions for the reverse proportional reading of *many* and *few*, and this has been achieved using only conservative determiners —namely, $\text{MANY}_{\text{prop}}$ and FEW_{prop}— and exploiting independently motivated properties of *POS*.[7]

---

[7] When cardinal $\text{MANY}_{\text{card}}$ and FEW_{card} are used, different readings are derived too depending on whether the associate is external or internal to the host NP. See Romero (2015) for details. The reason why no attention has been drawn towards the reverse cardinal reading is that it does not give the appearance of non-conservativity.

5 Further predictions

We have argued that *many* and *few* decompose into MANY+*POS* and FEW+*POS* respectively and that *POS* scopes out of its host NP to gain sentential scope. If the pair MANY/FEW behaves like other degree antonym pairs, e.g. *tall/short*, we expect FEW to further decompose into MANY plus the negative element LITTLE (cf. Heim, 2006), defined in (52):

(52) $[\![\text{LITTLE}]\!]$ $= \lambda d_d.\lambda D_{<d,t>}.D(d) = 0$

Furthermore, if *POS* behaves like other degree operators, we expect it to be able to take scope in its own clause or in a higher clause in the relevant configurations. Such an ambiguity is found e.g. for the comparative morpheme *-er* in example (53): *required* scoping over *-er* produces reading (54) and *-er* scoping over *required* produces reading (55) (Heim, 2006):

(53)   (This draft is 15 pages.) The paper is required to be less long than that.
    a. 'Being under 15 pages is a necessity.'
    b. 'Being under 15 pages is a possibility.'

(54)   a. LF: [require [$_{\text{DegP}}$ **-er** than 15pp] 1[[$_{\text{DegP}}$ t_1 little] 2 paper is long t_2]]
    b. $\lambda w. \forall w' \in \text{Acc}(w)$ [{15pp} $\subset \lambda d.\neg\text{long}(\text{paper}, d, w')$]

(55)   a. LF: [[$_{\text{DegP}}$ **-er** than 15pp] 1[[$_{\text{DegP}}$ t_1 little] 2 required paper is long t_2]]
    b. $\lambda w. \{15\text{pp}\} \subset \lambda d.\exists w' \in \text{Acc}(w)[\neg\text{long}(\text{paper}, d, w')]$

Hence, if the decomposition analysis of *many* and *few* is on the right track, the proposed analysis predicts more LFs to be possible than an analysis where all the meaning components are fused together and thus must scope together, e.g. an analysis where the truth conditions in (56) are packed in an undecomposable entry for *few*. The examples of FEW$_{\text{card}}$ below suggest that the additional power of the decomposition analysis is needed: while (57)-(58) lead to a reading where *POS* and LITTLE scope under *required*, as shown in (59), (60)-(61) prompt a reading where *POS* and LITTLE scope over *required*, as in (62).

(56)   REVERSE PROPORTIONAL reading of *Few Ps are Q*:
               $|P \cap Q|{:}|P|$ $<$ $\theta(\{|P' \cap Q|{:}|P'| : P' \in \text{ALT}(P)\})$

(57)   Scenario: Our grad students are stressed out. According Dr. Smith's prescription, the amount of reading they do should be lower than that of a regular grad student (whatever that is). That is, our grad students reading few papers (for a grad student) is a necessity.

(58)   Smith requires our grad students$_{\textbf{F/CT}}$ to read few papers (for a grad student).

(59)   a. LF: [S. requires [[**POS** C] 1 [[$_{\text{DegP}}$ t_1 LITTLE] 2 [the stud.$_{\textbf{F/CT}}$ to read t_2-MANY papers]]]]
    b. $\lambda w. \forall w' \in \text{Acc}(w)[L([\![C]\!]) \subseteq \lambda d.|\{x : \text{papers}(x, w')\} \cap \{x : \text{read}(\text{john}, x, w')\}| < d]$

(60)   Scenario: For all full professors except for Prof. Smith, the minimum requirement in their courses is for our grad students to read (any) 30 papers. For Prof. Smith, the minimum requirement in his courses is for our grad students to read (any) 10 papers. That is, in Prof. Smith's courses, our grad students reading few papers is a possibility.

(61)   (For how much full professors tend to require from grad students in their courses,) Smith$_{\textbf{F/CT}}$ requires our grad students to read few papers.

(62)   a. LF: [[**POS** C] 1 [[$_{\text{DegP}}$ t_1 LITTLE] 2 [Smith$_{\textbf{F/CT}}$ requires [the students to read t_2-MANY papers]]]]
    b. $\lambda w. L([\![C]\!]) \subseteq \lambda d.\exists w' \in \text{Acc}(w)[|\{x : \text{papers}(x, w')\} \cap \{x : \text{read}(\text{stud.}, x, w')\}| < d]$

6   Conclusions

By decomposing *many* into the positive degree operator *POS* and the parametrized determiner MANY, the so-called reverse proportional reading has been derived while appealing solely to independently motivated behavior of *POS* and while keeping a single, conservative lexical entry for MANY$_{\text{prop}}$. The same holds for *few*. Importantly, contrary to the analyses by Westerståhl (1985) and Cohen (2001), the proposed analysis derives the correct truth conditions for this reading. Furthermore, the proposed decomposition correctly predicts the existence of further scopal readings that are unexpected in non-decomposition analyses.

References

Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 5. 159–219.

Beaver, David & Brady Clark. 2008. *Sense and Sensitivity: How Focus Determines Meaning.* Oxford: Blackwell.

Cohen, Ariel. 2001. Relative readings of many, often and generics. *Natural Language Semantics* 69. 41–67.

van der Does, Jaap & Jan van Eicjk. 1996. Basic quantifier theory. In J. van der Does & J. van Eijck (eds.), *Quantifiers, Logic, and Language*, 1–45. Stanford, CA: CLSI Publications.

Fernando, Tim & Hans Kamp. 1996. Expecting many. In T. Galloway & Justin Spence (eds.), *Proceedings of SALT 6*, 53–68. Cornell University. Ithaca, NY: CLC Publications.

Hackl, Martin. 2000. *Comparative Quantifiers*: MIT dissertation.

Hackl, Martin. 2009. On the grammar and processing of proportional quantifiers: *Most* versus *More Than Half*. *Natural Language Semantics* 17. 63–98.

Heim, Irene. 1999. Notes on Superlatives. MIT lecture notes.

Heim, Irene. 2006. Little. In M. Gibson & J. Howell (eds.), *Proceedings of SALT 16*, 35–58. Cornell University. Ithaca, NY: CLC Publications.

Herburger, Elena. 1997. Focus and weak noun phrases. *Natural Language Semantics* 5. 53–78.

de Hoop, Helen & Jaume Solà. 1996. Determiners, context sets, and focus. In *Proceedings of the Fourteenth West Coast Conference on Formal Linguistics*, 155–167. Stanford, CA: CSLI Publications.

Keenan, Ed L. & Jonathan Stavi. 1986. A semantic characterization of natural language determiners. *Linguistics and Philosophy* 9. 253–326.

Partee, Barbara H. 1988. Many quantifiers. In J. Powers & K. de Jong (eds.), *Proceedings of the Fifth Eastern States Conference on Linguistics*, 383–402. Columbus: The Ohio State University.

Romero, Maribel. 2015. Pos and the inverse proportional reading of *many*. Talk at *NELS* 46.

Schöller, Anthea & Michael Franke. 2015. Semantic vales as latent parameters: surprising *few* and *many*. Talk at *SALT* 25.

Schwarz, Bernhard. 2010. A note on *for*-phrases and derived scales. Handout for talk at *Sinn und Bedeutung* 15.

von Stechow, Armin. 2009. The temporal degree adjectives *früher/später* 'early(er)'/'late(r)' and the semantics of the positive. In A. Giannakidou & M. Rathert (eds.), *Quantification, Definiteness and Nominalization*, 214–233. Oxford University Press.

Szabolcsi, Anna. 1986. Comparative superlatives. In N. Fukui, T. Rapoport & E. Sagey (eds.), *Papers in Theoretical Linguistics*, 245–265. Cambridge, MA: MITWPL 8.

Westerståhl, Dag. 1985. Logical constants in quantifier languages. *Linguistics and Philosophy* 8. 387–413.

Counterpossibles

Timothy Williamson

University of Oxford, Oxford, U.K

`timothy.williamson@philosophy.ox.ac.uk`

Abstract

According to orthodoxy, all counterpossibles (counterfactual conditionals with impossible antecedents) are true, or at least not false. Nevertheless, some counterpossibles look false. The problem is not just how best to tidy up an unimportant corner of the logic and semantics of counterfactuals. It has significant theoretical and methodological ramifications in several directions. This paper defends the orthodox view against some recent objections, and explains the most recalcitrant appearances to the contrary by our pre-reflective reliance on a fallible heuristic in the assessment of counterfactuals.

1   What is at Stake

Typically, we use counterfactuals to talk about what would have happened if something had been different from how it actually was. Still, despite the etymology, a counterfactual may have a true antecedent; 'If she were depressed, that would explain her silence' does not imply that she is not depressed. But a counterpossible is a counterfactual whose antecedent is impossible, and therefore false.

What kind of impossibility is relevant? It is not epistemic. For consider this counterfactual:

(1).   If thinking had never occurred, science would have flourished.

The antecedent of (1) is epistemically impossible, because it is incompatible with something we know: that we think. But that does not make the antecedent of (1) impossible in the relevant sense. Presumably, the universe could have been lifeless and thoughtless forever. In that case, science would *not* have flourished. Defenders of orthodoxy should agree that (1) is false. The special theoretical problem in evaluating 'If this had been so, that would have been so' arises when this *could not have been* so. The relevant sort of possibility is objective rather than epistemic or subjective. Moreover, what matters is the most inclusive sort of objective possibility, which we may call *metaphysical possibility*. For the special theoretical problem in evaluating a counterfactual does not arise when, although the antecedent could not *easily* have been so, it could still have been so.

It is convenient, though not crucial, to put the problem in terms of possible worlds. We take the worlds to be *possible* in the sense that it is metaphysically possible for any one of them to have obtained; metaphysical possibility as just explained is an appropriately inclusive standard for present purposes. The evaluation of the counterfactual $\alpha \:\square\!\!\rightarrow \beta$ depends on the truth-value of β at relevant possible worlds at which α is true. But what happens if β is true at *no* possible worlds?

We may equate the *intension* $|\alpha|$ of a sentence α (in a context C) with the set of possible worlds at which α is true (in C). In a compositional intensional semantics of the usual type for counterfactuals, the intension of a counterfactual is a function of the intensions of its antecedent and consequent:

(2). $|\alpha \:\square\!\!\rightarrow \beta| = f(|\alpha|, |\beta|)$

Indeed, we can surely be more specific, for all that should matter about the consequent is at which possible worlds *where the antecedent is true* the consequent is also true, in other words, the intersection of the intension of the antecedent with the intension of the consequent. If so, (2) implies (3):

(3). $|\alpha \:\square\!\!\rightarrow \beta| = f(|\alpha|, |\alpha| \bigcap |\beta|)$

The truth-value of the consequent at worlds where the antecedent is false should be irrelevant to the truth-value of the conditional, for it concerns only what hold if its antecedent held. But (3) yields (4):

(4).   If $|\alpha| = \{\}$ then $|\alpha \,\square\!\!\rightarrow \beta| = f(\{\}, \{\})$

Given (4), all counterpossibles have the same intension: they are indiscriminate. That is not yet to decide between making them all true and making them all false. However, we surely want any counterfactual (counterpossible or not) whose consequent merely repeats its antecedent to be a trivial necessary truth, true throughout the set of all possible worlds W:

(5). $|\alpha \,\square\!\!\rightarrow \alpha| = W$

Together, (4) (with $\beta = \alpha$) and (5) require $f(\{\}, \{\})$ to be W. Putting that back into (4), we get:

(6).   If $|\alpha| = \{\}$ then $|\alpha \,\square\!\!\rightarrow \beta| = W$

In other words, all counterpossibles are necessary truths, and so truths. This is just the conclusion reached by Stalnaker, Lewis, and their successors in the mainstream of intensional semantics.

Similar arguments can be made in the modal object-language, without reference to worlds. For example, instead of (2) we can just require that counterfactuals with necessarily equivalent antecedents and necessarily equivalent consequents are themselves necessarily equivalent:

(7). $(\square(\alpha \equiv \alpha^*) \,\&\, \square(\beta \equiv \beta^*)) \supset \square((\alpha \,\square\!\!\rightarrow \beta) \equiv (\alpha^* \,\square\!\!\rightarrow \beta^*))$

A plausible auxiliary assumption to complete the argument is simply that conjunctions necessarily counterfactually imply their conjuncts (if this and that were so, this would be so):

(8). $\square((\alpha \wedge \beta) \,\square\!\!\rightarrow \alpha)$

Together, (7) and (8) imply (9), the analogue in the object-language of (6):

(9). $\square\neg\alpha \supset \square(\alpha \,\square\!\!\rightarrow \beta)$

A much simpler argument in the object-language for the truth of counterpossibles just uses a plausible and attractive assumption linking metaphysical modality to counterfactuals. It is that strict implication materially implies counterfactual implication:

(10). $\square(\alpha \supset \beta) \supset (\alpha \,\square\!\!\rightarrow \beta)$

For, by elementary modal logic, an impossibility strictly implies anything:

(11). $\square\neg\alpha \supset \square(\alpha \supset \beta)$

By transitivity, (10) and (11) entail (12):

(12). $\square\neg\alpha \supset (\alpha \,\square\!\!\rightarrow \beta)$

If one assumes the necessitation of (10), one can also derive the necessitation of (12).

One can use (10) in deriving equivalents of metaphysical modalities in counterfactual terms, as part of an argument for understanding our cognitive capacities for handling metaphysical modalities as a by-product of our cognitive capacities for handling counterfactual conditionals. In such ways, issues about counterpossibles have significant knock-on effects for more general philosophical questions.

Thus strong theoretical pressures push towards orthodoxy about counterpossibles. It is required by the standard simple and natural approach to the semantics of counterfactuals, and it contributes to a simple and natural picture of how counterfactuals and metaphysical modality fit together. Nevertheless, those pressures are not obviously irresistible. If one is willing to countenance impossible worlds in addition to possible worlds, one might be able to retain the world-based semantic framework for counterfactuals while rejecting orthodoxy about counterpossibles, as Nolan, Brogaard and Salerno, and Kment do. For if the semantic value of a counterfactual is sensitive to its behaviour at impossible worlds, that makes it easier to deny assumptions such as (2), (7), and (10). Alternatively, one might seek a different semantic framework for counterfactuals that is less congenial to orthodoxy about counterpossibles.

The focus of resistance to orthodoxy about counterpossibles is usually on alleged counterexamples. Here is one (from Nolan). What are the truth-values of (13) and (14)?

(13).   If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared.

(14).   If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would not have cared.

If one responds in a theoretically unreflective way, the natural snap answers are presumably that (13) is false and (14) true. The sick children in South America were in no position to know about Hobbes's secret reasoning thousands of miles away; even if they had known, they had more urgent things to care

about. But, as we know, the shared antecedent of (13) and (14) is metaphysically impossible. Therefore, according to orthodoxy, (13) and (14) alike are true. According to the critics, (13) is a counterexample to orthodoxy: a false counterpossible. Such examples can be multiplied.

The temptation to deny (13) and similar counterpossibles is strong. But such inclinations are not always veridical. For the time being, we may treat them as defeasible evidence against orthodoxy.

For some metaphysicians, rejecting orthodoxy also has more theoretical attractions. Here is an example. Nominalists crave the scientific advantages that platonists gain from quantifying over numbers and other abstract objects. How to emulate them? A common strategy, in this and similar cases, is *fictionalist*. One treats the envied rival metaphysical theory as a useful fiction. The proposal deserves to be taken seriously only if accompanied by a properly worked-out account of how reasoning on the basis of a fiction can nevertheless be a reliably truth-preserving way of getting from non-fictional premises to a non-fictional conclusion. For instance, if one reasons validly from true premises purely about concrete reality plus a false (by nominalist lights) auxiliary mathematical theory about abstract objects to a conclusion purely about concrete reality, the conclusion needs to be true too. But why should it be true? One way of implementing the fictionalist strategy is to use counterfactuals. The nominalist reasons in effect about *how things would be if the mathematical theory were to obtain and concrete reality were just as it actually is*. Thus the conclusion corresponds to this counterfactual:

(15). $(M \wedge A) \,\square\!\!\rightarrow C$

Here M is the platonist mathematical theory, A says that concrete reality is just as it actually is, and C says something specific purely about concrete reality. Thus, the truth of the counterfactual seems to guarantee the truth of its consequent, even though its antecedent is false (by nominalist lights), because the relevant counterfactual worlds are the same as the actual world with respect to concrete reality, which C is purely about. The trouble is that the nominalist may well regard platonism as not just *false* but *metaphysically impossible*: for instance, the structure of the hierarchy of pure sets (if any) seems to be a metaphysically non-contingent matter. For such a nominalist, M is impossible, so the counterfactual (15) is a counterpossible. But, given orthodoxy about counterpossibles, the impossibility of the antecedent guarantees the truth of the counterpossible, irrespective of its consequent, so the mere truth of (15) is insufficient for the truth of C. Fictionalists who implement their strategy by means of counterfactuals and regard the rival metaphysical theory as a useful but impossible fiction have therefore been compelled to deny orthodoxy about counterpossibles.

We can be a little more explicit about the relation between the move from (15) to C, on one hand, and orthodoxy about counterpossibles, on the other. The natural route from (15) to C is this. Suppose that C is (actually) false. By hypothesis, A says that concrete reality is just as it actually is, and C says something specific purely about concrete reality, hence $\neg C$ does too. Therefore, we can treat $\neg C$ as part of what A in effect says. Thus the opposite counterfactual surely holds:

(16). $(M \wedge A) \,\square\!\!\rightarrow \neg C$

If (16) excludes (15) we can therefore derive $\neg$(15) from $\neg C$, and so C from (15) by contraposition. Conversely, we can derive (15) from C just as we derived (16) from $\neg C$ (without relying on the mutual exclusion of counterfactuals). But orthodoxy rejects the assumption that (15) and (16) exclude each other, for both are true if their shared antecedent is impossible.

Most orthodox theorists will hold that opposite counterfactuals such as (15) and (16) are compatible *only* if they are counterpossibles. For they are likely to accept the following two principles in the logic of counterfactuals. First, counterfactuals distribute over conjunction in the consequent:

(17). $(\alpha \,\square\!\!\rightarrow (\beta \wedge \gamma)) \equiv ((\alpha \,\square\!\!\rightarrow \beta) \wedge (\alpha \,\square\!\!\rightarrow \gamma))$

This holds on any standard semantics for counterfactuals. Second, no metaphysical possibility counterfactually implies a metaphysical impossibility:

(18). $(\alpha \,\square\!\!\rightarrow \beta) \supset (\Diamond\alpha \supset \Diamond\beta)$

If something is impossible which would obtain if something else were to obtain, then the other thing is impossible too. From (17) and (18) we can easily derive that the conjunction of opposite counterfactuals implies the impossibility of their antecedent:

(19). $((\alpha \,\square\!\!\rightarrow \beta) \wedge (\alpha \,\square\!\!\rightarrow \neg\beta)))) \supset \neg\Diamond\alpha$

Even opponents of orthodoxy about counterpossibles may grant (19), since their unorthodoxy may be confined to counterpossibles, and a counterexample to (19) would require a possible antecedent. What opponents of orthodoxy reject, and proponents accept, is the converse of (19).

2   Misconceptions about orthodoxy

In a recent critique of orthodoxy, Berit Brogaard and Joe Salerno characterize their target thus: "Counterpossibles are trivial on the standard account. By 'trivial', we mean *vacuously true and semantically uninformative*. Counterpossibles are *vacuously true* in that they are always true; an impossibility counterfactually implies anything you like. And relatedly, they are *uninformative* in the sense that the consequent of a counterpossible makes no contribution to the truth-value, meaning or our understanding of the whole." Of this conjunction, orthodoxy as characterized above corresponds only to the first conjunct, the claim of vacuous truth. Brogaard and Salerno handle even that conjunct somewhat oddly. For instance, they say of the counterpossible (14) above: 'The intuition is that [(14)] is true, but non-vacuously'. By their own definition, the non-vacuous truth of (14) consists only in its truth, which both sides acknowledge, and the falsity of at least one other counterpossible, such as (13). Thus the relevant 'intuition' is not directed at (14) at all, but at some other counterpossible. However, that is a minor point compared to their inclusion of the second conjunct, semantic uninformativeness. For we need to be quite clear that semantic uninformativeness is no part whatsoever of the standard account. Consequently, that counterpossibles are trivial in Brogaard and Salerno's sense is no part whatsoever of the standard account.

To see this, we must recall that the standard view of counterfactuals is that, as usual for complex sentences, they have a compositional semantics. Their meanings are built up out of the meanings of their constituents. On the standard view, as found in authors such as Stalnaker and Lewis, the meaning of the counterfactual $\alpha \,\square\!\!\rightarrow \beta$ is built up out of the meanings of the sentences α and β combined with the meaning of the counterfactual operator $\square\!\!\rightarrow$. Many linguists, notably Kratzer (2012), have a subtler view of the semantic structure of conditional sentences, but the points to be made here can be transposed to such alternative settings. On a fine-grained conception of meaning, any difference in meaning between the sentences β and γ makes a difference in meaning between the counterfactuals $\alpha \,\square\!\!\rightarrow \beta$ and $\alpha \,\square\!\!\rightarrow \gamma$, whatever the meaning of α. That applies just as much when α is impossible as when α is possible. For instance, the counterpossibles (13) and (14) differ by a 'not' in the consequent, which *ipso facto* makes a difference in meaning between (13) and (14). Thus it is just false that, on the standard view, 'the consequent of a counterpossible makes no contribution to the […] meaning […] of the whole'.

Equally objectionable is Brogaard and Salerno's claim that, on the standard view, 'the consequent of a counterpossible makes no contribution to […] our understanding of the whole'. For instance, consider these two counterfactuals:

(20).   If Plato had been identical with Socrates, Plato would have been snub-nosed.

(21).   If Plato had been identical with Socrates, 2 + 2 would have been 5.

We may assume that, by the necessity of distinctness, since Plato and Socrates are distinct, it is metaphysically impossible for Plato and Socrates to have been identical. Thus both (20) and (21) are counterpossibles. Nevertheless, we understand them by understanding their constituents and how they are put together. For instance, a failure to understand the constituent 'snub-nosed' prevents one from fully understanding (20), but does not prevent one from understanding (21). Thus, on the standard view, our understanding of the consequent of a counterpossible does contribute to our understanding of the whole counterpossible. Of course, if one happens to *know* that it is metaphysically impossible for Plato to have been identical with Socrates, and one accepts orthodoxy about counterpossibles, then one can work out that (20) is true even if one does not understand 'snub-nosed', but that is just an instance of the general point that one can know that a sentence states a truth without knowing what it states. For example, a trustworthy and trusted native speaker of Mandarin might utter a sentence of Mandarin and

tell me that it states a truth without telling me what truth it states. In any case, someone can understand (20) and (21) without knowing that it is impossible for Socrates to have been identical with Plato. Having spent too much time reading dodgy webpages, he might suspect that Plato *was* identical with Socrates. Alternatively, he might know that Plato was distinct from Socrates, but doubt the necessity of distinctness on faulty metaphysical grounds. In general, one can understand a counterpossible without knowing it to be a counterpossible, and one's understanding of it is relevantly like one's understanding of other counterfactuals. All these points arise naturally within the framework of a compositional approach to semantics, such as standard accounts assume.

What of the claim that, on the standard account, 'the consequent of a counterpossible makes no contribution to the truth-value […] of the whole'? At first sight, it looks more defensible, since truth-value is a more coarse-grained feature than either meaning or understanding. However, their claim about truth-values is unwarranted too. The only basis for making it is that, according to standard views, all counterfactuals with impossible antecedents have the same truth-value, because all are true, irrespective of their consequent. But, equally, according to standard views, all counterfactuals with *necessary consequents* have the same truth-value, because all are true, irrespective of their antecedent:

(22). $\quad \Box\beta \supset (\alpha \Box\!\!\rightarrow \beta)$

Both principles, (12) and (22), are corollaries of the quite general entailment (10) from any strict implication to the corresponding counterfactual; one can also derive the semantic analogue of (22) from (3) and (5). Thus, if the standard account implies that the consequent of a counterfactual with an impossible antecedent makes no contribution to the truth-value of the whole, by parity the standard account also implies that the antecedent of a counterfactual with a necessary consequent makes no contribution to the truth-value of the whole. But that combination is absurd. For consider a counterfactual such as (23) with an impossible antecedent *and* a necessary consequent:

(23). $\quad$ If 6 were prime, 35 would be composite.

By Brogaard and Salerno's style of reasoning, the standard account would imply that *neither* the antecedent *nor* the consequent of (23) makes any contribution to the truth-value of (23). That is absurd because, without its antecedent and consequent, all that is left of (23) is the bare counterfactual construction alone, which by itself certainly does not determine a truth-value. Obviously, standard theories of counterfactuals such as Stalnaker's and Lewis's have no such ridiculous consequence. Thus even Brogaard and Salerno's claim that, on the standard account, the consequent of a counterpossible makes no contribution to the truth-value of the whole is unwarranted. Such examples also tell in a parallel way against their already rejected claims that, on the standard account, the consequent of a counterpossible makes no contribution to the meaning and our understanding of the whole.

It is thus a misunderstanding of orthodoxy to suppose that it makes counterpossibles semantically uninformative or cognitively trivial. It simply makes them true. The misunderstanding is the source of many objections to orthodoxy. One such style of objection is this. According to orthodoxy, (24) is true, because Fermat's Last Theorem is a necessary truth:

(24). $\quad$ If Fermat's Last Theorem were false, 2 + 2 would be 5.

The critic then points out, correctly, that Andrew Wiles could not have simplified his famous proof by merely invoking (24) and thence deducing Fermat's Last Theorem by *reductio ad absurdum*. This does indeed refute the claim that (24) is uninformative or trivial, for given the latter claim it is harmless to rely on (24) in a proof. But it is hopeless as an argument against the claim that (24) is true, for the mere truth of a claim does not permit one to rely on it in a *proof*. For that, the claim must have some epistemically appropriate property: it must be an axiom, or have been already proved, or follow from previous steps in a way clear to expert mathematicians, or something like that. Since (24) has no such epistemically appropriate property, it offers no simplification of Wiles's proof. Thus the objection fails. More generally, assertibility requires some epistemically appropriate status, such as being known by the asserter, for which truth is insufficient. That point applies just as much to counterpossibles as to sentences of any other kind, and fits well with orthodoxy. Failure to appreciate it presumably comes from the confused idea that orthodoxy makes counterpossibles uninformative or trivial.

A subtler misconception about orthodoxy concerns speakers who know the impossibility of the antecedent. Consider (25):

(25).   If Hobbes had squared the circle, he would have become Lord Chancellor.

I know that the antecedent of (25) is impossible; given orthodoxy, I know that (25) is true. Epistemically, I am in a position to assert (25) on those grounds. But if I do so in a discussion of seventeenth century English politics, something is obviously amiss. However, that point does not tell against orthodoxy, for orthodoxy can easily explain what is amiss. Given orthodoxy, I was also in a position to assert the more informative and equally relevant (26) instead:

(26).   Hobbes could not have squared the circle.

Of course, (25) mentions political matters while (26) does not, but my grounds for asserting (25) make the mention factitious and misleading. Therefore, I should have asserted (26) — or, better, just kept quiet — instead, on Gricean grounds of conversational cooperation. Since I did not, my hearers may assume that I asserted (25) because I knew of some politically significant connection between squaring the circle and the Lord Chancellorship, and so be misled. If my hearers correctly identify my grounds for asserting (25), they will recognize the irrelevance of my contribution. Orthodoxy has no trouble in dealing with such cases.

In order to keep one's grip on the implications of orthodoxy, a salutary comparison is between the vacuous truth of counterpossibles and the vacuous truth of empty universal generalizations. The impossibility of the antecedent corresponds to the emptiness of the subject term. For it is widely agreed that 'Every N Vs' is true if and only if the extension of N is a subset of the extension of V. Thus, as a special case, if the extension of N is empty, it is a subset of the extension of V, whatever V is, so 'Every N Vs' is true. Consequently, since there are no golden mountains, 'Every golden mountain is a valley'. It would be obviously absurd to claim that, on this standard account of the universal quantifier, the predicates make no contribution to the truth-value, meaning or our understanding of such sentences. For universal generalizations have the same overall compositional semantic structure whether the subject term is empty or not, just as counterfactual conditionals have the same overall compositional semantic structure whether the antecedent is impossible or not. Similarly, it would be absurd to claim that, on the standard account, a sentence like 'Every golden mountain is in Africa' is cognitively trivial or semantically uninformative. One can understand it without knowing that there are no golden mountains.

Our reactions to counterpossibles are often similar to our reactions to analogous vacuous universal quantifications. In the latter case, we have learnt to override our immediate reactions. Perhaps we should learn to override our immediate reactions to counterpossibles in a similar way.

3 Counterfactual reasoning by reductio ad absurdum

The prime specimens of useful reasoning from an impossible supposition are arguments by *reductio ad absurdum* in mathematics. When we state them in everyday terms, it is natural to use counterfactual conditionals. Lewis gives these examples:

(27).   If there were a largest prime p, $p! + 1$ would be prime.
(28).   If there were a largest prime p, $p! + 1$ would be composite.

They summarize the classic proof that there is no largest prime: (27) holds because if p were the largest prime, $p!$ would be divisible by all primes (since it is divisible by all natural numbers up to p), so $p! + 1$ would be divisible by none; (28) holds because $p! + 1$ is larger than p, and so would be composite if p were the largest prime. To complete the proof, one can use Lewis's principle of Deduction within Conditionals to conjoin the consequents of (27) and (28):

(29).   If there were a largest prime p, $p! + 1$ would be both prime and composite.

Since the consequent of (29) is a contradiction, one can deny the antecedent, and conclude that there is no largest prime.

Of course, one does not strictly *need* to formulate the proof in terms of counterfactual conditionals. One could use material conditionals instead, because all standard mathematical reasoning can be

formalized in purely extensional terms. Nevertheless, it is surely legitimate, indeed natural and appropriate, to use counterfactual conditionals. They nicely convey the role of the antecedent in the reasoning. At the very least, on a good semantic theory, the counterpossibles (27)-(29) should come out *true*, for they are soundly based on valid mathematical reasoning.

Consider any non-obvious impossibility α that can be shown, by more or less elaborate deductive reasoning, to lead to an obvious impossibility ω. The general anti-orthodox strategy is to be charitable by evaluating counterfactuals with α as the antecedent at impossible worlds or situations not closed under such reasoning, precisely in order to falsify counterpossibles such as α $\square\!\!\rightarrow$ ω. But those are exactly the counterpossibles one needs to assert in articulating the argument by *reductio ad absurdum* against α. Thus the point generalizes, for instance to the use of counterlogical worlds.

Mathematical arguments by *reductio ad absurdum* are amongst the best arguments for counterpossibles we have. They tell us that if something non-obviously impossible were the case, something obviously impossible would be the case. We should accept the conclusions of those mathematical proofs. They provide strong evidence for orthodoxy. But how can we explain away the strongest evidence *against* orthodoxy, all the seemingly clear examples of false counterpossibles?

4   An error theory of apparently false counterpossibles

Processing a non-obvious counterpossible typically *feels* very like processing a non-counterpossible counterfactual. Consider (13) above, a good example of a seemingly false counterpossible ('If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared'). What goes on when we process it? In my case, before I consciously apply any theoretical considerations, it is something like this. I imagine Hobbes doing geometry in the secrecy of his room. I ask myself whether sick children in the mountains of South America at the time would have cared. I answer in the negative, because there was no way for them to have known about Hobbes's doings at the time, and even if they had known, they would hardly have cared. In the first instance, I assent to (14), the opposite counterfactual to (13), with the same antecedent but the negation of the consequent ('If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would *not* have cared'). My immediate inclination is then to deny (13), as excluded by (14). So far, the impossibility of the antecedent has played *no role whatsoever*. That is not to deny that I imagine Hobbes (secretly) squaring the circle. In some minimal, vague, unspecific way I do imagine him squaring the circle, but I could imagine him carrying out some genuine geometrical construction in much the same way. Now, in my case, theory kicks in. I remind myself that squaring the circle is impossible, and that opposite counterfactuals may both be true when their shared antecedent is impossible. I therefore countermand my inclination to deny (13).

What this suggests is that, in our unreflective assessment of counterfactual conditionals, we use a simple heuristic along these lines:

(HCC). Given that β is inconsistent with γ, treat α $\square\!\!\rightarrow$ β as inconsistent with α $\square\!\!\rightarrow$ γ

For instance, 'Sick children in the mountains of South America at the time cared' is obviously inconsistent with 'Sick children in the mountains of South America at the time did not care' (on the relevant readings), so in accordance with (HCC) we treat (13) as inconsistent with (14). Thus, having verified (14), we treat ourselves as having falsified (13). More generally, when drawing out the implications of a counterfactual supposition α, as soon as we have accepted α $\square\!\!\rightarrow$ γ, we take ourselves to be in a position to reject α $\square\!\!\rightarrow$ β for any β inconsistent with γ.

For many purposes, we can consider a simpler heuristic in place of (HCC):

(HCC*). If you accept one of α $\square\!\!\rightarrow$ β and α $\square\!\!\rightarrow$ ¬β, reject the other

(HCC*) has the advantage over (HCC) of not using 'inconsistent', a term which could do with some clarification.

(HCC) and (HCC*) are equivalent under a wide range of conditions. First, start with (HCC). Clearly, β is inconsistent with ¬β. Then (HCC) tells you to treat α $\square\!\!\rightarrow$ β as inconsistent with α $\square\!\!\rightarrow$ ¬β. Thus, if

you accept one of them, you should reject the other. In other words, you should obey (HCC*). Conversely, start with (HCC*). Suppose that you are given that β is inconsistent with γ. Thus γ entails ¬β. So, normally, from α □→ γ you can derive α □→ ¬β, by an informal analogue of Deduction within Conditionals. But (HCC*) tells you not to accept both α □→ β and α □→ ¬β. So, normally, you should not accept both α □→ β and α □→ γ. In other words, you should obey (HCC). However, since it is sometimes artificial to introduce an explicit negation when two sentences are obviously inconsistent, (HCC) may be the more natural heuristic.

There is psychological evidence that people reason in accordance with (HCC*), treating pairs of conditionals with the same antecedent and contradictory consequents as inconsistent, whether the conditionals are indicative or subjunctive. More generally, there is extensive psychological evidence that we tend to evaluate conditionals by evaluating their consequents on the supposition of their antecedents, with only subtle differences in treatment between indicatives and consequents. Thus we tend to treat cases where the antecedent is false as irrelevant to the evaluation. Our assessment of the probability of a conditional is highly correlated with our assessment of the conditional probability of its consequent on its antecedent. The well-supported suppositional model of our evaluation of conditionals predicts that it will conform to both (HCC) and (HCC*). For if β is inconsistent with γ, then it is inconsistent to accept both under the supposition of α. In probabilistic terms, the inconsistency of β with γ implies this relation between their conditional probabilities on α and the conditional probability of their disjunction on α:

(30).   Prob(β|α) + Prob(γ|α) = Prob(β ∨ γ|α) ≤ 1

Thus if Prob(β|α) is high, Prob(γ|α) is low, and *vice versa*. Given the close correlation between our assessments of conditional probabilities and our assessments of the probabilities of conditionals, this means that if we do not assess both conditionals in (HCC) or (HCC*) as probable.

For the orthodox, (HCC) and (HCC*) are only heuristics because they will lead you to reject true counterpossibles when α is impossible. However, it is plausible that usually, when counterfactual conditionals arise in practice, their antecedents are possible. In that case, (HCC*) will never lead you astray. For if you accept a true one of α □→ β and α □→ ¬β (which is not the responsibility of (HCC*)), (HCC*) will tell you to reject the other one, which will be false by (19). Given the near-equivalence of (HCC) and (HCC*), (HCC) will share much of the qualified reliability of (HCC*). Thus, on an orthodox logic of counterfactuals, both (HCC) and (HCC*) are reasonable though fallible heuristics.

One might wonder whether an unorthodox view of counterfactuals could treat (HCC) and (HCC*) as more than a fallible heuristic, by treating *no* counterpossibles as exceptions; that might even be an advantage of unorthodoxy. But that idea is unlikely to work. For consider counterpossibles with explicit contradictions as antecedents:

(31).   (α ∧ ¬α) □→ α   (32).   (α ∧ ¬α) □→ ¬α

Both (31) and (32) look highly plausible; surely conjunctions counterfactually imply their conjuncts. But if both (31) and (32) hold, then they are consistent, even though they have the same antecedent and inconsistent consequents. Now unorthodox theorists may reject some instances of (31) and (32), for instance when α itself is 'No conjunction counterfactually implies its conjuncts' or the like. But they cannot plausibly reject one of them in *all* cases. For instance, let α be 'The Liar is true', so α ∧ ¬α makes the dialetheist claim about the Liar paradox that the Liar is both true and not true. The dialetheist both asserts that the Liar is true *and* asserts that the Liar is not true. Presumably, therefore, both (31) and (32) should hold on this reading of α, even for the unorthodox. Thus even they should regard (HCC) and (HCC*) as fallible heuristics, not as marking exceptionless rules of the logic of counterfactuals.

Probabilistic considerations point in the same direction, even if we ignore the long series of results, initiated by David Lewis, which show that conditionals cannot in general be identified with propositions whose probability is the conditional probability of the antecedent on the consequent. Under the standard equation of the conditional probability Prob(β|α) with the ratio Prob(α ∧ β)/Prob(α) of unconditional probabilities, the conditional probability is undefined when Prob(α) is 0, as it is when α is a contradiction. If we treat conditional probabilities as primitive, we can sometimes assign Prob(β|α) a value even when Prob(α) = 0, but we still cannot do so when α holds at *no* point in the probability space, on pain of

violating basic principles of conditional probability. For Prob($\beta|\alpha$) should be 1 whenever α entails β, so for vacuous α Prob($\beta|\alpha$) should be 1 for every β, which violates the principle that Prob($\neg\beta|\alpha$) = 1 — Prob($\beta|\alpha$). Thus the structure of probability theory rules out vacuous conditional probabilities. Rejecting standard principles of conditional probability to allow for vacuous conditional probabilities would be a fool's bargain. Thus we should expect (HCC) and (HCC*) to have exceptions for at least some impossible antecedents.

How will the evaluation of $\alpha \;\square\!\!\rightarrow \beta$ and $\alpha \;\square\!\!\rightarrow \neg\beta$ go when α entails β and $\neg\beta$? Since both β and $\neg\beta$ would eventually emerge as we developed the counterfactual supposition α for long enough, (HCC) and (HCC*) make it a race between the contradictories as to which emerges first. If β emerges first, we accept $\alpha \;\square\!\!\rightarrow \beta$ and so reject $\alpha \;\square\!\!\rightarrow \neg\beta$ before $\neg\beta$ has time to emerge. If $\neg\beta$ emerges first, we accept $\alpha \;\square\!\!\rightarrow \neg\beta$ and so reject $\alpha \;\square\!\!\rightarrow \beta$ before β has time to emerge. On this view, the proponent of impossible worlds misinterprets this computational difference in terms of the relative closeness of impossible $\alpha \wedge \beta$ and impossible $\alpha \wedge \neg\beta$ worlds. One advantage of the heuristics account is that it explains our inattention to the impossibility of the antecedent in our cognitive processing of many counterpossibles. By contrast, accounts such as Brogaard and Salerno's that postulate a special standard of relative closeness for impossible worlds, apparently quite different from that appropriate for possible worlds, fail to explain the lack of felt adjustment to such a special standard in our cognitive processing of counterpossibles.

Of course, we are not completely helpless victims of our heuristics. Through conscious theoretical reflection, we can sometimes inhibit their operation. Our mastery of reasoning by *reductio ad absurdum* in mathematics shows our ability to defeat (HCC) and (HCC*). For example, we accept both the counterpossibles (29) and (30) in the proof that there is no largest prime, even though they have the same antecedent and mutually inconsistent consequents. Even in less formal settings, it is not psychologically compulsory to call off the search for β amongst the counterfactual consequences of α once $\neg\beta$ has turned up. If we are asked an open-ended question such as 'What would have been the consequences if α had been the case?', we can continue the search in a way that allows for mutually inconsistent counterfactual consequences to emerge. That is in effect what we do when asked 'Could α have been the case?'. Nevertheless, despite our ability to inhibit their operation, (HCC) and (HCC*) remain the default, to which we may always be liable to revert when off our guard. For instance, if one puts aside one's mathematical sophistication, it is not hard to feel that (27) and (28) are mutually inconsistent after all.

A useful analogy, noted above, is with our naïve reactions to vacuously true universal quantifications:

(33). Every dolphin in Oxford has arms and legs.

A natural inclination is to judge (33) false, even if one doubts that there are no dolphins in Oxford. That resistance is explicable by the hypothesis that we accept (34) on the basis of background information about dolphins, and are then inclined to reject (33) as inconsistent with (34):

(34). Every dolphin in Oxford lacks arms and legs.

That suggests heuristics for universal quantification analogous to (HCC) and (HCC*):

(HUQ). Given that φ is inconsistent with ψ, treat 'Every σ φs' as inconsistent with 'Every σ ψs'

(HUQ*). If you accept one of 'Every σ φs' and 'Every σ $\neg\varphi$s', reject the other

On the standard semantics for the universal quantifier, (HUQ) and (HUQ*) go extensionally wrong when and only when σ is empty in extension.

We can come to recognize the limitations of (HUQ) and (HUQ*) through natural reasoning. For instance, suppose that our rejection of (33) leads us to accept its negation:

(35). Not every dolphin in Oxford has arms and legs.

From (35) we can validly reason to (36), and thence to (37):

(36). Some dolphin in Oxford lacks arms and legs.

(37). There is a dolphin in Oxford.

But we know (37) to be false. That may lead us to realize that (33) is not false, though its utterance may induce a false presupposition. (HUQ) and (HUQ*) are fallible heuristics, defeasible by theoretical reflection, but they are still our default.

There may be a more general cognitive pattern underlying these heuristics. For example, it is plausible that we use analogues of (HCC) and (HCC*) for indicative as well as subjunctive conditionals, and

analogues of (HUQ) and (HUQ*) for generic as well as universal quantifiers. Very roughly, we ignore the empty case. We continue using heuristics based on that principle, even when the empty case is obviously relevant, until we resort to conscious reflection. Indeed, we may tend to use suppositional reasoning in evaluating universal and generic generalizations as well as conditionals. For instance, when asked to evaluate (33) or its generic analogue ('Dolphins in Oxford have arms and legs'), we may suppose that something is a dolphin in Oxford, and ask ourselves whether it has arms and legs.

Our theoretical grasp of universal quantification is currently more secure than it is of counterfactuals conditionals. We are consequently more comfortable in overruling (HUQ) and (HUQ*) than in overruling (HCC) and (HCC*). But it was not always so. Centuries of confusion about the existential import or otherwise of the universal quantifier bear witness to the difficulty of achieving a clear view of the truth-conditions of sentences of our native language formed using the most basic logical constants. Those who take themselves to have provided clear examples of false counterpossibles may be in a similar position to traditional logicians who took themselves to have provided clear examples of false universal generalizations with empty subject terms. Indeed, the primitively compelling nature of heuristics such as (HUQ) and (HUQ*) may have been the main obstacle to achieving a clear view of the truth-conditions of universal generalizations.

Imagine a philosopher attempting to craft a semantics for the universal quantifier to vindicate the heuristically driven judgments that (33) is false while (34) is true. He may invest immense patience and ingenuity in his project, but it is not going to end well. We should be similarly wary of attempts to craft a semantics for the counterfactual conditional to vindicate the heuristically driven judgments that some counterpossibles are false while others are true. There is a danger in semantics of unintentionally laundering cognitive biases into veridical insights, a danger evident in the semantics of generics.

With the universal quantifier, clear understanding was finally achieved through systematic, highly general semantic and logical theorizing, rather than by a more data-driven approach. The same may well hold for the counterfactual conditional. At any rate, it is methodologically naïve to take the debate over counterpossibles to be settled by some supposed examples of clearly false counterpossibles. As we have seen, a simple and mostly reliable heuristic would lead us to judge them false even if they were true.

On the view developed here, our assessments of counterfactuals are often based on fallible heuristics such as (HCC) and (HCC*). How far should that view make us sceptical more generally about reliance on pre-theoretic assessments of counterfactuals in philosophy, semantics and elsewhere?

The heuristics are reliable over wide ranges of cases. Just as we can gain lots of perceptual knowledge by relying on perceptual heuristics that are reliable over wide ranges of cases but fail under special conditions, so we can gain lots of modal knowledge by relying on heuristics such as (HCC) and (HCC*). Blanket scepticism is not a sensible response. Moreover, many assessments of counterfactuals will not rely on (HCC) and (HCC*) at all. (HCC) and (HCC*) are fundamentally devices for moving from the acceptance of one unnegated counterfactual to the rejection of another. They are not devices for accepting unnegated counterfactuals in the first place. Even if rejecting the latter counterfactual depends on the heuristic, accepting the former need not. Arguably, the key judgments in thought experiments, for instance that in such-and-such a Gettier case the subject would not know, involve the acceptance of unnegated counterfactuals, for which (HCC) and (HCC*) are not needed. Moreover, nothing said here impugns the reliability of counterfactual judgments made on the basis of mathematical reasoning. Even without theoretical reflection, we can inhibit the operation of the heuristics, as we sometimes need to do in order to maintain consistency. For instance, as already noted, we can continue the imaginative search for counterfactual consequences of a subjunctive supposition in an open-minded way that allows contradictions to arise. Nevertheless, the theorist who overrides the heuristic in favour of more reflective considerations should expect to feel some residual unease, at least at first. No matter how cogent the reflective considerations, the heuristic is too stupid to understand them; instead, it just goes on blindly pressing to have its way. If our access to the logic and semantics of our own language is essentially mediated by fallible heuristics, true theories may always feel Procrustean to us.

A Pure Logic-Based Approach
to Natural Reasoning

Lasha Abzianidze

TiLPS, Tilburg University, Tilburg, The Netherlands
L.Abzianidze@uvt.nl

Abstract

The paper presents a model for natural reasoning that combines theorem proving techniques with natural logic. The model is a tableau system for a higher-order logic the formulas of which resemble linguistic expressions. A textual entailment system LangPro, an implementation of the model, represents a tableau-based prover that directly operates on linguistic expressions. After training and evaluating on a textual entailment dataset, the prover shows accuracy comparable to the state-of-the-art results with almost perfect precision. Due to its reliable judgements, the system is also able to detect dubious problems in the dataset.

1 Introduction

A task of recognizing textual entailments (RTE) is one of the popular ways of testing reasoning capacity of NLP systems in natural language. RTE is usually a 3-way classification problem where a pair of T text and H hypothesis is classified based on whether T entails, contradicts or is neutral to H. A judgements shared by majority of human annotators are usually considered as a gold answer to an RTE problem. Hence the task is considered to evaluate the systems on human reasoning over natural language text.

While humans reason over natural language expressions, they heavily relay on the meaning of the expressions and use their reasoning capacity for drawing conclusions. The situation in RTE challenges is somewhat different from this. The systems are more concentrating on learning regularities from the data and later an RTE problem is classified based on how much it fits in the learned regularities. This type of approaches are usually robust but also imprecise. Their crucial drawback is inability of composing meanings of several facts and making conclusions based on them. On the other hand, approaches based on some logical language are brittle as translating from natural language into the formal language itself represents a difficult problem.

In this paper, we opt for a formal logic-based account for natural reasoning where the translation problem is facilitated with a highly expressive language. In this way, we mainly concentrate on the reasoning part and account for wide-coverage natural reasoning over linguistic text. First, the formal language and its calculus will be introduced. Then their extensions will be outlined based on which a theorem prover is constructed. We describe how the prover reasons over natural language expressions and demonstrate its performance over the SICK RTE [14] data set. The paper ends with a conclusion and the words about future work.

2 An analytic tableau system for natural logic

An analytic tableau system for natural logic [18] incorporates two main ideas according to which there are a formal semantic representation that resembles the linguistic surface form and a tableau proof system that acts on the structures of this representation. In this way, Muskens

$$\dfrac{\texttt{every}\ A\ B : [\ ] : \mathbb{T}}{A : [d] : \mathbb{F} \qquad B : [d] : \mathbb{T}}\forall_{\mathbb{T}}\ \text{where } d \text{ is } old \qquad \dfrac{\texttt{who}\ A\ B : [\vec{C}] : \mathbb{F}}{A : [\vec{C}] : \mathbb{F} \qquad B : [\vec{C}] : \mathbb{F}}\wedge_{\mathbb{F}}$$

$$\dfrac{\texttt{some}\ A\ B : [\ ] : \mathbb{T}}{\begin{array}{c} A : [c] : \mathbb{T} \\ B : [c] : \mathbb{T} \end{array}}\exists_{\mathbb{T}}\ \text{where } c \text{ is } fresh \qquad \dfrac{\texttt{who}\ A\ B : [\vec{C}] : \mathbb{T}}{\begin{array}{c} A : [\vec{C}] : \mathbb{T} \\ B : [\vec{C}] : \mathbb{T} \end{array}}\wedge_{\mathbb{T}} \qquad \dfrac{A\ B : [\vec{C}] : \mathbb{X}}{A : [B, \vec{C}] : \mathbb{X}}\text{PUSH}$$

$$\dfrac{\begin{array}{c} A : [\vec{C}] : \mathbb{T} \\ B : [\vec{C}] : \mathbb{F} \end{array}}{\times}\leq\!\times\ \text{where } A \leq B \qquad \dfrac{\begin{array}{c} A : [\vec{C}] : \mathbb{T} \\ B : [\vec{C}] : \mathbb{T} \end{array}}{\times}\text{DISJ}\times\ \text{where } A \mid B$$

Figure 1. The rules that are used by the tableau in Figure 2

in [18] regards natural language as a formal logical language and models natural reasoning in the same style as it is done for formal logics in terms of proof systems.[1]

The adopted semantic representation in [18] is a sort of functional type logic—the language of simply typed λ-calculus where terms are interpreted as functions.[2] For example, a term of type (α, β) is interpreted as a total function from objects of type α to objects of type β. Linguistic expressions are modelled by the terms, called Lambda Logical Forms (LLFs), built up mainly from lexical terms. Examples of LLFs are given below:

$$\text{Some runner who adores Mary won} \tag{1}$$

$$\texttt{some}_{(et)(et)t}\ \big(\texttt{who}_{(et)(et)et}\ (\texttt{adore}_{eet}\ \texttt{Mary}_e)\ \texttt{runner}_{et}\big)\ \texttt{won}_{et} \tag{1a}$$

Apart from resembling the surface forms, LLFs are comparable to the logical forms of generative grammar [10], the abstract terms of abstract categorial grammar (ACG) [9], or the terms built from multi-dimensional signs of λ-grammar [17].

A tableau system for natural logic, in short a natural tableau, is a signed tableau over LLFs. A tableau entry has three components: an LLF, a list of argument terms and a sign which is either true $\mathbb{T}$ or false $\mathbb{F}$. For instance, $\texttt{love} : [\texttt{Mary}, \texttt{John}] : \mathbb{T}$ is a well-formed entry (i.e. node) of a natural tableau. The semantics behind the entry is that a term resulted from applying an LLF to the arguments in a list order, $\texttt{love Mary John}$, is evaluated as the truth value of the attached sign, i.e. as true in this case.

The natural tableau comes with an inventory of inference rules. The rules are used to decompose complex nodes into shorter ones. Since the tableau aims to model natural reasoning, the inventory is expected to contain a plethora of rules. Several tableau rules for Boolean operators, determiners and lexical items with certain algebraic properties (including monotonicity) are already presented in [18]. Some of those rules are given in Figure 1.[3] The intuition behind the rules is simple; for example, the rule $\forall_{\mathbb{T}}$ asserts that if every A does B, then d is

---

[1] At first glance this decision is in the same vein as Montague's proposal in [15], but [18] gives the calculus over a formal language in contrast to Montague's solution to translate the English sentences into a formal logic.

[2] Actually [18] employs a relational type logic since interpreting terms as relations is more intuitive and simple according to [16]. Since entailment relations of both functional and relational type logics are the same [16], here we use a more common interpretation that is the functional one. While the original work [18] considers a type system with three basic types e, s and t (corresponding to entities, possible worlds and truth values, respectively), we will omit s type as no examples involving intentionality is discussed in this paper.

[3] While presenting the rules, we assume that: $\vec{C}$ stands for a sequence of terms; $A \leq B$ denotes $\forall \vec{X}(A\vec{X} \rightarrow B\vec{X})$ which informally says that A entails or is subsumed by B; and $A \mid B$ denotes $\neg \exists \vec{X}(A\vec{X} \wedge B\vec{X})$ which means

1 some (who (adore Mary) runner) won $: [\,] : \mathbb{T}$
2 every (who (love Mary) athlete) lost $: [\,] : \mathbb{T}$
|
$3^{\exists_{\mathbb{T}}[1]}$ who (adore Mary) runner $: [c] : \mathbb{T}$
$4^{\exists_{\mathbb{T}}[1]}$ won $: [c] : \mathbb{T}$
|
$5^{\wedge_{\mathbb{T}}[3]}$ adore Mary $: [c] : \mathbb{T}$
$6^{\wedge_{\mathbb{T}}[3]}$ runner $: [c] : \mathbb{T}$
|
$7^{\text{PUSH}[5]}$ adore $: [\text{Mary}, c] : \mathbb{T}$

$8^{\forall_{\mathbb{T}}[2]}$ who (love Mary) athlete $: [c] : \mathbb{F}$ $9^{\forall_{\mathbb{T}}[2]}$ lost $: [c] : \mathbb{T}$
$10^{\text{DISJ}\times[4,9]}$ $\times$

$11^{\wedge_{\mathbb{F}}[8]}$ love Mary $: [c] : \mathbb{F}$ $12^{\wedge_{\mathbb{F}}[8]}$ athlete $: [c] : \mathbb{F}$
| $13^{\leq\times[6,12]}$ $\times$
$14^{\text{PUSH}[11]}$ love $: [\text{Mary}, c] : \mathbb{F}$
$15^{\leq\times[7,14]}$ $\times$
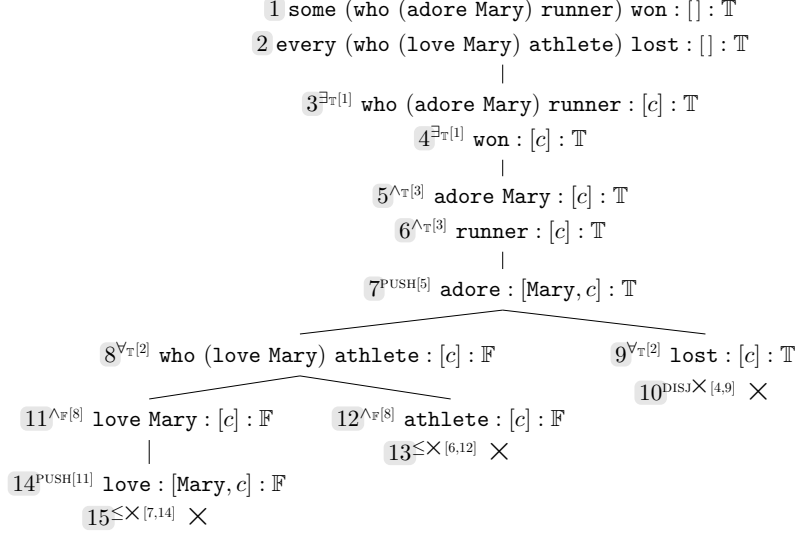
Figure 2. The tableau proving that *some runner who adores Mary won* contradicts *every athlete who loves Mary lost*. The nodes are enumerated and annotated with a source rule application.

not A or d does B given that d is some entity in a considered situation. With the help of the rules, it is possible to check LLFs on consistency and hence for entailment and contradiction too. For example, to check whether LLF_1 contradicts LLF_2, it is sufficient to check whether LLF_1 and LLF_2 together are consistent. The latter is done by starting a tableau with entries $LLF_1 : [\vec{C}] : \mathbb{T}$ and $LLF_2 : [\vec{C}] : \mathbb{T}$, where $\vec{C}$ is a sequence of terms feeding LLFs till the truth values. In this way, the tableau in Figure 2 proves inconsistency of sentences by showing that they are true together in no possible situation, i.e. all tableau branches are closed. All the rules used in the tableau can be found in Figure 1.

3   A wide-coverage natural language prover

3.1   Extending the natural tableau

For a wide-coverage tableau system we decide on what LLFs should look like for unrestricted sentences and also extend the format of tableau entries. We obtain LLFs automatically from the Combinatory Categorial Grammar (CCG) derivations [1, 2]. While doing so, it is possible to obtain LLFs typed with non-directional syntactic types prior to the final LLFs of semantic types. From the perspective of the tableau system, LLFs with syntactic types offer better matching between tableau nodes and antecedents of the rules than LLFs with semantic types.[4] Due to this reason, in the extended version of the tableau system LLFs are typed with syntactic

---

that A and B are disjoint concepts. The subsumption and disjoint relations for lexical constants are assumed to be a part of background knowledge.

[4]Note that an LLF of type *et* can ambiguously correspond to a bare singular noun phrase or an intransitive verb phrase. But further decomposition of the LLF in a tableau itself requires resolving this ambiguity which complicates the whole process. On the other hand, an LLF of syntactic type contains no such kind of ambiguity.

types. For instance, the LLF of syntactic type for (1) is (1b) which differs from (1a) only in types.[5] Note also that new LLFs now more resemble the abstract terms of ACG [9] and the signs of λ-grammar [17] due to their syntactic types.

$$\mathbf{some}_{n,vp,s} \; \big(\mathbf{who}_{vp,n,n} \; (\mathbf{adore}_{np,vp} \; \mathbf{Mary}_{np}) \; \mathbf{runner}_n\big) \; \mathbf{won}_{vp} \tag{1b}$$

Terms of semantic types are still needed while modelling certain phenomena. For example, an intersective adjective *red* is modelled as a term of type (n, n), but additionally a term of semantic type et is also necessary to assert redness of an object. Introduced fresh entity terms are also of type e. In order to accommodate both terms of syntactic type and semantic type in the same language, [2] introduces a subtyping relation over syntactic and semantic types. For example, given that $e \sqsubseteq np$, $s \sqsubseteq t$ and $n \sqsubseteq et$, terms like $\mathbf{love}_{np,vp} \, \mathbf{Mary}_{np} \, c_e$ and $\mathbf{athlete}_n \, c_e$ are well-typed terms of the new language with the extended type system.

Another extension to the natural tableau is to add a modifier set in tableau entries. As argued in [2], the set is used to save a modifier term that is indirectly applied to its head. A modifier is discharged from the set and applied to the main LLF of an entry when the LLF is a lexical term. This technique is used for the nouns with several adnominals or verbs with several adverbs. The trick with the modifier set solves the problem of event modification without introducing an event variable or the existential closure operator in LLFs, opposed to the approach in ACG [20]. Note that the extension is conservative in the sense that the tableaux generated with [18], e.g., the one in Figure 2, are still available in the extended version. This is done by considering entries with the old format as having an empty modifier set.

3.2   A theorem prover for natural language

Our next step is to present a theorem prover for natural logic [1] implemented based on the extended natural tableau system. The natural logic prover, like its theoretical model, has a modular architecture involving 3 main components: a knowledge base (KB), an inventory of rules (IR), and a proof engine (PE). The KB contains the hyponymy/hypernymy and antonymy relations of WordNet [8] as facts. The IR consists of around 80 rules part of which are taken from [18] and the rest are designed manually based on RTE training datasets [2]. Apart from new rules for passives, modifier-head constructions, copula, expletives, prepositional phrases, etc., there are also *admissible* (i.e. shortcut) rules in the IR which contribute to shorter proofs.

In order to reason over natural language expressions in an automatized way, the prover is paired with one of the state-of-the-art CCG parsers, C&C [6] or EasyCCG [11], and the LLF generator. The latter module produces LLFs from a CCG derivation by first replacing the CCG categories with syntactic types, then correcting inadequate analyses and finally type-raising determiners. The whole pipeline results in a theorem prover for natural language, called LangPro. The chart in Figure 3 shows how LangPro works.[6]

Compared to the forefront RTE system Nutcracker [3], based on first-order logic theorem provers and model builders, LangPro employs more expressive higher-order logic. As a result our system models intersective adjectives and higher-order quantifiers properly in contrast to Nutcracker. On the other hand, LangPro is backed up by a formal logical language opposed to NatLog [12], a prominent RTE system motivated by natural logic. Due to this reason NatLog is not able to reason over several premises and cannot account for logical laws like of De Morgan.

---

[5]An employed set of atomic types $\{n, np, pp, s\}$ corresponds to basic syntactic types of CCG for noun, noun phrase, prepositional phrase and sentence, respectively. Often (np, s) type is abbreviated as vp. Hereafter, a lexical term of syntactic type is denoted by its lemma and is written in boldface.

[6]See the online demo of LangPro at: `http://lanthanum.uvt.nl/labziani/tableau`

SICK-2865: `Nobody is riding a bike` ? `Two people are riding a bike`
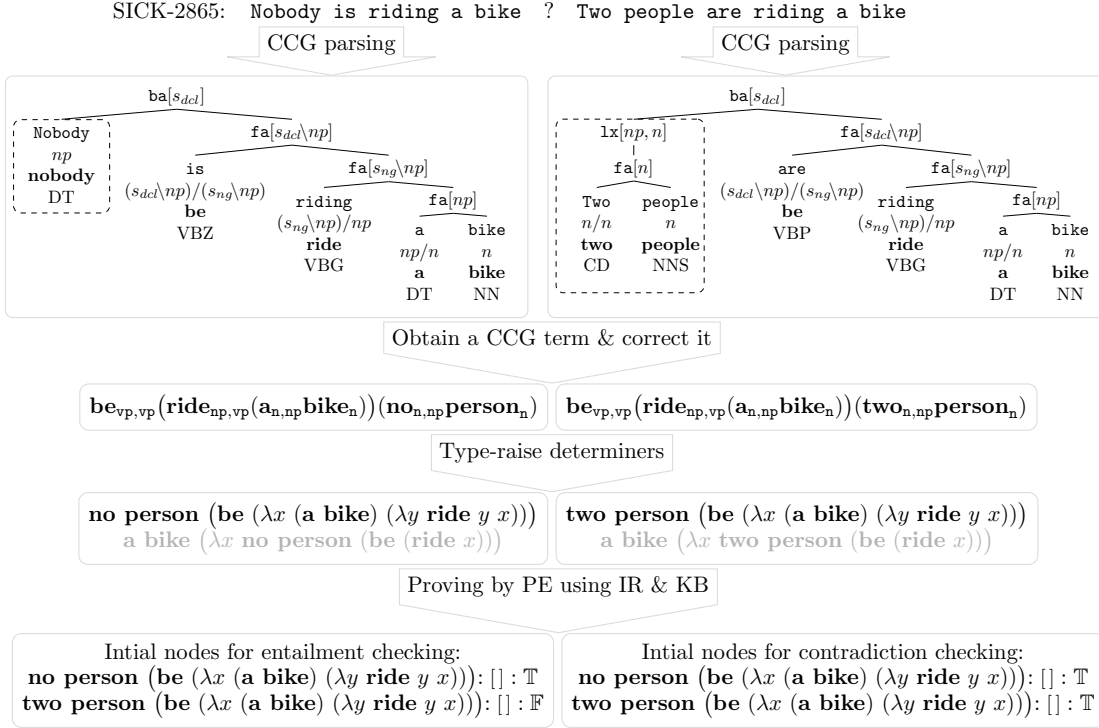
Figure 3. LangPro processing an RTE problem. To maintain the process efficient only one LLF from possibly several LLFs is employed by the tableau prover.

4   Adapting to SICK

4.1   Several new rules

The SICK (Sentences Involving Compositional Knowledge) dataset [14] is a set of about 10K text-hypothesis pairs annotated with three labels: entailment, contradiction and neutral.[7] The dataset is divided in tree parts: TRIAL (5%), TRAIN (45%) and TEST (50%). Following the RTE challenge [13], we keep the TEST portion unseen while using the rest of the data for development. The learning process, as described in [1], consists of three sub-learning components: improving the LLF generator, adding new facts to the KB and introducing new rules in the IR. The process is carried out manually while being facilitated with LangPro.

Analysis of false negatives (i.e. the non-neutral problems that are classified as neutral) reveals that often they are caused by inadequate CCG derivations. For instance, one of the most common mistakes the CCG parsers make is wrong PP-attachments. We design a couple of tableau rules to fix these mistakes. For example, the VP_PP converts a prepositional phrase from complements to modifiers and is used in problems like SICK-2171 (see Table 1). The closure rule PP_ATT, informally speaking, finds phrases like $((\mathbf{mix}\,A)\,\mathbf{in}\,B)$ and $(\mathbf{mix}\,(A\,\mathbf{in}\,B))$ equivalent, hence contributes to the proof of entailment pairs like SICK-4879.

---

[7] It was used as a benchmark for the RTE challenge [13] at SemEval-14: `http://alt.qcri.org/semeval2014/task1/`

ID	Gold/LangPro	Problem (T: text & H: hypothesis)
247	C/C	T: The woman is not wearing glasses or a headdress H: A woman is wearing an Egyptian headdress
344	N/C	T: An Asian woman in a crowd is not carrying a black bag H: An Asian woman in a crowd is carrying a black bag
410	E/E	T: A group of scouts are hiking through the grass H: Some people are walking
1696	E/E	T: A little cat is drinking fresh milk H: The milk is being drunk by a cat
1745	N/N	T: A man is pushing the buttons of a microwave H: A man is being pushed toward the buttons of a microwave
2171	E/E	T: An egg is being (cracked$_{\mathrm{pp,vp}}$ (into a bowl)$_{\mathrm{pp}}$) by a woman H: A woman is cracking an (egg$_{\mathrm{pp,n}}$ (into a bowl)$_{\mathrm{pp}}$)
3535	N/N	T: Someone is boiling okra in a pot H: Someone is being boiled with okra in a pot
3537	C/N	T: Nobody is cooking okra in a pan H: Someone is cooking okra in a pan
4443	N/E	T: A man is singing to a girl H: A man is singing to a woman
4879	C/C	T: There is no man ((mixing vegetables) (in a pot)$_{\mathrm{vp,vp}}$) H: A man is mixing (vegetables (in a pot)$_{\mathrm{n,n}}$)
5264	N/E	T: A person is folding a sheet H: A person is folding a piece of paper
8501	N/C	T: The person is not going into the water H: The man is going into the water

Table 1. Problems from TRIAL and TRAIN with gold and LangPro's answers

$$\frac{V_{\mathrm{pp},\alpha}\,(p^{\mathrm{IN}}_{\mathrm{np,pp}}\,D):[\vec{C}]:\mathbb{X}}{p^{\mathrm{IN}}_{\mathrm{np},\alpha,\alpha}\,D\,V_{\alpha}:[\vec{C}]:\mathbb{X}}\text{VP\_PP}
\qquad
\frac{\mathbf{crack}_{\mathrm{pp,vp}}\,(\mathbf{into}^{\mathrm{IN}}_{\mathrm{np,pp}}\,b_e):[c_e]:\mathbb{T}}{\mathbf{into}^{\mathrm{IN}}_{\mathrm{np,vp,vp}}\,b_e\,\mathbf{crack}_{\mathrm{vp}}:[c_e]:\mathbb{T}}$$

$$\frac{\begin{array}{c}[p^{\mathrm{IN}}\,b]:V:[d,\vec{C}]:\mathbb{F}\\ V:[d,\vec{C}]:\mathbb{T}\\ p^{\mathrm{IN}}_{\mathrm{pp}}:[b,d]:\mathbb{T}\end{array}}{\times}\text{PP\_ATT}
\qquad
\frac{\begin{array}{c}[\mathbf{in}^{\mathrm{IN}}_{\mathrm{pp,vp,vp}}\,b]:\mathbf{mix}_{\mathrm{np,vp}}:[d,c]:\mathbb{F}\\ \mathbf{mix}_{\mathrm{np,vp}}:[d,c]:\mathbb{T}\\ \mathbf{in}^{\mathrm{IN}}_{\mathrm{pp}}:[b,d]:\mathbb{T}\end{array}}{\times}$$

The dataset contains many expressions like *body of*, *group of*, *slice of*, etc. In order to correctly reason over the sentences involving these expressions, a new rule GRP_OF is introduced. Roughly speaking the rule asserts for *group members* whatever holds for a *group*. With the help of the rule it is possible to relate *group of scouts* to *people* in SICK-410 and prove H from T.

$$\frac{\begin{array}{c}G_{\mathrm{pp,n}}\,(\mathbf{of}^{\mathrm{IN}}_{\mathrm{np,pp}}\,b_e):[c_e]:\mathbb{T}\\ V:[c_e]:\mathbb{X}\end{array}}{V:[b_e]:\mathbb{X}}\text{GRP\_OF}
\qquad
\frac{\begin{array}{c}\mathbf{group}_{\mathrm{pp,n}}\,(\mathbf{of}^{\mathrm{IN}}_{\mathrm{np,pp}}\,b_e):[c_e]:\mathbb{T}\\ \mathbf{walk}_{\mathrm{vp}}:[c_e]:\mathbb{T}\end{array}}{\mathbf{walk}_{\mathrm{vp}}:[b_e]:\mathbb{T}}$$

where $G \in \{\mathbf{body},\mathbf{group},...\}$

Acc% on TRAIN	Not alig.	Aligned	Both
RAL=50	75.24	80.80	80.84
Disj & RAL=50	76.87	80.87	80.93
Disj & RAL=800	-	**81.33**	**81.44**

Gold \ LangPro	E	C	N
Entailment	**802**	0	612
Contradiction	1	**501**	218
Neutral	25	7	2761

(a) Results of the experiments on alignment and the disjoint relations. Employed LLFs are obtained from the C&C derivations.

(b) The confusion matrix of LangPro (with the best configuration) on TEST. LLFs are obtained from C&C and EasyCCG derivations.

Table 2. Results of the experiment on TRAIN and the evaluation on TEST

4.2 Experiments

Most of the sentences in SICK are generated by altering other sentences. As a result, a text and a hypothesis often have at least one multiword chunk in common (e.g., SICK-2865 in Figure 3). In general, aligning T and H and neglecting shared phrases are commonly used by RTE systems. To test the effect of alignment, we add an optional aligner to the LLF generator. The aligner finds a set of compound terms shared by all the LLFs of a problem and treats them as constant terms, i.e. with no internal structure. Since the PE is not able to expand these constant terms, which is believed to be worthless, this increases chances for finding a proof. During experiment the prover is limited with 50 rule applications and three options are tested on TRAIN. The first row of Table 2a shows the results of the experiment. The prover performs more than 5% better with aligned LLFs in contrast to the non-aligned ones. Note that alignment process can make some information inaccessible for the prover that might be crucial for closing a tableau. For this reason, we also test a combined method where if a proof is not found with aligned LLFs then non-aligned ones are tried by the prover. The combined system shows little improvement over the aligned one.[8]

The hypernymy relation is very important for identifying contradictions and closing tableau branches. With respect to the dataset, it seems that WordNet provides the prover with sufficient hypernymy information. In Figure 2, it was shown how disjoint concepts can contribute to reasoning. Unfortunately, it is not clear how to get the high quality disjoint relations from WordNet in order to check the contribution of the relation to the prover. For this reason, for each problem in TRAIN we collected pairs of nouns and annotated 500 most frequent ones (with around 30% of coverage) on disjointness. The annotations were asserted as facts in the KB. The results in Table 2a show that only the system with non-aligned LLFs got the highest improvement (1.6%). Little improvement on the aligned system was expected as it treats some chunks as constants and is less effected with lexical knowledge than the non-aligned version. In case of the effective rule application limit (800) [1] the improvement over the aligned version still stays minor. This suggests that the rule application strategy of the prover rarely needs the information about disjoint concepts.[9]

---

[8] The reason of the improvement is the problems with prepositional phrases like *into a bowl* in SICK-2171. Aligning such kind of phrases prevents the tableau rules, e.g. VP_PP, to further unfold the internal structure of the phrase.

[9] After introducing the disjoint relations, more than 20 sentences in TRAIN obtained inconsistent meaning. For example, the hypothesis in SICK-3537 is parsed by C&C in such a way that it entails *someone is an okra*. The prover finds this meaning inconsistent as *okra* and *person* are disjoint concepts. As a result an inconsistent sentence is an indicator for wrong parsing. To avoid misclassification due to this kind of sentences, first each sentence is separately checked by the prover on consistency and afterwards the whole problem is analyzed.

TEST LangPro+Disj	4927 problems			Systems on TEST	Prec%	Rec%	Acc%
	Prec%	Rec%	Acc%	Illinois-LH	81.56	**81.87**	**84.57**
Baseline (majority)	-	-	56.69	ECNU	84.37	74.37	83.64
C&C 800 aligned	97.42	56.51	80.56	UNAL-NLP	81.99	76.80	83.05
EasyCCG 800 aligned	97.60	57.22	80.88	**LangPro** (Comb.)	**97.53**	61.06	82.48
Combined 800 aligned	97.52	60.73	82.34	SemantiKLUE	85.40	69.63	82.32
C&C 800 both	97.43	56.84	80.70	Meaning Factory	93.63	60.64	81.59
EasyCCG 800 both	**97.61**	57.45	80.98	LangPro [1]	97.95	58.11	81.35
Combined 800 both	97.53	**61.06**	**82.48**	Nutcracker [19]	-	-	78.40

(a) Evaluation of the versions of LangPro (with disjoint facts) on TEST

(b) Comparing LangPro to the top RTE systems of SemEval-14

Table 3. Evaluation results of LangPro on the SICK TEST

5   Evaluation

For the evaluation we use TEST which was held out during the adaptation period. In addition to the LLFs generated from the C&C derivations, we also employ the LLFs obtained from another CCG parser, EasyCCG.[10] To evaluate the quality of LLFs with different source parsers, first the prover with the two best configurations is evaluated for each type of LLFs. The results of the evaluation are given in Table 3a. According to each measure, the prover reasons better with LLFs based on EasyCCG compared to those from C&C. This effect is explained by the architecture of EasyCCG. As the parser uses probabilistic approach only for supertagging (e.i. assigning the CCG categories to lexical items), it makes more probable that very similar T and H will be tagged and later parsed similarly.

The best performance is achieved by the combined prover that finds a proof if either the C&C-based or EasyCCG-based provers find it (in case of the different positive answers from the provers, the combined one classifies a problem as neutral). An extremely high precision (>97%) of the prover on unseen problems is explained by its sound rules. The analysis of false positives over TRAIN reveals that most of the *mistakes* represent dubious cases (e.g., SICK-344,8501,5264) or often caused by the multi-senses from WordNet (e.g., SICK-4443). The confusion matrix in Table 2b shows that it is almost never the case that entailment and contradiction are confused by the prover. Better performance over contradiction problems is explained by the fact that decomposing entries with true signs is more efficient and respectively a tableau starts with true entries while checking a problem on contradiction.

LangPro with these results makes in top 5 of the SemEval RTE challenge [13] while still being a pure logic-based system in contrast to the rest of the systems.[11] It is true that in comparison to statistical RTE systems, our system is more brittle as it is sensitive towards errors from the parsers and lack of tableau rules. But it can still prove relations that state-of-the-art systems are not able to account for. For example, SICK-247, 1696 were wrongly classified by almost all systems in the top 5 of SemEval-14[12], but LangPro is able to prove them correctly. A tableau proof for SICK-1696 is given in Figure 4. Also our system avoids to classify problems like SICK-1745, 3535 as positive in contrast to each system in the top 4.

---

[10]Unlike the C&C derivations, the EasyCCG derivations were not used in the development of the LLF generator [1]. In this way, we check whether the generator generalizes enough well for other CCG parsers.

[11]The brother system of Nutcracker, the Meaning Factory [4], obtains similar results as our system, but it employs machine learning techniques and similarity measures.

[12]Except Meaning Factory that correctly guesses only SICK-1696.

1 $\mathbf{a_{n,vp,s}}$ $(\mathbf{little_{n,n}}\ \mathbf{cat_n})$ $(\mathbf{be_{vp,vp}}(\lambda x.\,\mathbf{a_{n,vp,s}}\ (\mathbf{fresh_{n,n}}\ \mathbf{milk_n})\ (\lambda y.\,\mathbf{drink_{np,vp}}\ y_{np}\ x_{np}))) : [\,] : \mathbb{T}$

2 $\mathbf{the_{n,vp,s}}$ $\mathbf{milk_n}$ $(\mathbf{be_{vp,vp}}(\mathbf{be_{vp,vp}}(\lambda x.\,\mathbf{a_{n,vp,s}}\ \mathbf{cat_n}\ (\lambda y.\,\mathbf{by_{np,vp,vp}}\ y_{np}\ \mathbf{drink_{vp}}\ x_{np})))) : [\,] : \mathbb{F}$

|

3 $^{\exists_{\mathbb{T}}[1]}$ $\mathbf{little\ cat} : [c] : \mathbb{T}$

|

6 $^{\lambda\text{PULL, AUX, }\exists_{\mathbb{T}}[1]}$ $\mathbf{a}\ (\mathbf{fresh\ milk})\ (\lambda y.\,\mathbf{drink}\ y\ x) : [\,] : \mathbb{T}$

|

7 $^{\text{SUB-ADJ}[3]}$ $\mathbf{cat} : [c] : \mathbb{T}$

8 $^{\exists_{\mathbb{F}}[2]}$ $\mathbf{milk} : [c] : \mathbb{F}$ 9 $^{\exists_{\mathbb{F}}[2]}$ $\mathbf{milk} : [c] : \mathbb{T}$

| 10 $^{\exists_{\mathbb{F}}[2]}$ $\mathbf{be}(\mathbf{be}(\lambda x.\,\mathbf{a\ cat}\ (\lambda y.\,\mathbf{by}\ y\ \mathbf{drink}\ x))) : [c] : \mathbb{F}$

14 $^{\text{SUB-ADJ},\exists_{\mathbb{T}}[6]}$ $\mathbf{milk} : [m_e] : \mathbb{T}$ $\square$ $^{\text{DISJ}}\times^{[7,9]}\ \times$

13 $^{\lambda\text{PULL},\exists_{\mathbb{T}}[6]}$ $\mathbf{drink}\ m\ c : [\,] : \mathbb{T}$

|

18 $^{\lambda\text{PULL,AUX,AUX,THE}_{\mathbb{T}}[2,14]}$ $\mathbf{a\ cat}\ (\lambda y.\,\mathbf{by}\ y\ \mathbf{drink}\ m) : [\,] : \mathbb{F}$

|

20 $^{\lambda\text{PULL},\exists c_{\mathbb{F}}[7,18]}$ $\mathbf{by}\ c\ \mathbf{drink}\ m : [\,] : \mathbb{F}$

|

21 $^{\text{PUSH-ARG}[20]}$ $\mathbf{by}\ c\ \mathbf{drink} : [m] : \mathbb{F}$

|

22 $^{\text{PASS}[21]}$ $\mathbf{drink} : [m,c] : \mathbb{F}$

|

25 $^{\text{PUSH-ARG,PUSH-ARG}[13]}$ $\mathbf{drink} : [m,c] : \mathbb{T}$

$\square$ $^{\leq}\times^{[22,25]}\ \times$

Figure 4. The tableau proving that SICK-1696 represents entailment. Some nodes from the initial tableau proof are omitted as they are not relevant for the proof.

6   Conclusion

We have presented the extended theory of the natural tableau and showed that it is viable theory for wide-coverage natural reasoning. The natural language theorem prover LangPro, based on that theory, achieves high competitive results on the SICK dataset while still being as reliable as theorem provers used to be for formal logics. The prover overcomes common shortcomings of most of the RTE systems nowadays on the market. It is fluent in reasoning over Boolean operators and quantifiers, has a quite expressive language where higher-order quantifiers or terms, like subsective adjectives, can be expresses and its reasoning skills are not limited to single-premised arguments. The theory and the prover are also able to provide a counterexample of an argument; every open tableau branch could offer a candidate for it.

A transparent decision procedure is another advantage of our model. A combination of LLFs, the forms similar to surface forms, and tableau rules, intuitively interpretable schematic rules, presents a suitable framework for studying human reasoning and information processing, e.g., to explain a complexity of a certain entailment in terms of a structure of a closed tableau.

In future, we plan to explore possibilities of improving the LLF generator since the performance of the system significantly hinges on the quality of LLFs. One of the possibilities is to check whether the 2nd or 3rd best derivations of EasyCCG contribute to better LLFs. Testing the prover on another RTE dataset, for example, those of the RTE challenges [7] or a more recent SNLI dataset [5], further challenges each component of the prover. For instance, taking into account characteristics of the annotation in the SNLI dataset, the prover needs a new device to anchor the events occurring in a text and a hypothesis.

References

[1] Abzianidze, L: A Tableau Prover for Natural Logic and Language. In the proceedings of EMNLP, ACL (2015)

[2] Abzianidze, L: Towards a Wide-coverage Tableau Method for Natural Logic. In Murata, T., Mineshima, K., Bekki, D. (eds.), New Frontiers in Artificial Intelligence, LNCS, vol. 9067, pp. 66–82. Springer Verlag (2015)

[3] Bos, J., Markert, K.: Recognising Textual Entailment with Robust Logical Inference. In: J. Quinonero-Candela, I. Dagan, B. Magnini, F. d'Alch-Buc (eds): Machine Learning Challenges, MLCW 2005, LNAI (3944), pp 404-426 (2006)

[4] Bjerva, J., Bos, J., Van der Goot, R., Nissim, M.: The Meaning Factory: Formal Semantics for Recognizing Textual Entailment and Determining Semantic Similarity. Proceedings of the 8th SemEval, pp 642-646 (2014)

[5] Bowman, S. B., Angeli, G., Potts, C., Manning, C. D.: A Large Annotated Corpus for Learning Natural Language Inference In the proceedings of EMNLP, ACL (2015)

[6] Clark, S., Curran, J.R.: Wide-Coverage Efficient Statistical Parsing with CCG and Log-linear Models. Computational Linguistics, 33(4) (2007)

[7] Dagan, I., Glickman, O., Magnini, B.: The PASCAL recognising textual entailment challenge. Machine Learning Challenges: Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment. LNCS, vol. 3944, pp. 177-190. Springer Berlin Heidelberg (2006)

[8] Fellbaum, Ch. (eds.): WordNet: an Electronic Lexical Database. MIT press (1998)

[9] de Groote, Ph.: Towards Abstract Categorial Grammars. In Proceedings of the 39th ACL Conference, pp. 148-155 (2001)

[10] Heim, I., Kratzer, A.: Semantics in Generative Grammar. Blackwell, Oxford (1998)

[11] Lewis, M., Steedman, M.: A* CCG Parsing with a Supertag-Factored Model. In Proceedings of the EMNLP 2014, pp. 990-1000. ACL (2014)

[12] MacCartney, B.: Natural Language Inference. Ph.D. Dissertation, Stanfor University, USA (2008)

[13] Marelli, M. *et al.*: SemEval-2014 Task 1: Evaluation of Compositional Distributional Semantic Models on Full Sentences through Semantic Relatedness and Textual Entailment. Proceedings of the 8th SemEval (2014).

[14] Marelli, M. *et al.*: A SICK Cure for the Evaluation of Compositional Distributional Semantic Models. In Proceedings of LREC, Reykjavik (2014)

[15] Montague, R.: English as a formal language. In: Linguaggi nella societa e nella Tecnica, pp. 189-224. Edizioni di Comunita (1970); reprinted in: Formal Philosophy; Selected papers of Richard Montague. Thomason, R. H. (eds.), pp. 108–221. Yale University Press (1974)

[16] Muskens, R.: Meaning and Partiality. CSLI, Stanford (1995)

[17] Muskens, R.: Language, Lambdas, and Logic. In: Kruijff, G., Oehrle, R. (eds.) Resource-Sensitivity, Binding and Anaphora. Studies in Linguistics and Philosophy, vol. 80, pp. 23–54. Springer, Heidelberg (2003)

[18] Muskens, R.: An Analytic Tableau System for Natural Logic. In: Aloni, M., Bastiaanse, H., de Jager, T., Schulz, K. (eds.) Logic, Language and Meaning. LNCS, vol. 6042, pp. 104–113. Springer, Heidelberg (2010)

[19] Pavlick, E., Bos, J., Nissim, M., Beller, C., Van Durme, B. & Callison-Burch, C.: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, pp. 1512–1522. ACL (2015)

[20] Winter, Y., Zwarts J.: Event Semantics and Abstract Categorial Grammar. In Kanazawa, M., Kornai, A., Kracht, M., Seki, H. (eds.) The Mathematics of Language. LNCS, vol. 6878, pp. 174-191. Springer, Heidelberg (2011)

Novelty and Familiarity for Free

David Beaver and Elizabeth Coppock

[1] University of Texas at Austin
[2] University of Gothenburg & Swedish Collegium for Advanced Study

1   Introduction

In this paper we offer a novel resolution of a familiar tension, that between approaches in which the difference between definites and indefinites is based on uniqueness, and those in which it is based on novelty/familiarity. Advocates of uniqueness-based approaches such as Horn & Abbott (2013) and Coppock & Beaver (2015) have pointed out cases where the description is unique, but not familiar. For example, Coppock & Beaver (2015) use the following example:

(1)   Jane didn't score the only goal. #It wasn't a bicycle kick, either.
       (where *the only goal* serves as an antecedent for *it*)

The fact that the definite description cannot be an antecedent for a subsequent pronoun means that it cannot be picking up a familiar discourse referent; a familiar discourse referent would still be available for subsequent anaphora. On the other hand, proponents of a familiarity-based approach such as Heim (1982), Szabó (2000) and Ludlow & Segal (2004) can point to the fact that the descriptive content of an anaphoric definite description need not be unique in any obvious sense. Take the following example from Heim (1982):

(2)   A glass$_i$ broke last night. The glass$_i$ had been very expensive.

This does not seem to imply that there is just one contextually-relevant glass. Ideally, a theory of definite descriptions should be able to explain both of these kinds of examples.

In this paper, we show that familiarity of (short) definites and novelty of indefinites can be derived from uniqueness and non-uniqueness respectively. Specifically, we show that if an ordinary dynamic semantics is defined to allow tracking of discourse referents, and an indexing mechanism is defined to allow identification of descriptions with referents, then both novelty of indefinites and familiarity of definites can be derived without stipulating lexically that the articles have these properties. The derivation relies entirely on principles that are commonly used (e.g. the uniqueness requirement for definites, and Heim's Maximize Presupposition principle). However, the derivation does not predict familiarity for all definites; in particular, familiarity for definites in case of semantic uniqueness will not be required. Furthermore, the uniqueness requirement will effectively drop away in case of familiarity.

We will explicate the proposal by first defining a dynamic system that embodies many of the insights in Heim's (1982,1983) seminal work on definites, and then showing how basic properties of her treatment can be derived and improved upon.

2   Partial File Logic

As a tool for representing dynamic meanings, we specify a logic that we call Partial File Logic (PFL). PFL has a basic type l for labels in addition to s, e, and t, and three truth values, as well as undefined entities $\#_\alpha$ for every type α, denoted by constants $\star_\alpha$ in the logic.

Type		Variables
e	:	x, y, z
s	:	w
l	:	i, j
$\langle l, e \rangle \quad (= \sigma)$	:	f, g, h
$\langle \sigma, \langle \sigma, t \rangle \rangle \quad (= \tau)$	:	p, q
$\langle e, t \rangle$	:	P, Q
$\langle e, \tau \rangle$	:	$\mathcal{P}, \mathcal{Q}$

Table 1: Variables and their types in Partial File Logic

A dynamic system, which borrows from Heim (1982) and the Amsterdam tradition from Groenendijk & Stokhof (1989) on, is built on top of this static foundation. The Heimian notion of a *sequence* is implemented as a function of type $\langle l, e \rangle = \sigma$. A dynamic proposition relates two sequences, and is thus of type $\langle \sigma, \langle \sigma, t \rangle \rangle \; (= \tau)$. Dynamic properties are type $\langle e, \tau \rangle$.

The syntax of the language is mostly standard for typed lambda calculus. The variable naming conventions are given in Table 1. To denote the result of applying predicate π to argument α, we write $\pi(\alpha)$, except in the case of an expression of type τ, where τ is shorthand for $\langle \sigma, \langle \sigma, t \rangle \rangle$, effectively a relation between two sequences. In that case, instead of $\phi(f)(g)$, where ϕ is an expression of type τ, we write $f[\phi]g$, where f is the 'input' sequence and g is the 'output' sequence. We sometimes use a dot (.) to separate a binder (λ, $\forall$, or $\exists$) from its scope; the scope should in that case be interpreted as extending as far to the right as possible.

Expressions of PFL are interpreted with respect to a model, a world, and an assignment. The extension of an expression α with respect to model M, world w, and assignment a is written $[\![\alpha]\!]^{M,w,a}$; sometimes the model and the assignment parameters are suppressed. A model is a tuple $\langle D_e, D_t, D_l, W, I \rangle$ subject to the following constraints:

- The domain of individuals D_e contains at least one individual along with the undefined individual of type e, denoted by $\#_e$.

- The domain of truth values D_t contains three truth values: T, F, and $\#_t$. We use $\#$ as a shorthand for $\#_t$.

- The domain of labels D_l is the set of integers.

- W contains at least one possible world.

- I is an interpretation function, assigning an intension to all of the constants of the language. The intension of a constant of type τ is a function from W to D_τ.

An assignment a is a total function whose domain consists of the variables of the language such that if u is a variable of type τ then $a(u) \in D_\tau$. Note that these assignments are for interpreting variables of PFL; they should not be confused with *sequences*, which are objects in the model, functions from labels to individuals. PFL uses Weak Kleene connectives, and undefinedness of a functor or an argument yields undefinedness of an application: $[\![A_{\langle \beta, \alpha \rangle}(B_\beta)]\!]^{M,w,a} = \#_\alpha$ if $[\![A]\!]^{M,w,a} = \#_{\langle \beta, \alpha \rangle}$ or $[\![B]\!]^{M,w,a} = \#_\beta$, $[\![A]\!]^{M,w,a}([\![B]\!]^{M,w,a})$ otherwise.

Using these tools, we define dynamic connectives as follows:

ABBREVIATION 1. **And**$\equiv \lambda\phi\lambda\psi\lambda f\lambda g \,.\, \exists h \,.\, f[\phi]h \,\wedge\, h[\psi]g$

ABBREVIATION 2. **Not** $\equiv \lambda\phi\lambda f\lambda g \,.\, f = g \wedge \neg\exists h \,.\, f[\phi]h$

ABBREVIATION 3. ϕ **Implies** $\psi \equiv$ **Not**$[\phi$ **And Not** $\psi]$

Let ∂_s denote the static ∂-operator from Beaver & Krahmer 2001 (yielding undefinedness if its complement is not true); then the dynamic partial operator ∂_d may be defined as follows.

ABBREVIATION 4. $\partial_d\phi \,\equiv\, \lambda f\lambda g \,.\, \partial_s(f[\phi]g) \wedge f[\phi]g$

The PFL dynamic connectives yield standard dynamic presupposition projection behavior (cf. Beaver 2001).

3 Updating Heim

Heim's (1982) implementation does not use type theory, but we can loosely describe her analysis of definite and indefinite DPs as giving them the same type as VPs, with both being essentially propositional. However, Heimian propositions are not the propositions of old, but rather Context Change Potentials. That is, for Heim, both a DP and a VP can provide a way of updating contexts to produce new contexts, so that e.g. "a cat$_i$" provides a way of updating a context so that in the output context the referent i is established to be a cat, and a VP like "purrs" is taken to share the same index, becoming "i purrs", which, once again, can be used to update a context in the obvious way. Identity between the cat and the purrer in "a cat purrs" is established by sharing of indices (sequential updates with "a cat$_i$" and "i purrs"). Thus, for Heim, predicates apply directly to numeric labels, so that e.g. "7 smiles" could be part of an LF. While a familiar subscripted index should have a meaning given by the input context, a novel index has an unconstrained value, but has a side-effect of extending the context so that it is defined on the new index.

We also take a more Montagovian line in our version, using type theory and functional application as the primary means of composition. In our variant of her system, labels for referents have a distinct type from individuals, and (dynamic) predicates apply to individuals rather than applying to labels. The nominal that a definite or indefinite article may in principle be labeled (e.g. *glass*) or unlabeled (e.g. *glass$_i$*). In other case, it denotes a dynamic property, i.e., a function from individuals (type e) to dynamic propositions (type τ).[1] Let us represent the static $\langle e,t \rangle$ property of being a glass with the non-logical constant **glass**. (The extension of this predicate at world w, $[\![\mathbf{glass}]\!]^w$ will depend on w.) We will call the corresponding dynamic property **Glass**. The latter may be defined in terms of the former as follows.

ABBREVIATION 5. **Glass** $\equiv \lambda x\lambda f\lambda g \,.\, f = g \wedge \mathbf{glass}(x)$

Analogous abbreviations will be made for the dynamic version of all basic static predicates. The unlabeled common noun *glass* translates as **Glass**:

TRANSLATION 1. *glass* $\rightsquigarrow$ **Glass**

Subscripting a noun with an index adds an additional constraint, which we capture using the constant **Labeled**. This constant is then used in the definition of how subscripted predicates $\mathcal{P}_i$ are to be interpreted, essentially forming a new dynamic property from the label property and the predicate property:

TRANSLATION 2. *glass$_i$* $\rightsquigarrow$ **Glass**$_i$

---

[1]In the case of a simple common noun like *glass*, no use is made of the dynamic nature of the property, but it becomes important in cases like *glass given to me by a friend*, where the noun phrase introduces a discourse referent.

ABBREVIATION 6. $\mathcal{P}_i \equiv \lambda x . \mathbf{Labeled}(i)(x) \mathbf{\ And\ } \mathcal{P}(x)$

So subscripting amounts to a dynamic version of intersection, i.e. intersection of the predicate property and the label property (cf. Heim and Kratzer's "Predicate Modification" rule).

The definition of **Labeled** is as follows, where $g \geq_i f$ can be read 'g extends f'.[2]

ABBREVIATION 7. $\mathbf{Labeled} \equiv \lambda i \lambda x \lambda f \lambda g . x = g(i) \wedge g \geq_i f$

Given these definitions, we have the following equivalence:

$$\begin{aligned}
\mathbf{Glass}_i &\equiv \lambda x . \mathbf{Labeled}(i)(x) \mathbf{\ And\ } \mathbf{Glass}(x) \\
&\equiv \lambda x \lambda f \lambda g \exists h . f[\mathbf{Labeled}(i)(x)]h \wedge h[\mathbf{Glass}(x)]g \\
&\equiv \lambda x \lambda f \lambda g \exists h . [x = h(i) \wedge h \geq_i f] \wedge [h = g \wedge \mathbf{glass}(x)] \\
&\equiv \lambda x \lambda f \lambda g . x = g(i) \wedge g \geq_i f \wedge \mathbf{glass}(x)
\end{aligned}$$

The definition in Abbreviation 7 implements the intuition that a label denotes a property, namely the property of being identical to the referent of the index. It also implements Heim's strategy of adding discourse referents to contexts on an as-needed basis. The studious reader may recall that Heim defines the semantics of predication in an unusual way, which can be understood in terms of two cases. First, suppose that a referent is familiar. In that case, predicating something of that referent will simply update the context so as to constrain the value of the referent in the appropriate way. But suppose instead that something is predicated of a referent that is not defined in the context. In that case, each sequence in the context is extended so as to provide values for the new referent, and these values are appropriately constrained. Thus in Heim's system, "i purrs" defines an update both for contexts in which i is familiar, and for contexts in which i is novel. Abbreviation 7 does something similar: **Labeled** maps from an index to the dynamic property of being identical to the value of that index, with the possible additional side-effect that the output sequence is defined on that index even if the input is not.

Given that linkage between DPs and VPs is established using variable binding, we introduce an operator to perform that binding. This operator, **Ex**, will play an important role in this paper. The intuition in the following definitions is that **Ex** takes a (dynamic) nominal property ($\mathcal{P}_1$) and a (dynamic) verbal property ($\mathcal{P}_2$), and then produces a dynamic proposition, which is a relation between assignments f and g. Specifically, the context f must be first updated using the nominal property, and then using the verbal property. But how can we establish that both properties hold of the same individual? To do that, we existentially quantify over an individual (using the variable x), and predicate both properties of that individual:

ABBREVIATION 8. $\mathbf{Ex} \equiv \lambda \mathcal{P}_1 \lambda \mathcal{P}_2 \lambda f \lambda g . \exists x\, f[\mathcal{P}_1(x) \mathbf{\ And\ } \mathcal{P}_2(x)]g$

Thus whereas in Heim's original system the LF for "The glass$_7$ broke" would be roughly $glass(7)$ & $broke(7)$, in our variant of Heim's system the at-issue part of the translation to PFL will (after reduction of the type theoretic expressions) amount to $\exists x . \mathbf{Labeled}(7)(x) \mathbf{\ And\ } \mathbf{Glass}(x) \mathbf{\ And\ } \mathbf{Broke}(x))$.

We are now ready to turn to Heim's use of Novelty and Familiarity, which, in the Heimian analysis of DPs, involves a constraint on the relationship between the determiner used and the index born by the DP. Specifically, an indefinite DP must bear a novel index, and a definite DP must bear a familiar index. We define novelty and familiarity to be (dynamic) properties of an index, in terms of whether the input context provides the index with a defined value, i.e one other than $\#_e$. Novelty and familiarity are defined as *tests* in the sense of Beaver (2001)

---

[2]$g \geq_i f$ means that g and f agree on all indices other than i, that if f is defined on i then g gives the same value, and that if not then g may map i to any entity. Formally, $g \geq_i f \equiv \forall j(j \neq i \rightarrow f(j) = g(j)) \wedge (i \in \text{DOM}(f) \rightarrow f(i) = g(i)) \wedge i \in \text{DOM}(g)$.

(who adapts the notion from Veltman (1996)), in that when the relevant condition holds, they have no effect on the context, but when the relevant condition fails, they yield undefinedness. Let us write $i \in \text{DOM}(f)$ to mean $f(i) \neq \#_e$. Then:

ABBREVIATION 9. **novel** $\equiv \lambda i \lambda f \lambda g \, . \, \partial(i \notin \text{DOM}(f))$

ABBREVIATION 10. **familiar** $\equiv \lambda i \lambda f \lambda g \, . \, \partial(i \in \text{DOM}(f))$

We saw above that indices are to be treated as labels on the nominal complement of the determiner rather than on the determiner or the DP. This is important, because a core idea in our proposal is that indices in complex DPs provide properties, just like nouns and intersective adjectives. But in restating Heim's analysis, this analysis of indices creates a tension: the determiner needs access to the index, but the index is on a different constituent. As a result, in giving our reformulation of Heim's system, it will be convenient to introduce the articles syncategorematically. In the following definitions, both definite and indefinite DPs combine a condition on the status of the index (as novel or familiar) with a use of the **Ex** operator to link the nominal property to a lambda-abstracted property ($\mathcal{P}$), which can be thought of as the dynamic meaning of the verbal predicate:[3]

TRANSLATION 3. *a* $X_i \rightsquigarrow \lambda \mathcal{P} . \, \textbf{novel}(i) \textbf{ And } \textbf{Ex}(X_i)(\mathcal{P})$, where $X_i \rightsquigarrow X_i$

TRANSLATION 4. *the* $X_i \rightsquigarrow \lambda \mathcal{P} . \, \textbf{familiar}(i) \textbf{ And } \textbf{Ex}(X_i)(\mathcal{P})$, where $X_i \rightsquigarrow X_i$

We assume a rule of Function Application specifying that if $\alpha \rightsquigarrow \alpha'$ and $\beta \rightsquigarrow \beta'$ and α' denotes a function that can be applied to the denotation of β, then an English phrase consisting (only) of α and β translates as $\alpha'(\beta')$. We can now calculate the meaning of *The glass$_7$ broke*, for which the translation into PFL will be $\textbf{familiar}(7) \textbf{ And } \textbf{Ex}(\textbf{Glass}_7)(\textbf{Broke})$.

The glass$_7$ broke
$\rightsquigarrow$ $\quad \textbf{familiar}(7) \textbf{ And } \textbf{Ex}(\textbf{Glass}_7)(\textbf{Broke})$
$\equiv$ $\quad \textbf{familiar}(7) \textbf{ And } \textbf{Ex}(\lambda x \, . \, \lambda f \lambda g \, . \, x = g(7) \wedge g \geq_7 f \wedge \textbf{glass}(x))(\textbf{Broke})$
$\equiv$ $\quad \textbf{familiar}(7) \textbf{ And } \lambda f \lambda g \, . \, \exists x (x = g(7) \wedge g \geq_7 f \wedge \textbf{glass}(x) \wedge \textbf{broke}(x))$
$\equiv$ $\quad \lambda f \lambda g \, . \, \partial(7 \in \text{DOM}(f)) \wedge \exists x (x = g(7) \wedge g \geq_7 f \wedge \textbf{glass}(x) \wedge \textbf{broke}(x))$
$\equiv$ $\quad \lambda f \lambda g \, . \, \partial(7 \in \text{DOM}(f)) \wedge f = g \wedge \exists x (x = f(7) \wedge \textbf{glass}(x) \wedge \textbf{broke}(x))$
$\equiv$ $\quad \lambda f \lambda g \, . \, \partial(7 \in \text{DOM}(f)) \wedge f = g \wedge \textbf{glass}(f(7)) \wedge \textbf{broke}(f(7))$

Thus *The glass$_7$ broke* will be defined on any sequence that provides a value to the index 7. Relative to such a sequence, it returns the original sequence if 7 is mapped to a smiling glass, and returns no sequence otherwise.

Let a *file* be a set of pairs of worlds and sequences in which all sequences are defined on the same labels. Now in Heim's system, such files are used within the compositional semantics, so that the Heimian meaning of a sentence is a function from files to files. However, we have simplified the compositional semantics such that sentential meanings have an essentially lower type, the type of relations between sequences. Suppose that natural language sentence S has an LF that translates to expression S' in PFL of type τ. Then we can define the Heimian notion of update as follows:

DEFINITION 1: ACCEPTANCE. *F accepts S* iff for every pair $\langle w, f \rangle \in F$, there is a g such that $f[\![S']\!]^w g \neq \#_t$

---

[3]Note that a stronger presupposition could easily be given for *the*, whereby it is presupposed not only that there is a familiar referent, but also that the nominal property was familiar. This would be achieved with the following alternative translation: *the* $X_i \rightsquigarrow \lambda \mathcal{P} . \, \textbf{familiar}(i) \textbf{ And } \partial_d(\textbf{Ex}(X_i)(X_i)) \textbf{ And } \textbf{Ex}(X_i)(\mathcal{P})$. For example, *the glass$_i$* would presuppose that i was familiar and that something was a glass identical with i. Heim (1983) has only the weaker familiarity condition.

DEFINITION 2: UPDATE. $F + S$ is defined iff F *accepts* S, in which case
$$F + S = \{\langle w, g \rangle \mid \exists f \, \langle w, f \rangle \in F \text{ and } f[\![S']\!]^w g = \text{T}\}$$

The translation of *The glass$_7$ broke* into PFL will be accepted in any file whose sequences are defined on 7, and no others; this is what the familiarity presupposition amounts to. The update will remove world-sequence pairs in which 7 is not a glass or did not break.

This completes the core of a re-implementation of Heim (1982, 1983). Note that the system can straightforwardly be applied to cases of donkey anaphora using a dynamic translation of *if-then*, for example as the dynamic equivalent of a material conditional:

TRANSLATION 5. If X then Y $\rightsquigarrow$ **Not**(X' **And Not** Y'), where X$\rightsquigarrow X'$ and Y$\rightsquigarrow Y'$

The system will then produce an appropriate update for e.g. "If a farmer$_1$ owns a donkey$_2$, then the farmer$_1$ beats the donkey$_2$". Specifically, a file updated with this sentence will contain no worlds in which a farmer owns a donkey and fails to beat it.

In the next section we reconsider the semantics of definites and indefinites, and show how a variant of Heim's novelty/familiarity condition can be derived pragmatically from a system in which definites encode uniqueness, and there is no stipulation regarding familiarity or novelty. Crucially, again, familiarity in case of semantic uniqueness will not be required, and the uniqueness requirement will effectively drop away in case of familiarity.

4 Deriving familiarity from uniqueness

We now switch to the analysis of definite and indefinite descriptions in Coppock & Beaver (2015), an analysis that differs from Heim's in two crucial respects. First, the difference between the meaning of *the* and *a* does not involve novelty and familiarity, but rather involves what Coppock & Beaver (2015) term 'weak uniqueness', i.e. a presupposition that there is at most one entity in the extension of the complement of *the*. Second, the basic denotation of both definite and indefinite descriptions is as properties, so that for example *a table* denotes the property of being a table, and *the best friend you could ever ask for* denotes the property that such friends have.

Taking definite and indefinite descriptions to be property-denoting has the immediate benefit of providing a straightforward analyis for predicative uses of DPs, e.g. in *Mary is the best friend you could ever ask for*, but creates a problem when DPs are used in argument positions. When a property-donoting DP occurs in an argument position, Coppock & Beaver (2015) take the resulting type mismatch to trigger a shift, specifically, one of two shifts of Partee (1986). We maintain the same intuitions here, but will adapt the shifts to a dynamic framework. One of these two dynamic shifts is the **Ex** shift introduced above. The other shift, **Iota**, has the same effect on types, but carries an additional presupposition that there exists exactly one object with the property given by the description. To achieve this, we first define a static operator **one** which holds of a property if there is exactly one object in its extension, use this to define a dynamic variant **One** which provides an update just in case it is applied to a dynamic property that holds of exactly one individual, and then build that dynamic notion into the definition of **Iota**:

ABBREVIATION 11. **one** $\equiv \lambda P \,.\, \exists x \,.\, P(x) \wedge \neg \exists y \neq x \,.\, P(y)$

ABBREVIATION 12. **One** $\equiv \lambda \mathcal{P} \lambda f \lambda g \,.\, f = g \wedge \exists h \,.\, \mathbf{one}(\lambda x \,.\, f[\mathcal{P}(x)]h)$

ABBREVIATION 13. **Iota** $\equiv \lambda \mathcal{P}_1 \lambda \mathcal{P}_2 \,.\, \partial_d(\mathbf{One}(\mathcal{P}_1)) \, \mathbf{And} \, \mathbf{Ex}(P_1)(P_2)$

Following Coppock & Beaver (2015), we take the indefinite article to have only a trivial meaning, an identity operation on properties, except that these properties are now dynamic:

TRANSLATION 6. *a(n)* $\rightsquigarrow \lambda \mathcal{P} . \mathcal{P}$

The definite article must also denote a function from dynamic properties to dynamic properties. Let us assume that its meaning is captured by an abbreviation **The**:

TRANSLATION 7. *the* $\rightsquigarrow$ **The**

To specify how **The** is to be interpreted, we need to define weak uniqueness, which we again do in terms of a static predicate (**unique**) from which a dynamic variant (**Unique**) is defined.

ABBREVIATION 14. $\mathbf{unique} \equiv \lambda P . \neg \exists x \exists y \neq x . P(x) \wedge P(y)$

ABBREVIATION 15. $\mathbf{Unique} \equiv \lambda \mathcal{P} \lambda f \lambda g . f = g \wedge \exists h . \mathbf{unique}(\lambda x . f[\mathcal{P}(x)]h)$

We can now give the interpretation of the definite article, presupposing weak uniqueness:

ABBREVIATION 16. $\mathbf{The} \equiv \lambda \mathcal{P} \lambda x . \partial_d(\mathbf{Unique}(\mathcal{P})) \mathbf{And} \mathcal{P}(x)$

Thus, for example, *the glass* gets the meaning **The**(**Glass**), equivalent to the following dynamic property:

$$\lambda x . \partial_d(\mathbf{Unique}(\mathbf{Glass})) \mathbf{And} \mathbf{Glass}(x)$$

Similarly, for the labeled DP *the glass$_i$* we have the translation **The**(**Glass**$_i$), which is equivalent to:

$$\lambda x . \partial_d(\mathbf{Unique}(\lambda y . \mathbf{Labeled}(i)(y) \mathbf{And} \mathbf{Glass}(y))) \mathbf{And} \mathbf{Labeled}(i)(x) \mathbf{And} \mathbf{Glass}(x)$$

Thus, *the glass$_i$* presupposes that there is at most one entity which has both the property of being identical to whatever is labeled i, and the property of being glass.

Used in argument position, *the glass$_i$* will be translated either as **Iota**(**The**(**Glass**$_i$)) or **Ex**(**The**(**Glass**$_i$)). Two principles determine which is used. The first is essentially that of e.g. Heim (1991); Schlenker (2011); Percus (2006), while the second is a variant of the Principle of Informativeness of Atlas & Levinson (1981) and the Strongest Meaning Hypothesis of Dalrymple et al. (1998):

Maximize presupposition (production principle) Relative to a file F, if two sentences are identical except for one item which could take values A or B differing only in that A has stronger presuppostions than B, and if F accepts both versions, then prefer A.

Maximize presupposition (comprehension principle) Suppose two LFs for a sentence are identical except for one item which could take values A or B differing only in that A has stronger presuppostions than B. Then, in the absence of evidence to the contrary, prefer the interpretation with A, if necessary removing world-sequence pairs from the input file F which would not be compatible with that interpretation.

The production variant can be illustrated using semantically unique descriptions. Consider (3) and (4), in which the descriptions *only way* and *tallest mountain* are semantically unique:

(3)   {The/#an} only way is up.

(4)   {The/#a} tallest mountain is Everest.

Given that the descriptions in these sentences are semantically unique, it follows that in any context of interpretation, the uniqueness presupposition of the definite article would be satisfied. Since *the* has strictly stronger presuppositions than *a*, and since those presuppositions are guaranteed to be satisfied, the production principle then predicts that the weaker variant, the

indefinite, will be blocked by the stronger variant, and hence infelicitous. Quite generally, the principle predicts infelicity of indefinite articles with semantically unique descriptions.

Descriptions that are unique with respect to a background file are also prevented by the production variant of Maximize Presupposition from co-occuring with indefinite articles. This fact lets us derive novelty for indefinite descriptions, because familiarity implies uniqueness. More specifically, if i is familiar, then $glass_i$ will be unique. So the speaker should use a definite article in combination with $glass_i$ in that case. Indefinite, labeled descriptions will only be licenced when the label is novel, because only then can the description be non-unique.

Note that the production variant of Maximize Presupposition also predicts that indefinite descriptions should never be interpreted with **Iota**. If the speaker presupposes uniqueness, as he should if he or she intends an **Iota** interpretation, then the definite article should be used.

The comprehension variant of Maximize Presupposition implies that **Iota** is usually chosen over **Ex** in the case of definite descriptions, even if it requires accommodation, but not always – not if there is suffcient "evidence to the contrary", as discussed by Coppock & Beaver (2015). For a case like *Jane didn't score the only goal*, mentioned in the introduction, Coppock and Beaver assume that placement of focus on *only* is sufficient "evidence to the contrary" to rule out an **Iota** interpretation, because the salient focus-alternative to *only* is 'multiple', so the common ground must allow for multiple satisfiers of the predicate, which in turn means that there is *no* satisfier of the predicate 'only goal'. This goes against **Iota**'s existence presupposition, so **Ex** is the only option. In the present setting, the reasoning is somewhat more complicated, because *the only goal* may in principle carry an index. But the same results obtain; if the index is familiar, then there can only be one satisfier of the predicate, and this clashes with focus on *only*. If there is focus on *only*, then any index must be novel. With a novel index, this example behaves as it does in the absence of an index. (The failure to license anaphora observed in the introduction can be attributed in that case to the presence of negation, which caps the 'life-span' of any discourse referents in its scope, in Karttunen's (1973) terminology.)

Now let us consider what presuppositions arise under the **Iota** interpretation of a definite description like *the glass_i*. We have the following equivalences (where some steps in the derivation are left as an exercise for the reader):

Iota(**The**(**Glass**$_i$))(**Broke**)
$\equiv \partial_d(\textbf{One}(\textbf{The}(\textbf{Glass}_i)))$ **And** **Ex**(**The**(**Glass**$_i$))(**Broke**)
$\equiv \partial_d(\textbf{One}(\textbf{Glass}_i))$ **And** **Ex**(**The**(**Glass**$_i$))(**Broke**)
$\equiv \lambda f \lambda g . \partial_s(\textbf{one}(\lambda x . \exists h . h \geq_i f \wedge x = h(i) \wedge \textbf{glass}(x))) \wedge f = g \wedge \textbf{glass}(f(i)) \wedge \textbf{broke}(f(i))$

To consider when the presuppositions are met, recall the definition of acceptance: *F accepts S* iff for every pair $\langle w, f \rangle \in F$, there is a g such that $f[\![S']\!]^w g \neq \#_t$. It turns out that F will accept *The glass_i broke* on an **Iota** interpretation either if i is a familiar glass (regardless of how many other glasses there are in the worlds under consideration) or if every world has exactly one glass. To put it more simply, the presuppositions are met either if i is a familiar glass or if it is in the common ground that there is exactly one glass.

Let us see why this is so. First, suppose i is familiar in file F. This means that for all pairs $\langle w, f \rangle \in F$, i is in the domain of f. Any extension h of f will map i onto the same individual. So there is only one individual that can satisfy the property in the scope of **one**, regardless of how many glasses there are. Since the constraint inside the ∂_s operator in the last line of the derivation above is satisfied for all sequences f in the input file, it follows that for every pair $\langle w, f \rangle$ in F, there will be some g for which the update is defined; *F accepts* the sentence, in the technical sense.

Suppose on the other hand that i is not familiar. If there are worlds in the common ground where there are multiple glasses, then there are multiple individuals that an extension h of f could map i to. Therefore, the property in the scope of **one** will not be unique, and the presupposition will not be satisfied. Put more technically, there will be some pairs $\langle w, f \rangle \in F$ for which the constraint inside the ∂_s operator is not satisfied, and this will yield undefinedness. Therefore, F does not accept the sentence under such conditions. On the other hand, if there is exactly one glass in every world in F, then the update will be defined, even if the index is new. Thus only if the unlabeled property is presupposed to be unique can the presuppositions be satisfied when the discourse referent is new.

Note that we get similar presuppositions under the **Ex** interpretation. The only difference is that **one** is replaced by **unique**, so the presupposition may be satisfied even if there are no satisfiers of the predicate when the discourse referent is novel. But existence of a satisfier is part of the at-issue meaning, and may therefore be the target of negation.

5   Summary

From a uniqueness-based theory of definites, supplemented with a mechanism for interpreting indices on descriptions, we have derived novelty for labeled indefinite descriptions, and familiarity or semantic uniqueness for definite descriptions. The blocking principle Maximize Presupposition implies that indefinite descriptions must be novel, because if they are familiar, then they are unique, and a unique description should be accompanied be a definite article. The uniqueness requirement means that labeled definite descriptions must be familiar if they are not semantically unique, and effectively disappears in case the description is familiar, in the sense that the common ground may allow many satisfiers of the unlabeled predicate.

References

Atlas, Jay David & Stephen C Levinson. 1981. It-clefts, informativeness and logical form: radical pragmatics (revised standard version). In *Radical pragmatics*, 1–62. Academic Press.

Beaver, David. 2001. *Presupposition and assertion in dynamic semantics*. Stanford: CSLI Publications.

Beaver, David & Emiel Krahmer. 2001. A partial account of presupposition projection. *Journal of Logic, Language and Information* 10. 147–182.

Coppock, Elizabeth & David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* 38(5). 377–435.

Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo & Stanley Peters. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21(2). 159–210.

Groenendijk, Jeroen & Martin Stokhof. 1989. Type-shifting rules and the semantics of interrogatives. In Barbara H. Partee & Raymond Turner (eds.), *Properties, types and meanings: Semantic issues*, vol. 2, 21–68. Kluwer.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*: U. Mass Amherst dissertation.

Heim, Irene. 1983. File change semantics and the familiarity theory of definiteness. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use, and the interpretation of language*, 164–189. Berlin: Walter de Gruyter.

Heim, Irene. 1991. Artikel und Definitheit. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales handbuch der zeitgenössischen forschung*, 487–535. Berlin: Mouton de Gruyter.

Horn, Laurence R. & Barbara Abbott. 2013. <*the, a*>: (in)definiteness and implicature. In William P. Kabasenche, Michael O'Rourke & Matthew H. Slater (eds.), *Reference and referring*, 325–355. MIT Press.

Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4(2). 169–193.

Ludlow, Peter & Gabriel Segal. 2004. On a unitary semantical analysis for definite and indefinite descriptions. In Maria Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, Oxford: Clarendon Press.

Partee, Barbara H. 1986. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh & Martin Stokhof (eds.), *Studies in Discourse Representation Theory and the theory of generalized quantifiers*, 115–143. Dordrecht: Foris.

Percus, Orin. 2006. Antipresuppositions. In Ayumi Ueyama (ed.), *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science*, 52–73. Japan Society for the Promotion of Science.

Schlenker, Philippe. 2011. Maximize Presupposition and Gricean reasoning. Ms., UCLA and Institut Jean-Nicod.

Szabó, Zoltán Gendler. 2000. Descriptions and uniqueness. *Philosophical Studies* 101. 29–57.

Veltman, Frank. 1996. Defaults in update semantics. *Journal of philosophical logic* 25(3). 221–261.

Totally tall sounds *totally* younger. From meaning composition to social perception.*

Andrea Beltrama

University of Chicago, Chicago, Illinois, USA
andremormora@uchicago.edu

Abstract

Research on meaning typically separates *semantic* meaning − the linguistic content conventionally associated with an expression − from *social* meaning − the social qualities and attributes that expressions convey about their users. This divide, however, has recently been questioned by work pointing to a principled connection between these domains ([1], [9]). Using the intensifier *totally* as a test case, I extend this program by showing that, in contexts where only a speaker-oriented reading is possible, the social meaning of the intensifier − measured in terms of Age, Gender, Solidarity and Status perception − is more prominent than in those where a lexically compositional interpretation is licensed. I explain this difference in terms of the interactional pragmatic effects that speaker-oriented *totally* indexes, which are conversely missing for the lexical reading. These results reveal a principled connection between the two dimensions of meaning, meanwhile pointing to social perception as a novel methodology for research in experimental semantics.

1   Introduction

Linguistic expressions carry two distinct kinds of content. On the one hand, they convey a *semantic* meaning, with which they are conventionally associated in the grammar of a language. On the other hand, they convey a *social* meaning (e.g., [7]), that is, a package of typified socio-psychological qualities about the identity of language users, including demographic categories (e.g., gender, age) and more specific *social types/personae* (e.g., "Jocks", "Burnouts", "Yuppies" and similar. See [2], [19] for further discussion). These varieties of content, despite the common label, are seen ultimately as pertaining to independent domains. Recognizing their different status, however, does not entail that the two are inherently disjointed. In particular, social meaning, despite its contingent nature, has been shown to be highly systematic and readily available to listeners' cognition (e.g., [21], [6]). As such, a question remains open: do these two types of meaning interact with one another to determine what an expression "says" when used in communication?

To address this question, it becomes crucial to focus on cases in which social meaning attaches to expressions that *do* have non-trivial semantic and pragmatic content. [17] in particular argues that, for non-phonological kinds of variation, the relationship between linguistic forms and social meaning is crucially mediated by the pragmatic effects of the expression. For example, she claims that the association of command imperatives with male gender in American English is grounded in the activity of ordering pragmatically indexed by the form, which in turn becomes associated with a typical affective disposition of men. More recently, [1] build on

---

[12]'s observation that demonstratives like *this* and *that* index a sense of "emotional closeness between speaker and hearer" (p. 351), anchoring these social effects to the presupposition that the addressee must be able to consider the speaker's relation to the NP referent in the discourse context. [9] argues that *need to*, in comparison to *have to/got to*, encodes an obligation directed the hearer's well being, thus indexing an additional component of care/presumptuousness.

Taken together, these results provide evidence that social and semantic content are connected in a principled fashion. However, as the investigation of this area has just begun, a number of issues remain open. The present paper aims to cast light on the following: do the effects of semantic properties extend to attributes of social meaning that transcend the here-and-now of the interactional context and involve more durable categories about the social identity of the speaker?

2 *Totally*: a promising case study

The intensifier[1] *totally* features considerable empirical richness in terms of both its semantic and social content, thus emerging as a ripe testbed to investigate possible answers to the question posed above. On the social front, native speakers agree that this expression conveys a flavor of marked informality and reduced social distance, in that it suggests that the interlocutors are close to one another, share a set of norms or values and easily agree on the content of the conversation. Asides from these effects, the intensifier additionally conveys a set of social attributes about the social identity of its typical users, which track macro-social categories - e.g., *young* and *female* - as well as more specific personae and social types - e.g., "Valley Girl", "Cheerleader".[2] *Totally* likewise presents a rich empirical picture on the semantic and pragmatic front. On a general level, the intensifier combines with a bounded scale and requires that the scalar maximum on such a scale be reached. It is precisely through the way in which this scale is supplied that variation enters the picture. In standard cases, the scale is provided by the subsequent predicate, as in (1): both *full* and *agree* come with a bounded ordering hardwired in their lexical meaning, providing *totally* with an argument to operate on. I refer to these as instantiations of *lexical totally*. In other cases (as in (2)), though, *totally* combines with predicates that do *not* supply a scale, operating on the *commitment* that the speaker has towards the proposition ([14], [10]). I refer to these cases as *speaker-oriented totally*.[3]

(1) a. The bus is *totally* full.   Lexical

     b. She *totally* agrees with me.   Lexical

(2) a. You should *totally* click on that link! It's awesome.[4]   Speaker-oriented

     b. Man in "I have drugs" shirt *totally* had drugs.[5]   Speaker-oriented

Despite sharing reference to maximality, the speaker-oriented usage of *totally* is empirically distinct from the lexical one. First, because it does not combine with a lexical scale, it cannot be replaced by modifiers like *partially* and *almost* (in (3a)). Second, it contributes its meaning

---

[1]For a discussion of the sociolinguistic behavior and distribution of intensifiers, see [22]

[2]Entries from the website Urban Dictionary (http://www.urbandictionary.com/define.php?term=totally) include labels such as "ditzy young girls", "Valley Girl Speak", "girly girls, poppers, and rich spoiled little brats".

[3]The OED added a dedicated entry to this flavor of *totally* in 2005: "In weakened use, as an intensifier: (modifying an adjective) very, extremely; (modifying a verb) definitely, absolutely."

[4]https://www.facebook.com/TheBiscuitGames/posts/488916347870627 accessed on June 5th 2015

[5]http://www.miaminewtimes.com/news/wtf-florida-man-in-i-have-drugs-shirt-totally-had-drugs-6542858

at the non at-issue level, as shown by the fact that it cannot be challenged independently from the rest of the propositional content.[6]

(3) a. # You should *partially/almost* click on that link! It's awesome.
      b. She should *totally* click on that link!
         B: # **No!** She should click on that link, but you're not committed to saying that!

As far as the exact nature of the contribution of *totally*, the intensifier operates as a conversational operator, emphasizing the speaker's intention to make the proposition become shared knowledge for all the interlocutors, thus enriching the Common Ground of the conversation (see [3] for a full formal analysis). As a result, while primarily representing the speaker's perspective, *totally* carries implications for the hearer's position as well. This makes it crucially different from seemingly similar operators like *definitely* and *certainly*, which are instead grounded in private, individual certainty of the speaker towards the truth of the proposition. While subtle, this difference emerges in particular environments. In the following exchange, for example, while expressing individual certainty sounds somewhat deviant, *totally* can be used to dispel the doubt that the interlocutor expressed.

(4) Mark: I can't remember if your name is Emily.
      a. Emily: Oh, yes, it's *totally* Emily!
      b. Emily: #Oh, yes. It's *definitely/certainly* Emily.

That the commitment intensified by *totally* crucially bears on the hearer emerges even more clearly in contexts in which the intensifier is used out of the blue. Let us consider (2b) again. On the one hand, the presence of *totally*, by pushing for the addition of the proposition to the Common Ground, signals the shareworthiness of the event described (e.g., it is particularly outlandish or funny). On the other hand, without a previous discourse move introducing uncertainty, pure epistemic operators like *certainly* and *definitely* sound bizarre.

(5) Man in "I have drugs" shirt {*totally*/#*definitely/certainly*} had drugs.

Finally, an intermediate variety of *totally* occurs with *extreme* adjectives ([16]) − e.g.,*awesome*, *amazing* − which do *not* lexicalize a bounded scale but refer to properties to an inherently high degree, making it easier to coerce their open scale into a bounded one ([18]). Here, *totally* appears to be less deviant than pure speaker-oriented *totally* with respect to the diagnostics discussed above.[7]

(6) a. Bob is totally awesome
      b. ?# Bob is {**not** *totally*/**almost** *totally*/**completely**/**entirely**} awesome.

3   From semantic to social meaning: hypotheses

Previous work (e.g., [17], [1], [9]) has shown that expressions that affect the social positioning of the interlocutors as they proceed in the exchange are particularly suitable candidates to convey social information about the identity and attributes of language users. The question therefore emerges as to whether, in a parallel fashion, the different variants of *totally* also index a particular kind of alignment and range of interpersonal stances. On the one hand, the lexical

---

[6]From this perspective, it shows compositional properties similar to other expressions that specify the attitude of the speaker such as expressives ([20]) and certain evidentials ([8])

[7]The symbol ?# indicates a minor degree of deviance in comparison to cases marked with #.

variant hardly features any cue of this sort: by maximizing the degree to which a property applies, it serves a bare informational function. On the other hand, the pragmatic move associated with speaker-oriented *totally* contributes to highlight the conversation as a joint activity, underscoring the engagement of the speaker with establishing a stock of shared knowledge with the hearer. In addition, when used to strengthen commitment to propositions uttered out of the blue (see (5) above), *totally* also presupposes a shared evaluative stance between the interlocutors, requiring that they converge on the reason that makes the proposition shareworthy (e.g., its absurdity.) I hypothesize that, in light of its implications on the positioning of the interlocutors, speaker-oriented *totally* emerges as a more suitable linguistic resource to convey the social identity of its users than lexical *totally*. Moreover, if this hypothesis is confirmed, I hypothesize that, with extreme adjectives, the social meaning of *totally* should have intermediate salience between the lexical and the speaker-oriented use, mirroring the gradience emerging in the semantics.

4 The experiment

4.1 Methods

I test these hypotheses via a social perception experiment. This methodology, which has long been used to investigate language attitudes ([13]; see [4] for a literature review), is based on the assumpion that social evaluation is a proxy into the social meaning conveyed by an expression, as it allows us to measure fine-grained changes in social meaning in relation to a manipulated factor. In the first step, I conducted a study to construct the evaluation scales to be used in the perception experiment. The study was designed with the software Qualtrics and subsequently circulated on Amazon Mechanical Turk. 60 subjects, self-declared speakers of American English and between 18 and 35 years old, were paid $0.50 for participating. The subjects were asked to provide a series of adjectives to describe the social identity of the speaker of sentences containing *totally*. Based on the most recurring adjectives, eight evaluative dimensions were selected. Four of these dimensions, which I label Solidarity attributes, are predicted to be positively affected by the presence of *totally*. They include *Frendliness, Coolness, Outgoingness, Excitability*. The other four dimensions, which I label Status attributes, express identity categories that should be negatively affected by the presence of the intensifier and include *Articulateness, Maturity, Intelligence, Seriousness*.

4.1.1 Stimuli

2 factors were crossed in a 3x4 design. The first factor manipulates the semantic variant of *totally* along the lexical vs speaker-oriented axis of variation. To cue lexical *totally*, the intensifier was used next to bounded adjectives (or *absolute adjectives*, [11]), which lexicalize a bounded scale as part of their lexical meaning (e.g., "bald"). To cue the speaker-oriented reading, it was used next to unbounded adjectives (i.e., *relative*, e.g., "tall"), which offer a commitment scale as the only possible target. In addition, *extreme adjectives* (e.g., *awesome, amazing*) were used as an intermediate case between the two other categories. Following the hypotheses outlined above, I predict the following effects on the social perception of *totally*.

	Adjective type	Example	Semantic type	Likelihood to convey social meaning
(7)	Bounded	Bald	Lexical	Low
	Extreme	Awesome	Intermediate	Medium
	Unbounded	Tall	Sp-oriented	High

An additional factor, the type of modifier accompanying the adjective came in four different conditions: the target intensifier, *totally*; two control intensifiers: *really* and *completely*; and the positive, non-intensified form. On the one hand, *completely*, contrary to *totally*, is exclusively able to target lexical scales, resulting in ungrammaticality when used with a open-scale adjective. On the other hand, *really* is not sensitive to scale structure ([15], [5]): since all the adjectives used in the experiment are scalar, the intensifier should always operate at the lexical level. Accordingly, I predict that, if an effect of semantic type of *totally* is observed on social meaning, the same effect should not be observed on the two control intensifiers. Finally, the positive form serves as a baseline condition to isolate the social meaning of the intensifier from the one contributed by the other elements in the sentence, as discussed below. 12 items, each with a different set of adjectives, were crossed in a Latin Square Design.

4.1.2 Procedure

Every subject saw a total of 12 written sentences, one sentence for each condition. Each sentence was followed by ten questions. The first two questions are targeted at the demographic characteristics of the speakers and provide fixed alternatives as possible responses.[8] The other eight questions are aimed at assessing Solidarity-based and Status-based traits of social meaning. They were presented in the form of a 1-6 Likert scale, where 1 indicated the minimum value and 6 the maximum value. The study was created with Qualtrics and carried out online. 36 self-declared native speakers of American English, aged 18-35, were recruited on Amazon Mechanical Turk and compensated $2 for their participation.

4.2 Results

To conduct the analysis, the social evaluation of the sentence with the positive form of the adjective was subtracted from the evaluation of the sentence with either *totally*, *completely* or *really*. This made it possible to filter out the effects on the social meaning independently introduced by other elements, isolating those induced by the intensifier. On the resulting differences, statistical analysis was conducted on the following dimensions: Age, Gender, Solidarity attributes and Status attributes, obtaining four separate metrics for assessing the social meaning contributed by each intensifier. For each dimension, a two-way ANOVA was carried out to verify the effect of the two factors (Scale type and Intensifier type). In the event that a higher-level effect was present, planned paired t-test were carried out to verify the effect of Scale on each intensifier.

4.2.1 Age

Age was first converted from a categorial to a continuous variable.[9] A main effect of Adjective (by-subject $F_1(2,70)=18.1$, $p<0.0001$; by-item $F_2(2,22)=8.4$, $p<0.001$) and Intensifier (by-subject $F_1(2,70)=14.5$, $p<0.0001$; by-item $F_2(2,22)=9.4$, $p<0.001$) and an interaction effect

---

[8]For Age: Kid / Teenager / Adult / Elderly. For Gender: Male / Could be Either / Female

[9]Every life stage provided in the multiple choice response was assigned a numerical score with increasing value, where Kid=1, Teenager=2, Adult=3 and Elderly=4. This ensures that the higher the Age score, the higher the perceived age of the speaker.

between Adjective and Intensifier (by-subject $F_1(2,70)=10.4$, $p<0.001$; by-item $F_2(2,22)=11.3$, $p<0.001$) were found. The results are plotted in the graph below.
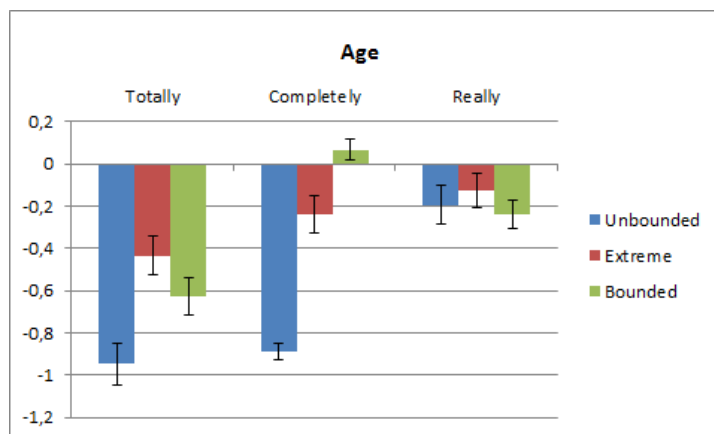
Figure 1: Age perception. The Y-axis indicates the value of the subtraction of the score of each intensifier from the positive form. The X-Axis groups the different intensifiers. The blue bar stands for Unbounded adjectives (e.g. *tall*). The red bar stands for Extreme adjectives (e.g. *awesome*). The green bar stands for Bounded adjectives (e.g. *bald*). Error bars indicate standard errors.

Concerning *totally*, for all adjective types the intensifier caused a lowering of the perceived age of the speaker. However, the effect is significantly stronger with unbounded adjectives than with extreme adjectives ($t(35) = 6.2$, $p < .0001$) or bounded adjectives ($t(35) = 4.1$, $p < .001$). No significant difference was found between Bounded and Extreme adjectives. *Completely* caused an increase in perceived age with bounded adjectives and a decrease with extreme/unbounded ones. No effect was found across adjective type for *really*.

4.2.2 Gender

As with Age, Gender was converted into a continuous variable.[10] An interaction effect was found between Adjective and Intensifier (by-subject $F_1(2,70)=8.2$, $p<0.001$; by-item $F_2(2,22)=7.3$, $p<0.001$). The results are plotted below.

---

[10] "Male" =1, "Could be either" = 2 and "Female" = 3. Hence, the higher the resulting score, the higher the likelihood that the person was perceived to be female.
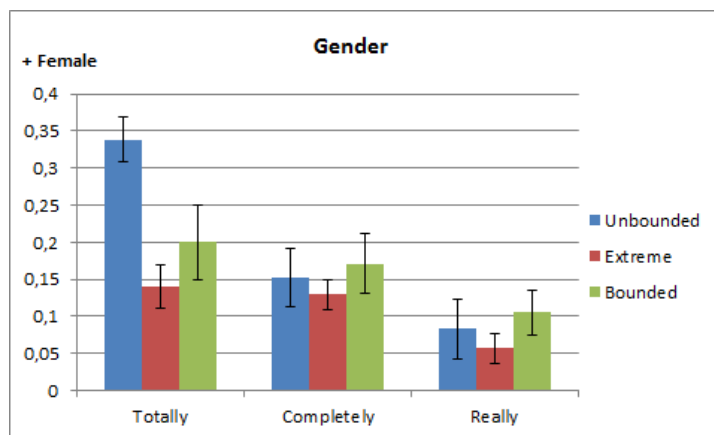
Figure 2: Gender perception.

All intensifiers increased the likelihood with which the speaker was perceived to be female. The effect of *totally* with unbounded adjectives was significantly stronger than with extreme adjectives ($t(35) = 7.2$, $p < .0001$) and bounded adjectives ($t(35) = 5.1$, $p < .001$). No difference was found for the other intensifiers.

4.2.3 Solidarity

Since their effects were highly similar, Solidarity attributes were analyzed together. Main effects of Adjective (by-subject $F_1(2,70)=11.1$, $p<0.001$; by-item $F_2(2,22)=6.7$, $p<0.001$), Intensifier (by-subject $F_1(2,70)=11.8$, $p<0.0001$; by-item $F_2(2,22)=7.5$, $p<0.001$) and an interaction effect between Adjective and Intensifier (by-subject $F_1(2,70)=8.6$, $p<0.001$; by-item $F_2(2,22)=10.3$, $p<0.001$) were found. The results are plotted below.
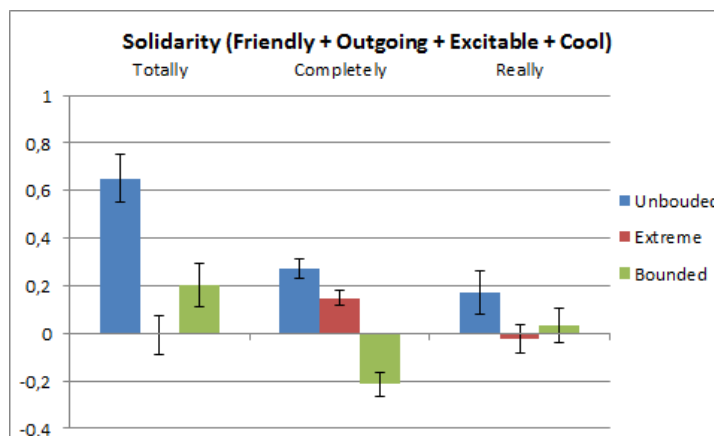
Figure 3: Solidarity perception.

Totally increased the perception of solidarity across the board, but the effect was significantly

stronger with Unbounded than bounded adjectives (t(35) = 7.5, $p < .0001$). No effect of *totally* on extreme adjectives was found. Concerning *completely*, unbounded adjectives record higher Solidarity value than Bounded ones (t(35) = 7.9, $p < .0001$). Concerning *really*, extreme adjectives recording a lower Solidarity value than bounded and unbounded ones (t(35) = 7.4 and t(35) = 6.5 respectively, $ps < .0001$).

4.2.4 Status

The four Status attributes also patterned similarly, and were thus analyzed together. A main effect of Adjective (by-subject $F_1(2,70)=12.1$, $p<0.001$; by-item $F_2(2,22)=6.9$, $p<0.001$) was found, along with a main effect of and Intensifier (by-subject $F_1(2,70)=12.3$, $p<0.0001$; by-item $F_2(2,22)=8.6$, $p<0.001$) and an interaction effect between Adjective and Intensifier (by-subject $F_1(2,70)=9.5$, $p<0.001$; by-item $F_2(2,22)=11.3$, $p<0.001$). The results are plotted below.
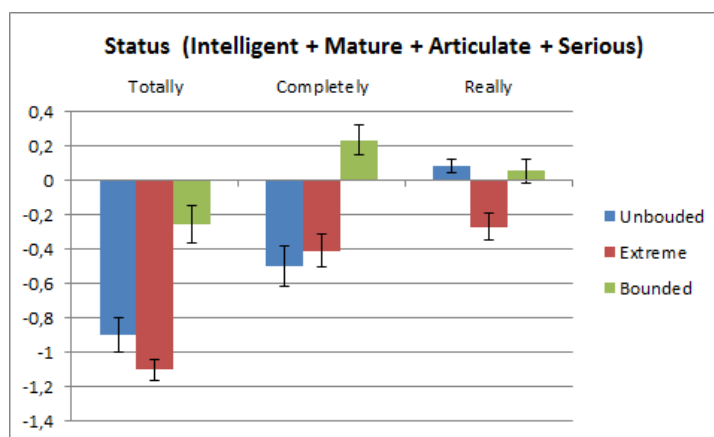
Figure 4: Status perception.

Totally lowered the Status perception across adjective types. Yet, the effect is significantly stronger with unbounded than bounded adjectives (t(35) = 8.1, $p < .0001$). Concerning *completely*, the intensifier with bounded adjectives raises the Status perception, while it lowers it with unbounded adjectives (t(35) = 7.7, $p < .0001$). Concerning *really*, extreme adjectives record a lower Status value than bounded and unbounded ones (t(35) = 7.4 and t(35)=6.5 respectively, $ps < .0001$).

4.3 Discussion

The current study was aimed at investigating how the social perception of *totally* is affected by variations in the semantic properties of the intensifier across different linguistic contexts. Two hypotheses were tested. First, I predicted that, by inviting the hearer to converge on the Common Ground of the conversation and on the evaluation of the anchor proposition, speaker-oriented *totally* would be more likely to convey social meaning than lexical *totally*, which merely modifies the descriptive content of the utterance. The prediction is confirmed: when *totally* occurs next to a unbounded adjective, an environment in which only a speaker-oriented reading is licensed, the speaker is perceived as younger, more likely to be female,

higher in Solidarity and lower in Status than when *totally* occurs next to a bounded adjective and could therefore receive a lexical interpretation. Crucially, the same pattern does not hold for the control intensifiers. As predicted, *really* presents no significant difference across the tested adjective types. Concerning *completely*, an effect of Adjective does emerge for Age, Solidarity and Status. Yet, with the exception of Age, the effect on *totally* is considerably stronger than the one on *completely*, suggesting that the two patterns are most likely not driven by the same source.[11] The second hypothesis study also tested whether the salience of the indexed social meaning reflects the continuum between lexical and speaker-oriented uses, predicting that the social meaning should have intermediate intensity with Extreme Adjectives. This prediction, however, is not borne out, as we observe that for none of the tested dimensions a continuum along these lines emerges.[12]

Taken together, these findings suggest that, when making social evaluations about language users, hearers keep track of the semantic and pragmatic properties of the form, such as the type of scale that the intensifier targets in a given context, pointing to another domain in which semantic and social meaning are connected. To further cast light on this relationship, though, a central issue must be addressed: What is the nature of the process whereby *totally* becomes a vehicle of those specific identity categories? In particular, is the high value of speaker-oriented *totally* along Solidarity-based attributes the result of a reinterpretation at the social level of the speaker-hearer convergence indexed at the pragmatic level? Is it grounded in social knowledge about the types of people that use the expression more frequently? To cast light on the issue, one could compare the social perception of uses of speaker-oriented *totally* that index different degrees of pragmatic inclusiveness between the interlocutors. For example, *totally* in sentences in out-of-the-blue contexts presupposes a shared evaluative stance between the participants that is instead less salient in uses of *totally* with modalized sentences or in responses, and should therefore be an even more salient carrier of social information. I plan to address this question in future research.

5   The broader picture

While representing a preliminary step, the current study opens up a novel area of research on the study of meaning, highlighting the interface between social and semantic content as a ripe, and largely uncharted, domain of investigation. On the theoretical level, this line of research can lead us to adopt a more comprehensive view of linguistic meaning, in which social meaning is seen as a *bona fine* type of content side by side with the logical properties of expressions. On a methodological level, it points to social perception studies as a promising technique to explore the behavioral correlates of semantic and pragmatic features, opening up a novel direction of inquiry for experimental investigations on meaning.

---

[11]I speculate that the effects of scale type on *completely* are grounded in the ungrammaticality of the combination, rather than to the particular semantic properties of the expression. As such, the fact that this expression has an effect on the social meaning can be reasonably predicted to being associated with whatever social features are associated with a "default other" who does not fully master the grammar of English, e.g. a particularly young speaker, or one with a low Status. I thank an anonymous NWAV 44 reviewer for suggesting this explanation.

[12]Concerning Solidarity, *totally* has virtually no effect, leaving the perception unchanged. A possible explanation for this could be that, because extreme adjectives have remarkable emotive charge on their own, they are independently associated with high level of solidarity, thus giving rise to a ceiling effect that neutralizes the impact of the intensifier. Concerning Status, *totally* with Extreme adjectives brings about a particularly pronounced lowering of the perception of the speaker. We suggest that the steep drop associated with *totally* might have to do with the fact that combinations like "totally awesome" are stereotypically associated with low-status social groups (e.g. Valley Girls), thus triggering a negative evaluation of the intensifier.

References

[1] Eric Acton and Chris Potts. That straight talk. sarah palin and the sociolinguistcs of demonstratives. *Journal of Sociolinguistics*, 18(1):3–31, 2014.

[2] Asif Agha. Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, 15:38–59, 2005.

[3] Andrea Beltrama. Totally between discourse and subjectivity. exploring the pragmatic side of intensification. In *Proceedings of CLS 51*, 2015.

[4] Kathryn Campbell-Kibler. Perception in sociolinguistics. *Language and Linguistics Compass.*, 4(6):377–389, 2010.

[5] Camelia Constantinescu. *Gradability in the Nominal Domain*. PhD thesis, Leiden University, 2011.

[6] Annette D'Onofrio. Persona-based information shapes linguistic perception: Valley girls and california vowels. *Journal of Sociolinguistics*, 19(2):241–256, 2015.

[7] Penelope Eckert. *Jocks and Burnouts: Social Identity in the High School*. New York: Teachers College Press., 2008.

[8] Martina Faller. *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. PhD thesis, Stanford University, 2002.

[9] Lelia Glass. Need to vs. have/got to: Four socio-pragmatic corpus studies. In *Selected papers from NWAV 43*, 2015.

[10] Patricia Irwin. So [totally] speaker-oriented: An analysis of "Drama SO". In Raffaella Zanuttini and Laurence R. Horn, editors, *Microsyntactic Variation in North American English.*, pages 29–70. Oxford University Press, Oxford, 2014.

[11] Christopher Kennedy and Louise McNally. Scale structure, degree modification and the semantics of gradable predicates. *Language*, 81(2)(2):345–381, 2005.

[12] Robin Lakoff. Remarks on this and that. In *Proceeding of the Chicago Linguistic Society 10*, 1974.

[13] R.C. Hodgson Lambert, W and, R.C. Gardner, and S. Fillenbaum. Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology*, pages 44–51, 1960.

[14] Eric McCready and Magdalena Kaufmann. Maximum intensity. Paper presented at the Semantics Workshop, Keio University, 29 November, 2013.

[15] Yaron McNabb. Cross-categorial modification of properties in Hebrew and English. In Anca Chereches, editor, *Proceedings of Semantics and Linguistic Theory (SALT) 22*, pages 365–382. eLanguage, 2012.

[16] Marcin Morzycki. Adjectival extremeness: Degree modification and contextually restricted scales. *Natural Language and Linguistic Theory*, 30(2):567–609, 2012.

[17] Elinor Ochs. Indexing gender. In Alessandro Duranti and Charles Goodwin, editors, *Rethinking Context: Language as an Interactive Phenomenon*. New York, Cambridge: Cambridge University Press, 1992.

[18] Carita Paradis. Adjectives and boundedness. *Congnitive Linguistics*, 12:47–65, 2001.

[19] Robert J. Podesva. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, 11.4:478–504, 2007.

[20] Christopher Potts. *The Logic of Conventional Implicature*. Oxford University Press, Oxford, 2005.

[21] E. Allyn Smith, Kathleen Currie Hall, and Benjamin Munson. Bringing semantics to sociophonetics: Social variables and secondary entailments. *Laboratory Phonology*, 1:121–155, 2010.

[22] Sali Tagliamonte. So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics*, 12(2):361–394, 2008.

Scope interactions between
modals and modified numerals[*]

Dominique Blok

Utrecht University
`d.blok@uu.nl`

1   Introduction

It is a widely observed phenomenon in the literature about modified numerals that sentences with an existential modal and *at most* cannot be used to merely give permission to do something (e.g. Geurts & Nouwen, 2007; Nouwen, 2010; Schwarz, 2011; Coppock & Brochhagen, 2013; Penka, 2014; Kennedy, 2015). (1-a), for instance, forbids the hearer from taking more than two biscuits. It cannot be used only to convey that she has permission to take one or two biscuits. When *at most* is replaced by *fewer than*, as in (1-b), we do get this 'pure permission' reading.

(1)   a.   You're allowed to take at most two biscuits.
      b.   You're allowed to take fewer than three biscuits.

This difference is not restricted to *fewer than* and *at most*. It is a general difference that holds between all so-called class A and class B modified numerals that set an upper bound. Class B numeral modifiers are characterised by the fact that they, unlike class A modifiers, give rise to ignorance inferences in unembedded contexts (Nouwen, 2010).[1]

When class B modified numerals are combined with a modal, two readings can arise: the *ignorance reading* and the *authoritative reading*. For example, (1-a) can be uttered when the speaker does not know the maximum number of biscuits the hearer is allowed to take (but she knows it is not more than two) or when the speaker has full knowledge of what is allowed.

The examples in (1) raise the question of why upper-bounded class B modifiers lack the pure permission reading that their class A counterparts do display when they are combined with an existential modal. Most current theories wrongly predict that sentences like (1-a) have the same truth conditions as sentences like (1-b). One notable exception is Penka (2014), who tackles the problem by positing a decomposition account for *at most*. However, as I will discuss in this paper, her account misses an important generalisation. As I aim to show here, class B numeral modifiers obligatorily outscope existential modals. This generalisation holds for upper-bounded

---

[1]To see this, consider (i).

(i)   a.   I know exactly how many books there are in this bookshop, #and it's { at most / maximally / no more than } 10,000.
      b.   I know exactly how many books there are in this bookshop, and it's { fewer than / less than / under } 10,000.

The class B modifiers in the a-sentence are incompatible with exact knowledge of the number of books in the bookshop. The class A modifiers in the b-sentences do not give rise to ignorance inferences and are felicitous in a context in which a speaker claims to have knowledge of the precise number of books.

modifiers like *at most* and *maximally* but also for their lower-bounded counterparts such as *at least* and *minimally*. As will become clear in this paper, this generalisation accounts for a variety of data involving lower-bounded and upper-bounded class B modifiers, root modals, epistemic modals, and islands. It also raises problems for existing theories of modified numerals.

In the following section I briefly discuss one of these existing theories (Schwarz, 2011) and Penka's (2014) modification of the theory. In section 3 I review the data with class B modifiers and existential modals that have led me to propose the generalisation mentioned above. I also explore the consequences of this generalisation for current theories of modified numerals. Section 4 concludes.

2   Previous accounts

2.1   The neo-Gricean approach

Current theories of modified numerals generally derive the ignorance and authoritative readings of sentences with numeral modifiers and modals with an implicature mechanism. I will use the account proposed in Schwarz (2011) to illustrate this. Schwarz assumes the denotations of *at least* and *at most* given in (2).

(2) a. $[\![\text{at least}]\!] = \lambda d_d \lambda P_{\langle d,t \rangle}.\text{MAX}\{n \mid P(n)\} \geq d$
     b. $[\![\text{at most}]\!] = \lambda d_d \lambda P_{\langle d,t \rangle}.\text{MAX}\{n \mid P(n)\} \leq d$

At least and *at most* take a degree of type d and a degree predicate of type $\langle d, t \rangle$ and express that the maximal degree n such that P holds of that degree is at least as high or at most as high as d. Schwarz assumes that there are two Horn sets that are used for the calculation of implicatures: the set of natural numbers; alternatives to the modified number, and a set containing the modifiers *at least*, *exactly*, and *at most*; alternatives to the modifier.

(3) a. { 1, 2, 3, 4, 5, ... }
     b. { at least, exactly, at most }

2.1.1   Lower-bounded modified numerals

Let us first turn to modified numerals that set a lower bound, like *at least*. A sentence with *at least* and a universal modal such as (4) has two possible LFs depending on where the modified numeral takes scope. These are given in (5). The narrow scope reading in (5-a) says that in all permissible worlds, Mary submits one or more abstracts. The wide scope reading in (5-b) says that the maximum number of abstracts such that Mary submits that many abstracts in all permissible worlds is one or higher.

(4) Mary is required to submit at least one abstract.

(5) a. □ [MAX { n | Mary submits n abstracts } ≥ 1]
     b. MAX { n | □ [Mary submits n abstracts] } ≥ 1

The stronger alternatives to (5-a) and (5-b) are given in (6) and (7) respectively.

(6) a. □ [MAX { n | Mary submits n abstracts } = 1]
     b. □ [MAX { n | Mary submits n abstracts } ≥ 2]

(7) a. MAX { n | □ [Mary submits n abstracts] } = 1
     b. MAX { n | □ [Mary submits n abstracts] } ≥ 2

The alternatives in (6) are not symmetric, so the primary implicatures that arise from them can be strengthened to secondary implicatures (Sauerland, 2004). The meaning of the assertion in (5-a) combined with these implicatures is that the speaker believes Mary is required to submit one or more abstracts but she is not required to submit exactly one and she is not required to submit two or more.[2] This reading does not involve any ignorance on the part of the speaker and is referred to as the authoritative reading.

The alternatives in (7), on the other hand, are symmetric, so only primary implicatures can be derived. The primary implicatures together with the assertion generate ignorance implicatures: the speaker believes that Mary is required to submit at least one abstract, but she is not sure whether Mary is required to submit exactly one and she is not sure whether Mary is required to submit at least one. Thus, Schwarz's account derives the authoritative reading when the modified numeral takes narrow scope and the ignorance reading when the modified numeral takes wide scope.

Now let us turn to a similar example with an existential modal. A sentence like (8) is taken to have the scope configurations given in (9). The narrow scope reading in (9-a) merely says that there is a permissible world in which Mary submits one or more abstracts. The wide scope reading in (9-b) conveys that there is an upper bound to the number of abstracts Mary is allowed to submit, and that upper bound is one or higher.

(8)   Mary is allowed to submit at least one abstract.

(9)   a. $\Diamond$ [MAX $\{\ n \mid$ Mary submits n abstracts $\} \geq 1$]
        b.   MAX $\{\ n \mid \Diamond$ [Mary submits n abstracts] $\} \geq 1$

The stronger alternatives to both (9-a) and (9-b), are symmetric, which leads to the ignorance implicatures in (10) for (9-a) and in (11) for (9-b) (ignorance is indicated with a question mark).

(10)   a.   ?$\Diamond$ [MAX $\{\ n \mid$ Mary submits n abstracts $\} = 1$]
        b.   ?$\Diamond$ [MAX $\{\ n \mid$ Mary submits n abstracts $\} \geq 2$]

(11)   a.   ?MAX $\{\ n \mid \Diamond$ [Mary submits n abstracts] $\} = 1$
        b.   ?MAX $\{\ n \mid \Diamond$ [Mary submits n abstracts] $\} \geq 2$

The narrow scope ignorance implicatures in (10) say that the speaker does not know whether submitting one abstract is allowed or whether submitting at least two abstracts is allowed. The wide scope ignorance implicatures in (11) do not indicate ignorance merely about which numbers are allowed, but about the upper bound of the allowed numbers. Thus, the wide scope reading is that the maximum allowed number of pages is at least one, but the speaker does not know if this maximum is exactly one or higher than one. As Schwarz admits, only the wide scope reading and not the narrow scope reading appears to be attested.

2.1.2   Upper-bounded modified numerals

Now let us turn our attention to upper-bounded modified numerals. Parallel to (8), (12) has the two denotations given in (13).[3] The narrow scope reading in (13-a) says that submitting one or two abstracts is allowed without excluding the possibility of submitting more abstracts. This is the unattested pure permission reading. The wide scope reading in (13-b) expresses that the maximum number of abstracts Mary is allowed to submit is two or less.

---

[2]This is the so-called *free choice* reading: Mary can choose freely between submitting one or more than one abstract.

[3]I do not discuss examples with upper-bounded modified numerals and universal modals here as they are not relevant for the current discussion.

(12)   Mary is allowed to submit at most two abstracts.

(13)   a. $\Diamond$ [MAX { n | Mary submits n abstracts } ≤ 2]
        b.   MAX { n | $\Diamond$ [Mary submits n abstracts] } ≤ 2

The stronger alternatives for (13-a) and (13-b) are given in (14) and (15) respectively.

(14)   a. $\Diamond$ [MAX { n | Mary submits n abstracts } $= 2$]
        b. $\Diamond$ [MAX { n | Mary submits n abstracts } ≤ 1]
(15)   a.   MAX { n | $\Diamond$ [Mary submits n abstracts] } $= 2$
        b.   MAX { n | $\Diamond$ [Mary submits n abstracts] } ≤ 1

Again, symmetry arises in both cases. This means that only primary implicatures are derived, which leads to ignorance inferences. Parallel to the *at least* cases, the weak narrow scope ignorance inferences (corresponding to (13-a)) convey that the speaker does not know if two is an allowed number or if one or less is allowed, whereas the stronger wide scope ignorance implicatures (corresponding to (13-b)) are about the upper bound: the speaker does not know if the maximum number of abstracts Mary is allowed to submit is two or lower than two.

There are two problems here. The first is that, as mentioned in the introduction of this paper, the weak narrow scope reading that this account derives is not attested. The second is that there is no way to derive the correct authoritative reading. This account has no hope of deriving any kind of an authoritative reading for (12) because existential modals lead to symmetric alternatives regardless of where they take scope, and this blocks secondary implicatures. As we have seen, primary implicatures lead to ignorance inferences, so only ignorance readings can be obtained. The chances of deriving the particular authoritative reading that is attested for (12) are even lower. As the neo-Gricean accounts derive ignorance readings when the modified numeral takes wide scope and authoritative readings when the modified numeral takes narrow scope, *at most* would have to scope under the existential modal in the authoritative LF. This kind of a scope configuration can never lead to a strong upper bound because it merely states that the upper bound is there in one permissible world and remains silent about other permissible worlds. In the next section, I show how Penka (2014) addresses these issues.

2.2   Penka's solution

Penka (2014) posits an account that derives the missing authoritative reading for sentences with *at most* and an existential modal. Her proposal is to decompose *at most* into an antonymising operator ANT and *at least*, as in (16).

(16)   [at most n] = [[n ANT] at least]

At least is defined as in (2-a), repeated here as (17-a), and ANT is defined as in (17-b).[4]

(17)   a. $[\![$at least$]\!] = \lambda d_d \lambda P_{\langle d,t \rangle}.\text{MAX}\{n \mid P(n)\} \geq d$
        b. $[\![\text{ANT}]\!] = \lambda d_d \lambda P_{\langle d,t \rangle}.\forall d' : d' > d \rightarrow \neg D(d')$

In this decomposition account, a sentence like (12) has the three LFs in (18).

(18)   a.   [allowed [ANT 2 abstracts [λd [at least d [$\lambda d'$ [Mary submits d' abstracts]]]]]]
        b.   [ANT 2 abstracts [λd [at least d [$\lambda d'$ [allowed [Mary submits d' abstracts]]]]]]
        c.   [ANT 2 abstracts [λd [allowed [at least d [$\lambda d'$ [Mary submits d' abstracts]]]]]]

---

[4]The ANT operator is thus equivalent to *at most* as defined in (2-b).

(18-a) and (18-b) have the same truth conditions as (13-a) and (13-b) respectively, but (18-c), where the modal takes scope between the antonymising operator and *at least*, is an LF that is not available in the orignial neo-Gricean account. This LF has the truth conditions in (19).

(19) $\forall d' : d' > 2 \rightarrow \neg\Diamond$ [MAX$\{ d \mid$ Mary submits d abstracts $\} \geq d'$]
         $= \neg\Diamond$ [MAX $\{d \mid$ Mary submits d abstracts $\} > 2$]

These are the desired strong truth conditions for the authoritative reading: it is not allowed for Mary to submit more than two abstracts. For the calculation of the implicatures, Penka assumes the same Horn sets as Schwarz along with the two Horn sets given in (20).

(20)   a.   { ANT , $\varnothing$ }
         b.   { $\Diamond$, $\square$ }

The stronger scalar alternatives derived using these four Horn sets are the ones in (21).

(21)   a. $\neg\Diamond$ [MAX $\{d \mid$ Mary submits d abstracts $\} > 1$]
         b. $\square$ [MAX $\{d \mid$ Mary submits d abstracts $\} = 2$][5]

These alternatives are not symmetric, so the secondary implicatures in (22) are derived. The implicatures are that the speaker believes Mary is allowed to submit more than one abstract and she is not required to submit exactly two. Leaving aside the question of whether these specific implicatures are empirically correct, the derived reading is clearly not an ignorance reading but an authoritative reading that sets a strong upper bound, as desired.

(22)   a.   B $\Diamond$ [MAX $\{d \mid$ Mary submits d abstracts $\} > 1$]
         b.   B $\neg\square$ [MAX $\{d \mid$ Mary submits d abstracts $\} = 2$]

Although this analysis solves one of the issues the neo-Gricean account suffers from, some problems remain. In the following section I will discuss these problems and argue that they stem from a common core: the fact that class B modifiers always outscope existential modals.

3   A new generalisation

3.1   The scopal behaviour of class B modifiers

At the end of section 2.1 I mentioned that the neo-Gricean account has two problems with examples with existential modals and *at most*: it yields an unattested weak ignorance reading (when the modal takes wide scope) and it does not generate a strong authority reading. While Penka solves the second issue, her account inherits the first one. Penka's LF in (18-a) corresponds to a weak reading that merely says that submitting zero, one, or two abstracts is allowed.

Kennedy (2015), whose account has the same problem, argues that the strong upper bound of the authoritative reading is a scalar implicature. In other words, (23) has the weak truth conditions that the neo-Gricean account predicts and the upper bound is calculated pragmatically because the speaker chose not to say *at most three*.

(23)   Mary is allowed to submit at most two abstracts.

However, as the upper bound is not cancellable, as illustrated in (24), this does not seem likely.

---

[5]Here ANT has been replaced by $\varnothing$, $\Diamond$ has been replaced by $\square$, and *at least* has been replaced by *exactly*.

(24)   Mary is allowed to submit at most two abstracts. #In fact, she can submit three.

I believe that (24) strongly suggests that the authoritative upper bounded reading should be accounted for in the semantics. The only way to arrive at a reading with a strong upper bound is to say that *at most* has to take scope over the modal. As we have seen, letting it take scope under an existential modal inevitably leads to truth conditions that are too weak.

A way to get rid of the reading where *at most* takes narrow scope in Penka's account is to say that ANT has negative features.[6] It has been observed in the literature (Iatridou & Zeijlstra, 2010) that negation outscopes existential modals. If ANT displays the same behaviour, this would rule out the LF in (18-a), where ANT occurs in the scope of the modal. Penka's antonymising operator does not have the semantics of negation in that it does not yield the complement of its prejacent, but Penka could still stipulate that ANT is an operator that displays the same syntactic behaviour as negation.

If she takes this route, a potential problem is the fact that *at least* is a PPI, as illustrated in (25) (Spector, 2014).

(25)   ??Mary didn't solve at least three problems.

Taking this into consideration, it would be curious if *at most* were decomposed into an operator with negative features and *at least*, where *at least* consistently occurs in the immediate scope of the negative operator. However, Penka could again argue that the semantics and the syntactic behaviour of the elements under discussion should be teased apart: perhaps the *at least* part of the decomposed *at most* is not the same *at least* as the *at least* that occurs by itself. It could be an operator that has the same denotation as *at least* but does not share its syntactic features, one of those features being its PPI-hood.[7]

While a few stipulations are needed, it seems as though Penka's theory can account for the way *at most* interacts with existential modals. However, there is one crucial fact that has been overlooked thus far: the data with *at least* and existential modals suggest that upper-bounded numeral modifiers like *at most* do not actually behave differently from lower-bounded numeral modifiers like *at least*. To see this, let us reconsider (8), repeated here as (26-a).

(26)   a.   Mary is allowed to submit at least one abstract.
          b.   Mary is allowed to submit more than one abstract.

As we saw in section 2.1, these kinds of examples are predicted to give rise to two different types of ignorance readings in Schwarz's neo-Gricean account (cf. (9)-(11)): one where the speaker claims that the maximal allowed number of abstracts Mary submits is at least one (but she does not know the exact maximum), and one where the speaker merely claims that submitting at least one abstract is allowed (but she does not know if submitting exactly one is allowed or if submitting more than one is allowed). This weaker reading seems absent for (26-a); one clearly gets the impression that there is a maximum number of abstracts Mary is allowed to submit. The weaker reading does seem to be available when we use *more than*, as in (26-b).

Schwarz claims that the weaker reading is not visible because it is blocked by the existence of the stronger one. To test this claim, let us consider a scenario where only the weak reading would be felicitous. Say that there is a high demand for instant formula, and the manifacturers are unable to keep up with this demand. For this reason, a particular retail chain has a rule that customers are only allowed to buy one box of instant formula at a time. After some

---

[6]I would like to thank Yaron McNabb for suggesting this to me.

[7]This would tie Penka to a syntactic story of PPI anti-licensing. Under a semantic account, it is not possible for two expressions to share the same truth conditions and for only one of those expressions to be a PPI.

time, the manifacturers manage to catch up with the high demand, so the rules are loosened. However, there is some confusion about what the new rules are. One store manager has said that customers are now allowed to buy a maximum of two boxes at a time, and that the boxes will be sold in packages of two to facilitate transport. In other words, it is only possible to buy exactly two boxes. Another store manager has announced that there no longer is a maximum, so customers can buy as many boxes of instant formula as they want. She has also said that the boxes will be sold in packages of three, which means customers can buy three or more boxes. David is talking to someone who is not aware of the recent changes; this person thinks that customers are still allowed to buy only one box of formula. David does know that the rules have been changed, but he does not know which store manager to believe. David says:

(27)   I'm not entirely sure about the current rules, but I know you're allowed to buy { more than one box / # at least two boxes } now.

In this scenario, David uses *at least two* in a situation where he is not sure whether buying exactly two boxes is allowed or whether buying at least three boxes is allowed, which corresponds exactly to the weak narrow scope reading. The fact that this sentence is not felicitous in the context indicates that the weak reading that arises when *at least* takes scope under an existential modal is simply not there. By extension, the data point towards the conclusion that lower-bounded class B modified numerals, like their upper-bounded counterparts, take scope over existential modals.

What about universal modals? *At most* seems to be able to scope both over and under universal modals. (28-a) has the reading in (28-b), where *at most* takes narrow scope. This says that Mary submits two or fewer abstracts in all permissible worlds. It also appears to have the wide scope reading in (28-c), which says that the number of abstracts Mary submits in every permissible world is two or lower (i.e. the minimum requirement is two or less). This is an ignorance reading that is compatible with Mary being allowed to submit more than two abstracts, as illustrated in (29).

(28)   a.   Mary is required to submit at most two abstracts.
         b. $\Box$ [MAX { n | Mary submits n abstracts } ≤ 2]
         c.   MAX { n | $\Box$ [Mary submits n abstracts] } ≤ 2

(29)   I'm not sure how many abstracts Mary has to submit, but I know she's required to submit at most two. She may actually choose to submit three abstracts though.

Unfortunately it is not possible to see where *at least* takes scope with respect to universal modals as both scope configurations lead to the same truth conditions.[8] Since *at most* and *at least* behave the same way in their interactions with existential modals, the null hypothesis should be that they also display the same behaviour in their interactions with universal modals.[9]

---

[8]The denotations corresponding to the two scope configurations of (i-a) are given in (i-b) and (i-c) (repeated from (4)-(5)). (i-b) says that Mary submits one or more abstracts in all permissible worlds. This is equivalent to (i-c), which says that the maximum number such that Mary submits that many abstracts in all permissible worlds is one or higher.

(i)   a.   Mary is required to submit at least one abstract.
         b. $\Box$ [MAX { n | Mary submits n abstracts } ≥ 1]
         c.   MAX { n | $\Box$ [Mary submits n abstracts] } ≥ 1

[9]The only differences one would expect have to do with the fact that *at most* is downward entailing. As some universal modals are said to be PPIs (Iatridou & Zeijlstra, 2013; Homer, 2015), this could mean that these modals have to outscope *at most* but not the upward entailing *at least*.

Thus, the most plausible hypothesis seems to be that class B numeral modifiers outscope existential modals and can scope both over and under universal modals. More evidence for these claims come from finite clause islands. It is known that quantifier raising over finite clause boundaries is not possible (e.g. Fox, 2000) so these kinds of islands are a good way to test if certain scope configurations are possible. Let us see what happens when we force modified numerals to take scope below existential modals by putting them in a finite clause island. As can be observed in (30-a), class A modifiers seem quite comfortable in this position. The class B modifiers in (30-b), on the other hand, do not accept being forced to take scope under existential modals.[10]

(30)   a.   It is allowed that you write { fewer than / more than } five pages.
        b. #It is allowed that you write { at least / at most } five pages.

The examples in (30-b) are only acceptable in echoic contexts, for example as a reply to the question if writing at least or at most five pages is allowed. As exemplified by the dialogue in (31), where a PPI is licensed in the scope of negation, echoic contexts allow all sorts of constructions that are normally ruled out. Therefore, the fact that echoic contexts may license the examples in (30-b) does not say much.

(31)   A: Did you see someone?
        B: No, I didn't see someone.

Now let us turn to cases where modified numerals are trapped in finite clause islands under universal modals. These examples are all felicitous and do not require an echoic context to be licensed. This shows that class B numeral modifiers do not mind occurring in the scope of a universal modal.

(32)   a.   It is required that you write { fewer than / more than } five pages.
        b.   It is required that you write { at least / at most } five pages.

A final piece of evidence for the generalisation I am defending comes from interactions with epistemic modals. (33) shows that *at most* also takes scope over epistemic existential modals (pace Kennedy, 2015), yielding a reading where the evidence rules out the possibility that there are more than four burglars in the building.

(33)   Police evidence suggests there may be at most four burglars in the building.

In sum, the data discussed in this section point towards the conclusion that class B modifiers must outscope existential modals but can occur in the scope of universal modals. As lower-bounded and upper-bounded class B modifiers appear to take scope the same way, a theory like Penka's that decomposes *at most* but not *at least* is probably not on the right track. In the next section I discuss some more consequences of these observations.

---

[10]Some of my anglophone informants found 'it is allowed that φ' a bad construction. In Dutch, however, the construction is perfectly fine, and the Dutch data are the same as the English data provided by those speakers who did accept the 'it is allowed that' construction:

(i)   a.   Het is toegestaan dat je   { minder / meer } dan vijf pagina's schrijft.
              It   is permitted that you { fewer   / more } than five pages   write.
              'It is permitted that you write { fewer / more } than five pages.'
        b. #Het is toegestaan dat je   { minstens / hoogstens } vijf pagina's schrijft.
              It   is permitted that you { at least   / at most   } five pages   write.
              'It is permitted that you write { at least / at most } five pages.'

3.2   Further consequences

As far as I am aware, all current theories of modified numerals use scope to derive the two readings that arise when a modified numeral interacts with a modal (the neo-Gricean accounts in Büring, 2008; Schwarz, 2011, 2013; Kennedy, 2015 but also accounts in other frameworks such as Nouwen, 2010; Coppock & Brochhagen, 2013). If class B modifiers must indeed take scope over existential modals, this presents a problem for all these accounts. While some authors argue that sentences with *at least* and an existential modal only give rise to ignorance readings (e.g. Schwarz, 2011; Penka, 2014), it is clear that this is not so for cases with *at most* and an existential modal. Examples like (12), repeated here as (34), have both an authoritative reading and an ignorance reading.

(34)   Mary is allowed to submit at most two abstracts.

If the modified numeral always takes scope over the modal, these two readings cannot be said to be derived from two different scope configurations. We must therefore find another way of accounting for the two readings of examples like (34).

  As for cases with universal modals, which display the same ambiguity as (34), there are two theoretical possibilities. Either we propose that their two readings, too, must be accounted for in a different way, or we maintain that their ambiguity arises from scope interactions. The latter option involves proposing two different analyses for authoritative and ignorance readings: one for cases with universal modals and another for cases with existential modals. For this reason, the former option seems preferable. This would mean that we need a new theory of modified numerals that treats the scope interactions between modified numerals and modals and the authoritative and ignorance readings they give rise to as two separate phenomena.

  Another argument for such a theory comes from the Heim-Kennedy generalisation, which states that degree quantifiers can bind their trace across modals but not across nominal quantifiers (Heim, 2000). Assuming, as I have done in this paper, that modified numerals are indeed degree quantifiers, this would mean that they should be unable to outscope universal quantifiers in examples such as (35).

(35)   Every student submitted { at least / at most } two abstracts.

This seems to be the case: the wide scope reading, given in (36), is that the maximum number of abstracts such that all students submitted that many abstracts is two or lower, which means that the number of abstracts submitted by the student who submitted the least amount of abstracts is two or lower. This lower bound reading is not attested.

(36)   MAX $\{ n \mid \forall x \, [\, \text{student}(x) \rightarrow x \text{ submits } n \text{ abstracts} \,] \, \} \leq 2$

It seems to me that sentences like (35) also have two readings (pace Penka, 2014). The most obvious reading is the reading where the speaker has full knowledge of the situation (parallel to the authoritative reading) and conveys that different students submitted different numbers of abstracts, and they all submitted two or more/less. The other reading is the ignorance reading: there is a specific number such that every student submitted that many abstracts. The speaker does not know what that number is, but she knows it is two or more/less.

  Class B modified numerals thus appear not to be able to outscope universal quantifiers, but they do give rise to two readings when they occur in sentences with universal quantifiers. Therefore, examples with universal quantifiers and modified numerals are another case where we do see an ignorance reading and a non-ignorance (variation) reading but where we do not want to say that scope is responsible for this. Both the examples with existential modals and the

sentences with universal nominal quantifiers point towards the conclusion that the ambiguity we observe must be accounted for in a different way.

4   Conclusion

I have argued that class B modified numerals always take scope over existential modals. This means that a decomposition story à la Penka (2014), in which *at most* has different scopal properties than *at least*, does not account for the data we observe. More importantly, it means that the two readings modified numerals display when they are combined with modals and the way modified numerals take scope with respect to modals should be treated as two separate phenomena. More evidence for the latter claim comes from examples with modified numerals and universal nominal quantifiers. These sentences, like the ones with existential modals, allow only one scope configuration but do give rise to two readings. The next question is what such a 'scope-independent' theory of modified numerals might look like. I leave this issue for future research.

References

Büring, D. (2008). The least *at least* can do. In *Proceedings of the west coast conference on formal linguistics* (Vol. 26, p. 114-120).

Coppock, E., & Brochhagen, T. (2013). Raising and resolving issues with scalar modifiers. *Semantics & Pragmatics*, *6*(3), 1-57.

Fox, D. (2000). *Economy and semantic interpretation.* Cambridge, MA: MIT Press.

Geurts, B., & Nouwen, R. (2007). At least *at al.*: the semantics of scalar modifiers. *Language*, *83*(3), 533-559.

Heim, I. (2000). Degree operators and scope. In *Proceedings of SALT 10* (p. 40-64). Ithaca, NY: Cornell Linguistics Club.

Homer, V. (2015). Neg-raising and positive polarity: The view from modals. *Semantics and Pragmatics*, *8*(4), 1-88.

Iatridou, S., & Zeijlstra, H. (2010). On the scopal interaction of negation and deontic modals. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.), *Proceedings of the 17th Amsterdam Colloquium.* Berlin: Springer Verlag.

Iatridou, S., & Zeijlstra, H. (2013). Negation, polarity, and deontic modals. *Linguistic Inquiry*, *44*(4), 529-568.

Kennedy, C. (2015). A 'de-Fregean' semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics*, *8*(10), 1–44.

Nouwen, R. (2010). Two kinds of modified numerals. *Semantics & Pragmatics*, *3*(3), 1-41.

Penka, D. (2014). 'At most' at last. In E. Csipak & H. Zeijlstra (Eds.), *Proceedings of Sinn & Bedeutung 19.*

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics & Philosophy*, *27*(3).

Schwarz, B. (2011). *Remarks on class B numeral modifiers.* Handout of workshop *Indefinites and beyond*, Universität Göttingen.

Schwarz, B. (2013). *At least* and quantity implicature: Choices and consequences. In M. Aloni, M. Franke, & F. Roelofsen (Eds.), *Proceedings of the 19th Amsterdam Colloquium.*

Spector, B. (2014). Global positive polarity items and obligatory exhaustivity. *Semantics & Pragmatics*, *7*(11), 1-61.

Selection Function Semantics for *Will*

Fabrizio Cariani[1] and Paolo Santorio[2*]

[1] Northwestern University, Evanston, Illinois, U.S.A.
`f-cariani@northwestern.edu`
[2] University of Leeds, Leeds, U.K.
`P.Santorio@leeds.ac.uk`

Abstract

We develop a new semantics for the English auxiliary *will* that exploits a Stalnaker-style, single-world selection function. Unlike existing theories, the resulting analysis succeeds in satisfying three *desiderata*: it accounts for the modal character of *will*, it predicts its peculiar lack of scope interactions, and it vindicates intuitive judgments about the probabilities of *will*-claims.

We develop a new semantics for the English auxiliary *will* that exploits selection functions, similarly to Stalnaker's [24] semantics for conditionals. The semantic is motivated by three constraint, which we state in §1. In §2, we show that no existing account satisfies all three constraints. We state our analysis in §3. In §4 and §5, respectively, we show how this analysis accommodates indeterminacy about the future and intuitive judgments about the probability of *will*-claims. We close, in §6, by discussing the assertability of *will*-claims.

1 Three constraints

A plausible semantics for *will* must satisfy three constraints.

Constraint 1: The modal character of *will*. There is evidence that *will* is a modal rather than a tense. By this we mean that *will* manipulates a world parameter, similarly to modal auxiliaries like *must* or *might*. We understand the modal analysis as compatible with the claim that *will* also manipulates a time parameter; the salient contrast is with a view on which *will* manipulates exclusively a time parameter. The literature has provided three pieces of evidence for the modal view. Taken together, they seem to us compelling.

Shared morphology with *would*. It is widely accepted [1, 5, 13] that *will* shares morphology with the modal *would*. In particular, *will* and *would* have in common a modal morpheme, often represented as 'WOLL': *will* is PRESENT + WOLL; *would* is PAST + WOLL. The assumption of common morphology helps explain some semantic facts. For example, it explains why we can replace *will* with *would* in indirect reports of past utterances of *will*-sentences. If, on Tuesday, Harriet says "I will come to work tomorrow", then on Wednesday we would report Harriet's utterance by saying "Harriet said she would come to work today".

Epistemic readings of *will*. *will* has uses where it performs no temporal shift, but rather flags that the speaker is making a prediction on the basis of evidence [22]:

(1)   John will be in London by now.

will in (1) doesn't perform temporal shift: as is made clear by the modifier *by now*, the pre-jacent of (1) has its reference time in the present. Moreover, in (1), *will* works as a marker of evidentiality. To see this, notice that (1) is infelicitous if uttered by someone who is looking directly at John, even if both are indeed in London. These facts are hard to explain on a purely temporal view, but easily accommodated by a modal view.

Modal subordination. As Peter Klecha [14] has argued, *will* allows for modal subordination. Roughly, this means that *will*-sentences can inherit domain restrictions from previous elements of the discourse.

(2)   If the supplies arrive tomorrow, it will be late in the day. They will contain three boxes of cereal.

(3)   The supplies might arrive tomorrow. It will be late in the day.

This makes these sentences pattern with modals, and unlike tenses.

(4)   a.   If the supplies arrive tomorrow, it might (/must) be late in the day. They might (/must) contain three boxes of cereal.
       b.   #If the supplies arrived yesterday, it was late in the day. They contained three boxes of cereals.

The availability of modal subordination may be explained (as we do below) by treating *will* as a modal, and postulating that modals can stand in anaphoric relations to each other. By contrast, the pattern in the data seems hard to account for on a temporal analysis.

Constraint 2: Scopelessness.   *will* displays no significant scope interactions with other lexical items: i.e., changes in the relative syntactic scope of *will* and other items don't affect truth conditions. Here we focus on negation, but the point generalizes. For a basic illustration, notice that (5-a) and (5-b) are truth conditionally equivalent.

(5)   a.   It will not rain.
       b.   It is not the case that it will rain.

The phenomenon persists in *will*'s interactions with items that lexicalize negation, like *doubt* (≈*believe that not*) and *fail* (≈*not pass*), as the equivalence of (6-a) and (6-b) shows.

(6)   a.   I doubt that Sam will pass his logic exam.
       b.   I believe that Sam will fail his logic exam.

The lack of scope interactions with negation yields a logical constraint, which for the moment we state by appealing to an intuitive notion of logical truth:

   Will Excluded Middle (preliminary take): ⌜*will* A ∨ *will not* A⌝ is a logical truth

Constraint 3: The reasonableness of future uncertainty.   Speakers often have nonextreme credences in propositions expressed by *will*-claims. Moreover, this uncertainty seems sometimes rational. Consider an example:

   Sports Fan: Cynthia comes to work each day wearing a Warriors cap or a Giants cap, depending on the outcome of a fair coin toss. You are certain that for each of the two caps, it is an open possibility that Cynthia wears that cap tomorrow.

In this scenario, it seems permissible for you to assign credence .5 to an utterance of (7).

(7)   Cynthia will wear a Warriors cap.

This fact seems obvious, but (as we point out in the next section) it is hard to accommodate on existing modal accounts.

2   Existing analyses

In the the philosophical literature, *will* is usually analyzed as a tense (often in combination with a supervaluational treatment of the world parameter; see [25], [26], [3]). Obviously, these analyses violate Constraint 1, and hence fail to account for the data motivating a modal analysis. For reasons of space, we don't discuss accounts of this sort in this paper. Our goal is developing a modal account that meets in full Constraints 2 and 3.

  In the linguistics literature, *will* is usually (though not invariably; see e.g. [27]) analyzed as a modal. Following Kratzer [15, 16], we take the semantic value of *will* to be relativized to a modal base and an ordering source, which throughout this section we denote via, respectively, the metavariables 'f' and 'g'. The modal base is function from the world of evaluation to a set of worlds (the domain of quantification of the modal); the ordering source is a function from the world of evaluation to a pre-order $\preceq_w$ which is defined on the domain. (For simplicity, we'll allow ourselves to be sloppy at times, talking of the modal base simply as a set of worlds, and of the ordering source simply as a partial ordering on that set.) In this framework, all existing modal analyses share a schematic form. Under the limit assumption, this form is:

(8) $[\![\textit{will } \mathsf{A}]\!]^{w,f,g} = 1$ iff $\forall w' \in \text{BEST}_{g(w)}(f(w))$, $[\![\mathsf{A}]\!]^{w'} = 1$

where the set of 'best' worlds is the set of worlds ranked as maximal by the ordering source:

(9) $\text{BEST}_{g(w)}(f(w)) = \{ w' \in f(w) : \forall w'' \in f(w), w' \preceq_{g(w)} w'' \}$

Together, (8) and (9) specify the schematic form of modal analyses. These analyses generally agree in taking the modal base of *will* to include at least all possible worlds that count as 'open possibilities' at the time of utterance. The relevant notion of openness is metaphysical, and connects to the intuition that future events, unlike past events, are not settled. For current purposes, we understand openness in this way (following [2]):[1]

> **Openness.** At least some contingent facts about the future are not settled at the present moment in time.

While modal analyses generally agree about the modal base of *will*, they differ widely in their treatment of the ordering source. The ordering is understood as capturing a notion of likelihood (e.g., [13]), normality (e.g., [6]), or match the speaker's knowledge (e.g., [10]). The choice between these options has significant empirical consequences. We set it aside because the basic problem concerns the schematic form that they all share. Any theory that implements the schematic analysis in (8) and assumes that the domain of the universal quantifier is larger than a singleton fails to satisfy at least one of our constraints, and in particular Constraint 3.

  A first, potential difficulty concerns Constraint 2. By treating *will* as a universal quantifier, we predict that *will* (on a par with other universal modals, like *must* and *have to*) has nontrivial scope interactions with negation. This problem is well-known in the literature. The standard solution is Copley's [6]. Building on work by von Fintel on generics [7], she adds a presupposition

---

[1] This notion of openness is conveniently neutral on philosophical views about determinism and bivalence.

to the meaning of *will*: *will* presupposes that its domain of quantification is uniform with respect to the prejacent—either all worlds satisfy it, or none does.[2]

(10) $\llbracket$*will* A$\rrbracket^{w,f,g}$ is defined iff: either $\forall w' \in \text{BEST}_{g(w)}(f(w))$, $\llbracket$A$\rrbracket^{w'} = 1$,
      or $\forall w' \in \text{BEST}_{g(w)}(f(w))$, $\llbracket$A$\rrbracket^{w'} = 0$;
      if defined, it is true iff $\forall w' \in \text{BEST}_{g(w)}(f(w))$, $\llbracket$A$\rrbracket^{w'} = 1$

This analysis validates **Will Excluded Middle**, at least on a notion of entailment that makes appropriate reference to presupposition (such as, e.g., von Fintel's Strawson Entailment [8]).

Building a uniformity presupposition in the lexical entry of *will* satisfies Constraint 2—we get the correct predictions about *will*'s compositional interactions with negation. But, crucially, (10) still fails to satisfy Constraint 3. To see this, consider an example.

Coin. I'm about to toss a coin. The outcome of the coin toss is genuinely open. Moreover, you and I are both aware of this.

Suppose that I say:

(11) The coin will land tails.

What credence should you assign to (11)?[3] The natural answer is '1/2'. Even if you have doubts about this specific answer, it is clear that the correct answer is not 'zero', or 'near-zero'. We take this to provide both a descriptive data point—the degree of belief that most speakers of English assign to the content expressed by (11) is substantially higher than zero—and a normative one—the appropriate degree of belief in the content of (11) as uttered in the Coin scenario is also substantially higher than zero. Yet, modal theories that conform to (8), whether or not they incorporate a uniformity presupposition, are committed to this answer.

To see why, notice that, on all the available glosses for the ordering source, the set of best worlds that (11) quantifies over includes both heads- and tails-worlds.[4] Speakers are in a position to know this, hence they should know that (11) is false—since (11) requires that *all* worlds quantified over are tails-worlds. As a result, theories conforming to (8) predict that speakers should assign zero, or in any case very low credence to it. Moreover, they predict that this credence assignment should count as the only rational one. These predictions seem incorrect.

Notice that the uniformity presupposition doesn't help here. The effect of the presupposition is to make (11) undefined in our scenario. It is unclear what credence, if any, one should assign to utterances made in these circumstances. But it seems both that speakers don't assign them positive credences, and that it would be irrational to do so. For a comparison, consider (12).

(12) The King of France is bald.

It seems irrational to assign positive credence to that proposition, while also being certain that France is not a monarchy. And in fact, ordinary speakers have no temptation to do so.

Summing up: all existing modal theories violate Constraint 3. In the next section, we state a new semantics for *will* that satisfies all three constraints. The key move is dropping the idea that *will* is a universal modal. In fact, we deny that *will* has quantificational force at all.

---

[2]Copley's implementation differs in a number of details that don't make a difference for current purposes.

[3]We assume that it makes sense to talk about the credence attaching to an utterance, and that this credence derives from the credence attaching to the proposition expressed by the utterance.

[4]If you think that this assumption is incorrect for this specific example, just switch examples. There will be analogous cases on any view on which the domain of quantification of *will* is larger than a singleton.

3   Selection function semantics for *will*

1. Overview.   Our semantics exploits an extended analogy with Stalnaker's [24] semantics for conditionals. We assume that *will* denotes a selection function, i.e. a function that maps a pair of a world w and a set of worlds S to a 'selected' world w'. Intuitively, the world selected represents the way things will actually be; this world is selected out of the set of worlds that are compatible with history up to the time of utterance. Before proceeding, a word about notation: we use sans-serif letters ('A', 'B', etc.) as metalinguistic variables over sentences; and boldface letters ('$\mathbf{A}$', '$\mathbf{B}$', etc.) as metalinguistic variables over sets of worlds.

2. Selection function semantics.   Let's start by defining selection functions.

A function $s : W \times \mathcal{P}(W) \mapsto W$ is a *selection function* iff

  i. **Inclusion**: if $\mathbf{A}$ is non-empty, $s(w, \mathbf{A}) \in \mathbf{A}$, and
  ii. **Centering**: if $w \in \mathbf{A}$, then $s(w, \mathbf{A}) = w$.

Informally, a selection function is a function s that maps a world w and a proposition $\mathbf{A}$ to a world w' and that satisfies Inclusion and Centering. Inclusion says that the world selected must verify the input proposition (provided that the input proposition is non-empty). Centering says that, if the input world verifies the input proposition, then s selects the input world itself.[5]

We treat *will* as a sentential operator, i.e. an operator that takes a full clause as argument. (This is a simplification, but a harmless one.) We relativize interpretation to a context c (generally suppressed to avoid clutter), a world of evaluation w, an assignment g, and a selection function s. In addition to its prejacent, *will* takes a modal-base-type argument. We assume that modal bases are the semantic values of covert object language pronouns; we represent the latter as 'f_i', and the set of worlds individuated by a modal base and a world of evaluation w as '$\mathbf{F}_{i,w}$'. For shorthand, we often represent modal bases just as subscripts of modals. Hence we write '$will_f$' as a shorthand of '$will\ [f]$'.

At this point, we are ready to state the meaning of *will*. The schematic truth conditions of $\ulcorner will_f\ \mathsf{A} \urcorner$ are in (13), and the lexical entry for *will* in (14):

(13) $[\![will_f\ \mathsf{A}]\!]^{w,s,g} = 1$ iff $[\![\mathsf{A}]\!]^{s(w,g(f)),s,g} = 1$

(14) $[\![will]\!]^{w,s,g} = \lambda F_{\langle s,st \rangle}.\ \lambda p_{\langle s,t \rangle}.\ p(s(w, F(w))) = 1$

Informally, $\ulcorner will_f\ \mathsf{A} \urcorner$ is true (relative to w, s, g) just in case A is true relative to the world v that is selected by s when the input is the world of evaluation and *will*'s modal base (and s and g, which stay unshifted). Notice that we are leaving temporal shift entirely out of the meaning of *will*. We do this for simplicity. Introducing temporal shift would be easy,[6] but would be a distraction from our main innovation, which concerns the way in which *will* manipulates the world of evaluation parameter.

3. Modal base and historical alternatives.   Before stating truth conditions, we need to specify what set of worlds is determined by the modal base of *will*. We said that current modal accounts, building on an assumption of Openness, identify the modal base with the set of possible worlds that are open possibilities at the time of utterance. Our proposal is similar, but

---

[5]We leave it open whether further conditions apply to selection functions, as in Stalnaker's original theory, or whether (i) and (ii) are the only conditions in place.

[6]For example, following Kaufmann [13], we could let *will* extend forward the time interval at which the prejacent is evaluated.

not quite analogous. We assume that, as a default, the modal base of *will* at world w and time t is the set of **historical alternatives** of w at t. Here is how we define historical alternatives:

> Two worlds w and v are historical alternatives at t iff w and v match perfectly in their history (i.e., iff they match perfectly in matters of particular fact) up to t.

The notion of perfect match in matters of particular fact is borrowed from David Lewis [17]. Two worlds that perfectly match in matters of particular fact up to a certain point in time are duplicates—indiscernible copies of each other—up to that point. Two observations: first, historical alternatives form an equivalence class; second, since each world is a historical alternative of itself, the world of evaluation is always a member of the modal base when *will* is unembedded. Notice that we are not assuming that the modal base of *will* is invariably historical. We leave it open that (as it happens for, e.g., *have to* and *may*) *will* may also have a modal base of different flavor (more about this shortly).

We can now state the truth conditions of a sample sentence:

(15) $[\![\text{It will}_{f_i}\ \text{rain}]\!]^{w,s,g}$ is true at w iff it rains at $s(w, [g(f_i)](w))$

Notice a consequence of our theory. Given our assumptions about the modal base, we know that the set $[g(f_i)](w)$ includes w, i.e. the world of evaluation itself. Moreover, given the Centering condition, we know that the selection function will select w itself, i.e. $s(w, [g(f_i)](w)) = w$. Hence the truth conditions in (15) simplify to:

(16) $[\![\text{It will}_{f_i}\ \text{rain}]\!]^{w,s,g}$ is true at w iff it rains at w

More generally: in combination with our background assumptions, our semantics makes unembedded occurrences of *will* semantically vacuous with respect to the modal parameter.

(17) $[\![will_f\ \mathsf{A}]\!]^{w,s,g} = 1$ iff $[\![\mathsf{A}]\!]^{w,s,g} = 1$

Thus, when *will* occurs unembedded, our semantics effectively collapses on a simple nonmodal semantics which treated *will* as a mere tense. Nevertheless, as we show in the next paragraphs, our semantics differs from the nonmodal analysis in important respects.

4. Predictions.   Our theory provides the tools to predict all the data presented in section 1, at least in outline. For reasons of space, we can't give a full account of all these predictions, which are discussed in more detail in [4]. Here we give a brief survey of how the theory meets Constraints 1 and 2 (immediately below) and 3 (in section 5).

Shared morphology with *would*. On a Stalnakerian semantics for *would*, the *will*/*would* connection is vindicated in a simple and elegant way: *will* and *would* realize the very same modal morpheme. The connection is less straightforward on an analysis on which *would* works as a universal quantifier; but let us notice that a selection semantics is anyway better equipped to account for the connection than any nonmodal theory.

Epistemic readings of *will*. We noticed that *will* can have epistemic readings:

(1)   John will be in London by now.

Above, we suggested that the modal base of *will* defaults to a set of historical alternatives of the world of evaluation. But we don't assume that this is the only kind of modal base available for *will*. Rather, we allow that *will* may take an epistemic modal base. This alone doesn't account for the epistemic readings of *will*, but (presumably, in combination with an account of the evidential features of epistemic modals) is a first step towards such an account.

Modal subordination. Consider again our example of modal subordination involving *will*:

(2)   If the supplies arrive tomorrow, it will be late in the day. They will contain three boxes of cereal.

Our analysis can predict the relevant interpretation of (2), if combined with appropriate assumptions of *if*-clauses and a general account of modal subordination. As for the former: we assume (following Kratzer [16]) that *if*-clauses work as semantics restrictors of modal bases. As for the latter: we assume that modal subordination is generated by anaphoric connections between modal base pronouns (for an account in this style, see [9]). Against this background, all we need to do is assume that the modal bases of the two occurrences of *will* are coindexed:

(18)   If the supplies arrive tomorrow, it will$_{f_3}$ be late in the day. They will$_{f_3}$ contain three boxes of cereal.

Scopelessness. Let us start by defining a notion of validity.

> **Validity***: $\mathsf{A}_1, ..., \mathsf{A}_n \vDash \mathsf{B}$ iff, for any triple $\langle w, s, g \rangle$ such that $[\![\mathsf{A}_1]\!]^{w,s,g} = 1$, ..., $[\![\mathsf{A}_n]\!]^{w,s,g} = 1$, $[\![\mathsf{B}]\!]^{w,s,g} = 1$

We can now formulate **Will Excluded Middle** in a precise way:

> **Will Excluded Middle** (precise take): $\ulcorner will\ \mathsf{A}\ \vee\ will\ not\ \mathsf{A} \urcorner$ is logically true*.

(As usual, we take a sentence to be logically true* iff the inference from the empty set of premises to it is valid*.) It's easy to see that **Will Excluded Middle** holds.

> PROOF: let $\langle w, s, g \rangle$ be an arbitrary point of evaluation. We have that $\ulcorner will_f\ \mathsf{A} \vee will_f\ not\ \mathsf{A} \urcorner$ is true at $\langle w, s, g \rangle$ iff either **A** is true at $s(w, \mathbf{F}_w)$ or false at $s(w, \mathbf{F}_w)$. But the right-hand side of the biconditional is true for any choice of w, s, and $\mathbf{F}_w$. Hence, $\ulcorner will_f\ \mathsf{A} \vee will_f\ not\ \mathsf{A} \urcorner$ is true at $\langle w, s, g \rangle$. Since $\langle w, s, g \rangle$ was arbitrary, $\ulcorner will_f\ \mathsf{A} \vee will_f\ not\ \mathsf{A} \urcorner$ is true at any point of evaluation.

Validity* is a very strong kind of validity (it captures preservation of truth at a point of evaluation). Hence the fact that **Will Excluded Middle** holds in the form we state it immediately entails that sentences of the form $\ulcorner will\ \mathsf{A}\ \vee\ will\ not\ \mathsf{A} \urcorner$ are valid on a number of weaker notions of validity, including validity in the Kaplanian sense (i.e., preservation of truth at a context). See the full discussion in [4, §7] for further important logical consequences of our semantics.

4   Indeterminacy

Our semantics for *will* assumes that, at the time of utterance, there is a unique, fully specified way things will actually be. This assumption is controversial. Several theorists object that we have no right to assume that there is a fully specified way things will be. On the one hand, it might be that the future is open and that there is no fact of the matter about what world is actual [25, 3]. On the other, even if the future is not open, it is unclear that semantics can legitimately presuppose metaphysical claims about Openness [18, 19, 20].

Even if one agrees with these concerns, we don't think that the compositional semantics for *will* needs to be changed. Openness should be accommodated not by changing the lexical entry of *will*, but rather by allowing that there is indeterminacy in the value of contextual parameters that *will* appeals to. In particular, our denotation for *will* assumes that there is a single world of evaluation. We can preserve this assumption, but allow that it may be indeterminate which world this is. Let us show how this idea can be implemented at the technical level.

We start by defining a notion of truth at a context.[7] We do this in the standard contextualist fashion deriving from Kaplan's [12]. Take contexts to be fully specific situations that include (at least) a world, a time, and a speaker. (Crucially, we refrain from assuming that a fully specific situation in this sense corresponds to a concrete situation of utterance; more on this in a moment.) We define truth at a context by fixing the values of index parameters to the coordinates of the context. Formally:

Truth at a Context. A is true as uttered at c iff $[\![A]\!]^{w_c, s_c, g_c} = 1$

It follows that, given a context, every *will*-claim has a determinate truth value. At the same time, if Openness is correct, it may be indeterminate which context corresponds to the actual situation of utterance, since it is unsettled which world the utterance takes place at.

One consequence of implementing Openness in this way is that both the supporter and the opponent of Openness can make use of our formalism. The disagreement over Openness is moved out of the semantic apparatus entirely. While we don't assume that semantics has to be neutral between different metaphysical options, we regard it as a welcome feature of our theory that theorists in both metaphysical camps can help themselves to it.

5   Probability of *will*-claims

Recall our Constraint 3. Ordinary agents are uncertain about the future. On one natural way to understand this uncertainty, this means that ordinary agents have nonextreme degrees of belief in the propositions expressed by *will*-claims. Moreover, at least in some cases, it seems that this uncertainty is rationally permissible, if not rationally required. Now we are equipped to show how our theory meets this constraint.

Recall our Sports Fan scenario: every day, Cynthia tosses a fair coin and, on the basis of the outcome, decides whether to wear a Giants hat or a Warriors hat. Consider an agent who assigns credence $1/2$ to each of the two possibilities. Against this background, what we want to show is that this agent also has credence $1/2$ in the proposition expressed by (7).

(7)   Cynthia will wear a Warriors cap.

To extract a verdict from our system, we need to specify (a) what agents' credences attach to, and (b) how we individuate the propositions expressed by *will*-claims.

As for the first point: we simply assume that credences are defined over sets of worlds. In particular, we assume that an agent's credences at a given time may be modeled by a probability function μ satisfying the usual constraints. For example, let μ model an agent's credences at the current point in time. Let $\mu(\mathbf{A}) = 1/2$, where $\mathbf{A}$ is the set of worlds where Cynthia wears a Warriors cap. Our task is to check that $\mu(\mathbf{PROP}_W) = 1/2$, where $\mathbf{PROP}_W$ is the proposition expressed by an utterance of (7).

As for the second point: we simply identify the proposition expressed by the utterance of a *will*-sentence at a given context with the set of worlds such that the utterance is true as evaluated at those worlds. Formally:

Content of A at c: $\|A\|_c = \{w : [\![A]\!]^{w, s_c, g_c} = 1\}$

In what follows, we suppress reference to the context to avoid clutter.

---

[7] Here we adopt a specific account of how indeterminacy affects the semantics, i.e. the one defended by [2] (see also [11]). Barnes & Cameron's view is part of a family of views that draw inspiration from supervaluatism but retain a bivalent semantics; the inspiration for accounts of this sort comes from [21].

It's easy to see that our semantics yields exactly the verdict we want. $\|$Cynthia will wear a Warriors cap$\|$ is just the set of worlds in which Cynthia wears a warriors cap. On the assumption that the credence that our agent assigns to Warriors-cap-worlds is $1/2$, she will also assign credence $1/2$ to the proposition expressed by (7). More generally:

Transparency: For any prejacent A, $\|\mathit{will}_f\ \mathsf{A}\| = \|\mathsf{A}\|$.[8]

6 Assertion of *will*-claims

We close by noting an open problem for our semantics. If the future is genuinely open, then, for at least some future claims, it is not settled that those claims are true and it is not settled that they are false. Yet some of these claims seem okay to assert (and others not). For example:

(19)   The 2022 World Cup will take place in Qatar.

(19) seems assertable (unlike its negation). Yet, if Openness is true, presumably there are open possibilities where the 2022 World Cup takes place elsewhere, or doesn't take place at all.

This is a problem for all views of the future, not just ours. Nevertheless, we want to explore how it can be solved within our analysis. We briefly present two strategies. None of them succeeds as it stands, but perhaps one of them can be developed into a viable solution.

Strategy 1 allows that the modal base may be restricted to a subset of the historical alternatives to the world of evaluation. Depending on how it is sharpened, this proposal suffers from two potential problems. First, it might overgenerate. If we allow that the modal base might be restricted in this way, we risk predicting that an utterance of "The coin will land tails" concerning a genuinely open coin toss is perfectly appropriate—provided that the speaker intends to leave heads-worlds out of the modal base. This seems wrong. One might try to fix this by limiting the worlds one may leave out to low-probability or far-off worlds. Even with this patch, the proposal suffers from a second problem, i.e. it makes some wrong truth value predictions. Suppose that you utter (19) leaving out of the modal base non-Qatar-worlds, but that a chain of fluky events brings it about that the World Cup takes place in Iceland instead. In this case, your utterance is predicted to be true; but this seems incorrect.

Strategy 2 consists in weakening the link between assertability and truth. Following Stalnaker's account of assertion [23], think of assertion as a proposal to narrow down a set of worlds that we regard as live epistemic possibilities (roughly, what Stalnaker calls 'context set'). In general, the purpose of narrowing down the context set is locating the actual world with greater precision. But, if Openness is correct, there is no such thing as the actual world—rather, there are a number of equally viable candidates, and it unsettled which of them counts as the actual world. In this situation, making a *will*-claim may be seen as akin to placing a bet that the actual world will be in a subset of these candidates (the odds related to this bet may be given by the probability attaching to the relevant claim). Of course, this strategy needs to be developed in a rigorous fashion to be viable. In addition, since it doesn't connect the assertability of *will*-claims to any element that is syntactically localized in the modal *will*, it incurs another burden: it should explain why present tense sentences with a future reference time (like "The 2022 World Cup takes place in Qatar") are not assertable in the same range of circumstances as the corresponding *will*-claims.

---

[8] PROOF: $v \in \|\mathit{will}_f\ \mathsf{A}\|$ iff $[\![\mathit{will}_f\ \mathsf{A}]\!]^{v,s,g} = 1$ iff $[\![\mathsf{A}]\!]^{s(v,\mathbf{F}_v),s,g[f\to\mathbf{F}]} = 1$ iff $v \in \|\mathsf{A}\|$. The first equivalence follows from our definition of content; the second from the truth-conditions of *will*, the third from $s(v,\mathbf{F}_v) = v$, which in turn follows from centering and $v \in \mathbf{F}_v$.

References

[1] Dorit Abusch. Sequence of tense and temporal de re. *Linguistics and Philosophy*, 20(1):1–50, 1997.

[2] Elizabeth Barnes and Ross Cameron. The open future: bivalence, determinism and ontology. *Philosophical Studies*, 146:291–309, 2009.

[3] Nuel Belnap and Mitchell Green. Indeterminism and the thin red line. *Philosophical perspectives*, 3:365–388, 1994.

[4] Fabrizio Cariani and Paolo Santorio. *Will* done better: Selection semantics, future credence, and indeterminacy. Manuscript, Northwestern University and University of Leeds, 2015.

[5] Cleo Condoravdi. Temporal interpretation of modals: Modals for the present and for the past. In D. Beaver, L. Casillas, B. Clark, and S. Kaufmann, editors, *The Construction of Meaning*. CSLI Publications, Palo Alto, CA, 2002.

[6] Bridget Copley. *The Semantics of the Future*. Routledge, 2009.

[7] Kai von Fintel. Bare plurals, bare conditionals and *Only*. *Journal of Semantics*, 14(1), 1997.

[8] Kai von Fintel. Npi licensing, strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148, 1999.

[9] Bart Geurts. *Presuppositions and Pronouns*. Elsevier, 1999.

[10] Anastasia Giannakidou and Alda Mari. Predicting the future in greek and italian: objective and subjective dimensions. Manuscript, University of Chicago and Institut Jean Nicod, CNRS, 2015.

[11] Andrea Iacona. Ockhamism without the thin red line. *Synthese*, 191:2633–2652, 2014.

[12] David Kaplan. Demonstratives. In J. Almog, J. Perry, and H. Wettstein, editors, *Themes from Kaplan*. Oxford University Press, 1989.

[13] Stefan Kaufmann. Conditional truth and future reference. *Journal of Semantics*, 22(3):231–280, 2005.

[14] Peter Klecha. Diagnosing modality in predictive expressions. *Journal of Semantics*, 31(3):443–455, 2013.

[15] Angelika Kratzer. The notional category of modality. In B. Partee and P. Portner, editors, *Formal Semantics: the Essential Readings*. Blackwell, 1981.

[16] Angelika Kratzer. *Modals and Conditionals*. Oxford University Press, 2012.

[17] David Lewis. Counterfactual dependence and time's arrow. *Noûs*, 13(4):455–476, 1979.

[18] John MacFarlane. Future contingents and relative truth. *Philosophical Quarterly*, 53:321–336, 2003.

[19] John MacFarlane. Truth in the garden of forking paths. In M. Kölbel and M. Garcia-Carpintero, editors, *Relative Truth*. Oxford: Oxford University Press, 2008.

[20] John MacFarlane. *Assessment Sensitivity*. Oxford University Press, 2014.

[21] Vann McGee and Brian McLaughlin. Distinctions without a difference. *The Southern Journal of Philosophy*, 33(S1):203–251, 1995.

[22] F. R. Palmer. *Mood and Modality*. Cambridge University Press, 1987.

[23] Robert Stalnaker. Assertion. In *Context and Content*. Oxford University Press, 1978.

[24] Robert C Stalnaker. A theory of conditionals. *Studies in Logical Theory, American Philosophical Quarterly*, Monograph Series 2:98–112, 1968.

[25] Richmond H. Thomason. Indeterminist time and truth-value gaps. *Theoria*, 18(3):264–281, 1970.

[26] Richmond H Thomason. Combinations of tense and modality. In D. Gabbay and F. Guenthner, editors, *Handbook of philosophical logic (Volume II: Extensions of Classical Logic)*, pages 135–165. Springer, 1984.

[27] Arnim Von Stechow. On the proper treatment of tense. In *Proceedings of Semantics and Linguistic Theory 5*, 1995.

Divine foreknowledge, time and tense

Pablo Cobreros*

Department of Philosophy
University of Navarra, Pamplona, Spain
`pcobreros@unav.es`

If God's omniscience entails knowing all things, including those that did not occur yet, how is it possible that humans act freely? This very much discussed old question had a sudden revival with Nelson Pike's paper fifty years ago. In this paper we provide an analysis of the argument for "theological fatalism" under the light of some assumptions about the structure of time and the semantics of tensed sentences.

We present, in particular, Prior-Thomason semantics for indeterminist time (second section). This semantics motivates the distinction between time of evaluation and *perspective* which, we argue, is required for an appropriate definition a truth-predicate in the context Prior-Thomason semantics. Third section shows how the previous language and semantics can be used to formalize the argument for theological fatalism. The argument thus formalized is quite robust and we argue that the only way to scape its conclusion makes essential use of the distinction between time of evaluation and perspective which was independently motivated in the previous section. We intend to show that this solution to the argument is a precise way to implement the *God as timeless* solution, making this a live option in this debate.

1 The argument for theological fatalism

Suppose Jones mowed his lawn last Saturday and God foreknew that. Then, at some point before Saturday, say on Thursday, God believes that Jones will mow his lawn on Saturday. Then Jones' ability to refrain from mowing his lawn before Saturday, say on Friday, is either (a) the ability of making God having a false belief or (b) the ability to influence on someone's past beliefs or (c) the ability to turn into non-existence someone who existed in the past. Neither of (a) to (c) describes a real ability of Jones'. Therefore either Jones does not have the ability to refrain from mowing his lawn on Saturday or God does not foreknow that Jones will mow his lawn on Saturday (see Nelson Pike's [14]).

Pike's argument share a strong affinity with previous arguments on the incompatibility of foreknowledge and future contingency and arguments about the incompatibility of truth and future contingency [17, 113-7]. We formulate the argument making use of the following abbreviations,

J = Jones mows his lawn
$\boxdot A$ = It is *accidentally necessary* that A
$\Box A$ = It is necessary that A
$[G]A$ = God believes A

$t_n : A = A$ occurred in time t_n

Accidental necessity corresponds to the idea of the fixity of the past: it's not necessary that Caesar crossed the Rubicon but once he did, there's nothing we can do today to change that it happened. The new version of the argument makes use, in addition, of the following principles (some extra parentheses added for readability):

NP $t_n : A \vDash \boxdot(t_n : A)$ (for any $n < p$ where t_p is present time)

EO $\Box(t_n : [G]A \supset A)$ (for any n)

TN $\boxdot A,\ \Box(A \supset B) \vDash \boxdot B$

The first principle NP expresses the "necessity of the past": that anything that already occurred is accidentally necessary (that occurred). The second, EO, expresses the (tensed) essential omniscience of God: if God believes A (at any time n) then A. The third, TN, is the principle of *transfer of necessity* according to which if A is accidentally necessary and necessarily if A then B, it follows that B is also accidentally necessary. Making use of these principles, the reformulation of Pike's argument run as follows:[1]

1. $t_0 : [G](t_2 : J)$ [Assumption]

2. $\boxdot(t_0 : [G](t_2 : J))$ [From 1 by NP]

3. $\Box(t_0 : [G](t_2 : J) \supset t_2 : J)$ [Instance of EI]

4. $\boxdot t_2 : J$ [From 2, 3 by TN]

Now if it is accidentally necessary that Jones will mow his lawn at t_2, then he cannot do otherwise.

2   Time: structure and semantics

In this section we talk about the structure of time and the semantics for tensed sentences. We make a number of choices about these questions corresponding, mostly, to the semantics and logics defended in [20].[2] We will give a reason for each choice we make although we don't intend to settle the question about each such issue.

---

[1] The argument in the text is a formalization of steps (1) to (6) of the argument in [22, sec. 1]. The argument seems to go back to, at least, the American philosopher Jonathan Edwards as reported by Prior in [17, 113-4].

[2] I discovered, after writing much of this paper, that [4] also make use of the formal framework of Thomason to discuss the question of foreknowledge. Although we share strong affinities in the philosophical background, the discussions are quite different in, at least, two respects. We both add further structure to Thomason's semantics, but they add a dynamic view of models that interact with a "NOW" operator whereas I intend just to see how to define a truth-predicate coherent with Thomason's semantics. Second, the scope of the paper is different for, whereas theirs discuss a broader range of topics, this paper restricts the attention to Pike's argument.

The language we will be dealing with is a propositional language containing classical connectives $\neg, \vee, \wedge$ and $\supset$ and tense operators $\langle + \rangle, \langle - \rangle, [-]$ and $[+]$. $\langle + \rangle A$: informally reads "it will be the case that A", that is, "A takes place at some future instant" and similarly for the other tense operators.

An interpretation a structure $\langle T, <, [\![\cdot]\!] \rangle$ where,

$T \not\equiv \emptyset$

$<$ is a tree-order on T

$[\![\cdot]\!]$ is a function : $Var \times T \longrightarrow \{0, 1\}$

A (strict) partial order $<$ on T is a relation between the elements of T that is irreflexive and transitive. Such an order is a tree-order on T if, in addition, for any t, t' and t'' if $t' < t$ and $t'' < t$ then either $t' < t''$ or $t'' < t'$. This last condition amounts to the idea that a tree-order is a partial order that is "linear to the left". This idea, in turn, seem to capture the intuitive asymmetry between past and future: that for any $t \in T$ there is a set H_t of histories containing t; these histories agree up to t and possibly disagree after t. (A *history* is a subset $T' \subseteq T$ that is linearly ordered by $<$; what it is sometimes called a *maximal chain*.) The tree-order, therefore, seem to rule out *Ockhamist solutions*.[3]

The function $[\![\cdot]\!]$ is a bivalent assignment of truth-values to propositional variables relative to each $t \in T$. We need to explain now how a given interpretation extends to complex formulas. Given that the same time t might belong to multiple histories, the truth-conditions for a formula A will be relative, not just to a given time t but also to a given history h such that $t \in h$:

- $[\![\neg A]\!]_t^h = 1$ just in case $[\![A]\!]_t^h = 0$

- $[\![A \wedge B]\!]_t^h = 1$ just in case $[\![A]\!]_t^h = [\![B]\!]_t^h = 1$

- $[\![\langle - \rangle A]\!]_t^h = 1$ just in case $\exists \mathtt{t} \in h$ such that $\mathtt{t} < t$ and $[\![A]\!]_{\mathtt{t}}^h = 1$

- $[\![\langle + \rangle A]\!]_t^h = 1$ just in case $\exists \mathtt{t} \in h$ such that $t < \mathtt{t}$ and $[\![A]\!]_{\mathtt{t}}^h = 1$

This semantics works fine for either classical or '$\langle - \rangle$' operators but there is an issue with '$\langle + \rangle$' formulas. Since a tree-order might be *non-linear to the right*, a formula '$\langle + \rangle$A' might receive different truth-value relative to different histories. Now if '$\langle + \rangle A$' is true in t relative to history h [$t \in h$] and false relative to h' [$t \in h'$], what is the final truth-value of '$\langle + \rangle A$' in t? Thomason's strategy [20, 272] is considering that, in such a case, the formula is neither true nor false.[4]

A *supervaluation* is a partial valuation based on a set of complete valuations. A time t determines a set H_t of histories (the set of histories that *pass over* t). Each history $h \in \mathsf{H}_t$, in turn, provides a complete valuation for formulas of the language (including '$\langle + \rangle$' formulas) relative to time t. For this reason, the tree-like structure of time is a natural context to define a supervaluation:

---

[3]See [18], [15], [1], [7], [9], [16], [6].

[4]Alternatively, we could consider that it is *both true and false*. The dual theory of supervaluationism has precisely this effect; see [3] for an introduction to subvaluationism.

A is *supertrue* at time t just in case for all $h \in \mathsf{H}_t$, $[\![A]\!]_t^h = 1$.

In general, logical consequence is a matter of necessary preservation of truth. Since our relevant notion of truth is that of *supertruth*, logical consequence is defined accordingly (the subscript PT is for *Prior-Thomason tense* logic):

$\Gamma \vDash_{PT} A$ just in case, there is no interpretation $\langle T, <, [\![\cdot]\!] \rangle$ with $t \in T$ such that:

$$\forall h \in \mathsf{H}_t\ \forall B \in \Gamma,\ [\![B]\!]_t^h = 1 \quad \wedge \quad \exists h \in \mathsf{H}_t\ [\![A]\!]_t^h = 0$$

In words, an argument is valid just in case there is no interpretation and time t such that all premises are supertrue and the conclusion is not.

A characteristic feature of supervaluationism in general is that it makes compatible classical logic with truth-value gaps.[5] In the case of Prior-Thomason's temporal logic supervaluationism comes with some additional validities,[6] one of which is the following inference:

$A \vDash_{PT} [-]\langle+\rangle A$

For suppose A is true in actual time t. Then $\langle+\rangle A$ is true in any time $t' < t$ of any history $h \in \mathsf{H}_t$; that is, $\langle+\rangle A$ is supertrue at any such time t' (relative to the set of histories H_t).

Note that the validity of this inference involves a strong form of truth-relativism. It is well known that the truth of a sentence might be relative to a context of utterance. The sentence 'I am a smart philosopher' might be false when uttered by me but true when uttered by you. The reason for this shift in truth-value is that part of the content of the sentence is fixed by contextual factors and, therefore, each utterance of the sentence expresses a different content. The form of truth-relativism involved in the validity the inference above, however, is more radical. If on Friday I utter 'Jones will mown his lawn tomorrow', what the sentence says is on Friday and from my perspective on Friday, untrue. If on Saturday Jones mows his lawn, what the sentence says is true on Friday, from my perspective on Saturday.

This form relativity is, in fact, what invokes Thomason to distinguish the notion of *supertruth* from the notion *unavoidability*. Given the previous structure of time and the semantics for tensed sentences, it is natural to define the notion of *unavoidability* this way [20, 275] (where '$\mathbb{U}A$' stands for 'A is unavoidable'):

$$[\![\mathbb{U}A]\!]_t^h = 1 \text{ just in case } \forall h' \in \mathsf{H}_t\ [\![A]\!]_t^{h'} = 1$$

As Thomason points out, under this definition *truth* seems to collapse with *unavoidability* since the definition of $\mathbb{U}$ seems to mirror in the object language the definition above of *supertruth*. Furthermore, the following pair of inferences are valid:

- $\mathbb{U}A \vDash_{PT} A$

- $A \vDash_{PT} \mathbb{U}A$[7]

---

[5]In addition to Thomason's paper [21] and [5] are classical examples of the application of supervaluationism in different contexts. See [10] for a more contemporary defense of supervaluationism.

[6]The validity of $\langle+\rangle A \vee \langle+\rangle \neg A$ is a particularly nice example.

[7]The inferences don't entail the triviality of $\mathbb{U}$ because of a failure of the deduction theorem; in particular, $\nvDash A \supset \mathbb{U}A$.

Thomason, however, argues that, despite initial appearances, truth and *unavoidability* (or *inevitability* as he says) are different. He seeks to show this difference defining a new operator $\mathbb{T}$ for *truth*:

$$[\![\mathbb{T}A]\!]_t^h = 1 \text{ just in case } [\![A]\!]_t^h = 1 \qquad [20, 278]$$

This is a transparent truth-predicate, allowing for full substitutivity between A and '$\mathbb{T}A$'. This means, among other things, that the inference: $A \vDash_{PT} [-]\mathbb{T}\langle+\rangle A$ remains valid. However, the corresponding inference involving $\mathbb{U}$: $A \nvDash_{PT} [-]\mathbb{U}\langle+\rangle A$ is not. With this different logical behavior, Thomason nicely puts the subtle difference between truth and unavoidability:

> Our theory thus allows (indeed forces) us to say that *having been true* is different from *having been inevitable*, as far as future-tensed statements go. The latter is not a consequence of the former, $[-]\mathbb{T}\langle+\rangle A \nvDash_T [-]\mathbb{U}\langle+\rangle A$, because in an assertion that it was *true* that a thing would come about, truth is relative to events up to the present, whereas in an assertion that it was *inevitable* that a thing would come about, inevitability is judged relative to some time in the past. (p. 279)

I think this explanation hits the nail on the head. I don't think, however, the previous characterization of the notion of truth (the definition of $\mathbb{T}$) is coherent with the full story. The definition has the advantage of substitutivity, which is often considered a desiderata for a theory of truth,[8] but it betrays the original spirit of supervaluationism as motivated by truth-value gaps. The notion of truth in the semantics (the notion of *supertruth*) employed to handle the issue of indeterminist time, allows for truth-value gaps. It would then be reasonable to think that if $\langle+\rangle A$ is neither supertrue nor superfalse, the statement '$\langle+\rangle A$ is supertrue' is false. However, '$\mathbb{T}\langle+\rangle A$' and '$\langle+\rangle A$' share identical truth-conditions in the above characterization and, therefore, '$\mathbb{T}\langle+\rangle A$' is neither supertrue nor superfalse if '$\langle+\rangle A$' is. In short, $\mathbb{T}$ is not an object language expression of *supertruth*.[9]

Despite this fact, I think the explanation of the difference between truth and unavoidability in the quotation above is correct. The driving idea is the following. The evaluation, in a given time, of a sentence containing tense operators requires moving forwards or backwards along the time structure. If times are linearly ordered, there is a single relevant history and the "movement" in search of times for evaluation reduces always to that history. If time has a tree-like structure, the point at which we start the evaluation, what we might call the *perspective*, determines the histories relevant for the evaluation of the sentence. The sentence "Jones will mow his lawn" is neither true nor false on Friday, when Friday is the starting point of evaluation (when Friday is the perspective). The sentence "Jones will mow his lawn" is true on Friday, when Saturday is the perspective.

The idea can be expressed in a more formal style saying that, in tree-like structures, the evaluation of tensed sentences involves a double reference to times[10] One of the reference times

---

[8]See [12]

[9]Perhaps Thomason is aware of this question and that's the reason why he speaks of formalizing a locution '... *was* true' (italics mine) intending to restrict the application of the truth predicate $\mathbb{T}$ only within the scope of a past tense operator.

[10]This idea of double-time-reference is explicitly endorsed and defended in [2] and [13]. These authors, however, reject the supervaluationist treatment of future contingent statements.

is the time at which we evaluate a given sentence (what we will call the *evaluation time*), the second reference time is the perspective.

In the previous semantics, we made explicit reference to the evaluation time, but we didn't do the same with the perspective. The reason, I think, is that we tend to take for granted that the perspective, the time at which the evaluation process begins, is identical with the time at which the sentence under evaluation is uttered, the time of assertion (see [13]). This makes perfect sense. As temporal beings, our actions, and particularly speech acts like assertions, take place in a particular time and are therefore connected to a particular perspective. It is not unthinkable, however, that perspective and time of assertion be different, as we shall point out later.

Given the previous remarks, we redefine the truth predicate as relative to a perspective and also redefine unavoidability accordingly.

$\llbracket \mathbb{T}A \rrbracket^h_{t\,[p]} = 1$ just in case $\forall h^* \in \mathsf{H}_p \; \llbracket A \rrbracket^{h^*}_{t\,[p]} = 1$

$\llbracket \mathbb{U}A \rrbracket^h_{t\,[p]} = 1$ just in case $\forall h^* \forall p^* (t \leq p^* \wedge h^* \in \mathsf{H}_{p^*} \supset \llbracket A \rrbracket^{h^*}_{t\,[p^*]} = 1)$.

The intended meaning of this truth operator is the following: a sentence A is true, relative to a given perspective p, just in case A is "settled" according to that perspective. Following our example, the statement 'It is true that Jones will mow his lawn' is *false* from Friday's perspective but true from Saturday's perspective. The definition of *unavoidability* intends to register the fact that, unlike truth, this notion is relative to the time of evaluation (as Thomason points out 'inevitability is judged relative to some time in the past.') and hence the predicate takes into account all histories from the time of evaluation on. $\mathbb{T}$ and $\mathbb{U}$ are certainly similar in that both are defined relative to a set of histories, but might differ in exactly what set. That set is guaranteed to be the same when time of evaluation and perspective coincide but not otherwise.

Summing up, we have the following properties of time and tensed sentences that might help our analysis of the problem of divine foreknowledge. First, time need not be linearly ordered but have a tree-structure; this fact will help us to understand the ideas of the fixity of the past versus the openness of the future. A sensible approach to the open future is provided by Thomason's supervaluations about histories; in this approach a proposition might be untrue in time t under a given perspective but true in the same time t under some other perspective. An adequate semantics for a tensed language including a truth predicate, therefore, should appropriately distinguish between the time of evaluation of a given sentence and the perspective. With this ingredients we also redefined the notion of *unavoidability* as linked to the time of evaluation.

3   Foreknowledge and free will

The argument for theological fatalism, as stated above, has the following shape,

1. $t_0 : [G](t_2 : J)$ [Assumption]

2. $\boxdot(t_0 : [G](t_2 : J))$ [From 1 by NP]

3. $\Box(t_0 : [G](t_2 : J) \supset t_2 : J)$ [Instance of EI]

4. $\boxdot t_2 : J$ [From 2, 3 by TN]

The principles used in the argument correspond to *necessity of the past*, God's *essential omniscience* and the *transfer of necessity*,

NP $t_n : A \vDash \boxdot(t_n : A)$ (for any $n < p$ where t_p is present time)

EI $\Box(t_n : [G]A \supset A)$ (for any n)

TN $\boxdot A$, $\Box(A \supset B) \vDash \boxdot B$

The first target of this section is adapting the argument to the language employed in section 2. The issue has its difficulty because the argument, as stated above, makes use of indexes for times and that is something that brings us beyond the expressivity of a simple modal language. I propose considering just two times: a present time (the time of God's beliefs about Jones) and a future time (the time at which Jones mows his lawn). In this way we can make use of the simple future tense in order to talk about Jones' future action.

1. $[G]\langle+\rangle J$ [Assumption]

2. $\boxdot([G]\langle+\rangle J)$ [From 1 by NP]

3. $\Box([G](\langle+\rangle J) \supset \langle+\rangle J)$ [Instance of EI]

4. $\boxdot\langle+\rangle J$ [From 2, 3 by TN]

The next task is finding an appropriate characterization of the principles used in the argument, that is, finding appropriate interpretation for '$\Box$' and '$\boxdot$'. The first, I think, is more straightforward. $\Box$ is invoked as a strong form of necessity (Pike claims that the necessity in question is a form of analyticity [14, 35]). In the context of branching time this amounts to the idea that $\Box A$ is true at time t and history h just in case it is true *everywhere* (at any time of any history). This is, at least, the strength of analytic necessity.

The case of $\boxdot$ is harder. It is generally agreed in discussions concerning Pike's argument that at least part of the problem is how can we, with actions in present time, be able to change facts about the past. So the idea seems to be: if something happened, then it does not matter how the world evolves, it won't be the case that it didn't happen. This is partially captured by the inference:

$A \vDash [+]\langle-\rangle A$

which informally reads: if A is true, then it will always be true that A was true. The proposal is then reading accidental necessity in this way:

$\boxdot A =_{df} [+]\langle-\rangle A$

Now this interpretation is not fully faithful to the idea of the necessity of the past in a, I want to argue, harmless way. The idea of the necessity of the past involves reference to a particular time: if A took place at time t in the past, then it will always be the case that A took

place **at time** t. The inference above only states that if A is true, it will always be true that it was (but not necessarily at the same past time). This difference rests crucially in a expressive limitation of our simple modal language; the difference, however, is harmless in the sense that we can add some information to reach a similar effect to that of naming times. Suppose for example that I want to express that it is accidentally necessary that Jones mown his lawn (due to the fact that Jones is now doing that). Then I can write $\boxdot J \wedge \neg\langle-\rangle J$. In this case, if $\boxdot J$ is true, that is due to something that is happening today. We can see this strategy as a way to force J refer to the relevant fact (a fact that is happening today) without explicit naming of times.

Given the above qualifications, the final shape of the argument is the following,

0. $\neg\langle-\rangle J \wedge \neg J$

1. $[G]\langle+\rangle J$ [Assumption]

2. $\boxdot([G]\langle+\rangle J)$ [From 1 by NP]

3. $\Box([G](\langle+\rangle J) \supset \langle+\rangle J)$ [Instance of EI]

4. $\boxdot\langle+\rangle J$ [From 2, 3 by TN]

5. $\mathbb{U}\langle+\rangle J$ [From 0 and 4]

We added premise 0 for the reasons just given; in that way, we guarantee that if $\langle+\rangle J$ is accidentally necessary, that is due to something that is happening in the immediate future (and not, for example, because J already occurred in the past). Steps 1 to 4 are based on a rewriting of the principles above according to our previous remarks on $\Box$ and $\boxdot$,

NP $A \vDash [+]\langle-\rangle A$

EI $\vDash \Box([G]A \supset A)$

TN $[+]\langle-\rangle A,\ \Box(A \supset B) \vDash [+]\langle-\rangle B$

The principle of *necessity of the past* is valid, so that proposition (1): 'God knows J', entails proposition (2): 'it will always be true that God knew J'. Our interpretation of $\Box$ as *true everywhere* is strong enough to guarantee the validity of *transfer of necessity*, since the truth of $[+]\langle-\rangle A$ and the falsity of $[+]\langle-\rangle B$ at the same time t require some time t' where A is true and B false, contrary to the truth of $\Box(A \supset B)$.[11] The final step, from 0 and 4 to 5 is, once again, valid.[12] Informally, if J is not true today, nor in any past time (assumption 0: $\neg\langle-\rangle J \wedge \neg J$), then that $[+]\langle-\rangle\langle+\rangle J$ is supertrue means that all future histories contain one

---

[11]Here is a more detailed explanation. Suppose there is some time t at which $[+]\langle-\rangle A$ and $\Box(A \supset B)$ are supertrue but $[+]\langle-\rangle B$ is superfalse. According to the last, for any history $h \in \mathsf{H}_t$ there is some time $t < t'$ such that $\langle-\rangle B$ is false in t'; this last in turn means that B is false at all times prior to t'. Call h^* to any such history. Since $[+]\langle-\rangle A$ is supertrue at t, it is true, in particular, relative to history h^*. Therefore, at time t' $\langle-\rangle A$ is true and this requires a time $t'' < t'$ such that A is true in t''. But by what we said above B is false at t'' and, therefore, the conditional $A \supset B$ is also false at t'', contrary to the assumption that $\Box(A \supset B)$.

[12]The validity of this step rests on Thomason's supervaluationist version of Prior's logic; that inference might not be valid without the assumption that the property preserved by logical consequence is *supertruth*.

immediately succeeding time where J, in which case $\mathbb{U}\langle+\rangle J$, it is unavoidable that Jones will mow his lawn.

If we endorse Thomason's semantics from section 2 and the interpretation given for $\Box$ and $\boxdot$, then, as far as I can see, the only way left to question the soundness of the argument is questioning the characterization of God's *essential omniscience* EO. This, I think, can be done on the grounds of the relativity of truth to a perspective.

If what we called the perspective coincides with the time of assertion then we have $\vDash \mathbb{T}A \supset A$ (and also that $\mathbb{T}A \vDash A$). We might view this inference expressing the idea that truth is *factive*: if A is true, then A is a fact. Factivity, in turn, entails unavoidability: $A \vDash \mathbb{U}A$. Now the validity of factivity of truth rests crucially in the idea that the perspective of truth coincides with the perspective at the time of assertion, to which logical consequence is connected. But we can think on a situation where the truth-predicate and the time of assertion do not share the same perspective, that of time travel. When old Tannen gives the Sports Almanac to young Tannen in 1954 what he asserts is true, but not relative to young Tannen's perspective (not relative to the time of assertion, which is some point in 1954).[13] Old Tannen insist that it is true that the UCLA will win the mach, but this does not entail that it is *a fact* (that it is true relative to that point in 1954). In short, the inference 'It is true that A therefore A' is guaranteed to hold when the perspective of truth coincides with the perspective of assertion (implicit in logical consequence) but not if it is some future perspective, like that of a time-traveller. I want to suggest that this is the situation when we consider God's knowledge about the (our) future. God's knowledge of A, when A refers to some future event relative to us, need not entail A (as stated in EO above) but only that A is true (true, relative to the divine perspective).

- $\vDash [G]A \supset \mathbb{T}A$

The idea of truth relative to a perspective provides a nice way to characterize the difference between truth and unavoidability. From a given perspective, it is true on Friday that Jones will mow his lawn and, still, it is not unavoidable on Friday that Jones will mow his lawn. This difference makes room for the idea of God's knowing that Jones will mow his lawn, making true the statement without making the corresponding action unavoidable. We want to stress that the relativity of truth to a perspective is not an *ad hoc* maneuver to address Pike's argument, rather, it emerges from independent considerations about the logic and semantics of tensed sentences.

The considerations so far leave still open much questions about the nature of time and its connection to the truth of tensed sentences. I want to point out just that the foregoing picture about God's omniscience is congenial with some classical ideas about why God's knowledge of our future does not determine contingent facts.

According to some authors like Anselm, Boethius, Aquinas and Schleiermacher (see [14, 29]) God's eternality should be understood as God's existing "outside time" so that it cannot be properly said that God knows *now* what will happen. The problem about God's foreknowledge and determinism, therefore, is rooted to the fact that it cannot be properly said that God *fore*knows the future, that has some knowledge in advance, because that would be mistakenly attributing temporal properties to God. I think this line of thinking is attractive and it was very popular in the past, but it lost support in the recent debate ([8] catalogs this as one of

---

[13]The example comes from the second part of Robert Zemeckis' *Back to the future*.

the "minor" solutions to the problem).[14] I see a fundamental reason for this decline: the view of God as timeless seems to make impossible almost all, if not simply all, human theological discourse. As temporal beings, we cannot make assertions outside time and we make assertions with the intention of reaching the truth. If we say "God is Holy" we would be mistakenly attributing to God holiness today. The foregoing remarks about relative truth and omniscience may serve, I think, to reconcile God's timelessness with the temporal dimension of our acts in general and with assertions in particular. God does not properly *fore*knows what will happen, He knows them from a peculiar perspective. But we can consistently say that God knows today that Jones will mow his lawn, the sense in which God knows *in advance* being just a reflection of our limited perspective.

References

[1] M. Adams. Is the existence of God a 'hard' fact? *Philosophical Review*, 76:492–503, 1967.

[2] N. Belnap and M. Green. Indeterminist and the thin red line. *Philosophical Perspectives*, 8:365–388, 1994.

[3] Pablo Cobreros. Vagueness: Subvaluationism. *Philosophy Compass*, 8(5):472–485, 2013.

[4] C. De Florio and A. Frigerio. In defense of the timeless solution to the problem of human free will and divine foreknowledge. *International Journal for Philosophy of Religion*, 78:5–28, 2015.

[5] K. Fine. Vagueness, truth and logic. *Synthese*, 1975.

[6] John Martin Fischer. Ockhamism. *The Philosophical Review*, 94(1):81–100, 1985.

[7] J. M Fisher. Freedom and foreknowledge. *Philosophical Review*, 92:67–79, 1983.

[8] William Hasker. The foreknowledge conundrum. In E. Th. Long, editor, *Issues in Contemporary Philosophy of Religion*, volume 50, pages 97–114. Springer, 2001.

[9] Joshua Hoffman and Gary Rosenkrantz. Hard and soft facts. *The Philosophical Review*, 93(3):419–434, 1984.

[10] R. Keefe. *Theories of Vagueness*. Cambridge University Press, 2000.

[11] B. Leftow. *Time and Eternity*. Ithaca, NY: Cornell University Press, 1991.

[12] H. Leitgeb. What theories of truth should be like (but cannot be). *Philosophy Compass*, 2(2):276–290, 2007.

[13] M. Perloff and N. Belnap. Future contingents and the battle tomorrow. *The Review of Metaphysics*, 64:581–602, 2011.

[14] N. Pike. Divine omniscience and voluntary action. *Philosophical Review*, 74(1):27–46, 1965.

[15] Nelson Pike. Of god and freedom: A rejoinder. *The Philosophical Review*, 75(3):369–379, 1966.

[16] Nelson Pike. Fischer on freedom and foreknowledge. *The Philosophical Review*, 93(4):599–614, 1984.

[17] A. Prior. *Past, Present and Future*. Oxford at the Claredon Press, 1967.

[18] John Turk Saunders. Of god and freedom. *The Philosophical Review*, 75(2):219–225, 1966.

[19] Eleonore Stump and Norman Kretzmann. Eternity. *The Journal of Philosophy*, 78(8):429–458, 1981.

[20] R. Thomason. Indeterminist time and truth-value gaps. *Theoria*, 36(3):264–281, 1970.

[21] B. C. van Fraassen. Singular terms, truth-value gaps and free logic. *Journal of Philosophy*, 1966.

[22] L. Zagzebski. Foreknowledge and free will. *Stanford Encyclopedia of Philosophy*, 2011.

---

[14]See, however, [19], [11] and [4].

Knowledge, Justification and Reason-Based Belief *

Paul Egré[1][†] Paul Marty[2] and Bryan Renne[3][‡]

[1] Institut Jean-Nicod, CNRS, ENS, PSL
`paul.egre@ens.fr`
[2] MIT, Linguistics Department
`pmarty@mit.edu`
[3] ILLC, Amsterdam
`bryan@renne.org`

Abstract

Can the ordinary concept of knowledge be defined in terms of justified true belief ('JTB')? We argue that the Gettier cases need not invalidate every analysis of knowledge in terms of justified true belief, depending on how the notion of justification is understood. We present a framework for the representation of reason-based belief and use it to define a notion of true belief supported by adequate reasons. An adequate reason is characterized as being infallible, namely as a justification supporting only true propositions. We give a tentative definition of knowledge in those terms. We present two variants of our axiomatics for belief, which differ mostly on how subjective and accessible the notions of reason and support are taken to be. A more detailed version of our approach and its model theory appears in an expanded version of this paper.

1 Knowledge and Justified True Belief

Can the ordinary concept of knowledge be defined in terms of justified true belief ('JTB')? Since Gettier (1963), the answer to this question is widely considered negative. Gettier produced two cases convincingly suggesting that a belief can be true, justified, and yet fall short of knowledge.

Our point of departure in this paper is the following: even though we agree with the force of Gettier's cases (see Machery et al. 2015), we share with others (in particular Chisholm 1977; Dretske 1971; Goldman 1979; Sosa 1974, 1979; Turri 2012) the intuition that those examples do not invalidate every analysis of knowledge in terms of justified true belief, depending on how the notion of justification is understood. What Gettier's cases teach us is that an agent can have a justification for believing a proposition that is *plausible* on internal grounds, without that justification being properly *adequate* to the truth of the proposition in question. But if so, then Gettier cases only show that knowledge is not identical with JTB under an internalist conception of justification.

In the wake of work done in Justification Logic (Artemov 2008; Artemov and Fitting 2011; Baltag et al. 2014), we propose an explicit treatment of reasons for belief, which we use to tease

apart two notions of justified true belief, and to defend an externalist version of the equation between knowledge and JTB. The gist of our account lies in the distinction between reasons that *(merely) support* belief in a proposition and reasons that are not only supportive but are also what we call *adequate.* Fundamentally, we view a reason as adequate only if every proposition it supports is true. We thereby endorse a form of infallibilism about knowledge (see Dutant 2010, 2015 for similar views).

In this paper, we give mostly an outline of our system: we refer to Egré et al. (2015) for more details, both technical and philosophical. To make our paper informative relative to the extended version, we present two versions of the axiomatic treatment of the notion of reason-based belief. In the more extended version, we introduce only the second of those versions, the system QRBB, but we find it worthwhile to also present the first system we came up with, in order to better situate our approach. Below we also briefly mention two objections that can be made against our approach: (i) the objection that agents have knowledge only if every reason they have for a proposition is adequate, and (ii) the objection that there appears to be knowledge from false lemmas (Warfield 2005).

2 Reason-based belief

2.1 Motivations

The main motivation behind our approach is the observation that there are two kinds of justified true beliefs: on the one hand true propositions that are believed on the basis of good or adequate reasons, and on the other true propositions believed on the basis of bad or inadequate reasons. A Gettierized belief is when a proposition φ is true and believed on the basis of some reason that is not adequate. A Gettier-proof belief is when a proposition φ is true and believed on the basis of an adequate reason. To represent both notions, we need to quantify over reasons, and first of all to represent the notion of an adequate reason as well as the notion of believing a proposition on the basis of some supporting reason.

Adequacy, as we use it, is a term of art: instead of saying that a reason is adequate, we could say that a reason is *good*, or *infallible*, provided goodness again is understood in externalist terms, independently of how good the reason appears to the agent. An important remark is that the notion of adequacy of a reason is not relative to a specific proposition. This is an important difference with Dretske (1971)'s account of knowledge in terms of *conclusive* reasons. For Dretske, a reason is only conclusive *for* a given proposition. In our approach, reasons are adequate or not, without this explicit relativity to a specific proposition.

2.2 Syntax and quantificational axioms

The two systems of reason-based belief to be presented rest on the following syntax. F is the set of formulas φ defined by the following grammar, in which R is a set of reason symbols, and P a set of propositional symbols, both sets being disjoint:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid (r : \varphi) \mid r \mid B\varphi \mid r = r \mid (\forall r)\varphi$$
$$p \in P, \, r \in R$$

Other Boolean connectives we shall use are defined as abbreviations, and brackets are removed when no ambiguity results. The main innovation relative to a system of standard epistemic logic is the use of reason variables. $(r : \varphi)$ is shorthand for: 'reason r supports the proposition

φ'; r alone is shorthand for 'reason r is adequate'. The quantificational core of both systems involves the axioms in Table 1.

<div style="text-align:center">

QUANTIFICATIONAL AXIOMS

</div>

(UD) $(\forall r)(\varphi \to \psi) \to (\varphi \to (\forall r)\psi)$, where r is not free in φ
(UI) $(\forall r)\varphi \to \varphi[s/r]$, where s is free for r in φ
(EP) $r = r$
(EN) $\neg(r = s)$, where r and s are syntactically different

<div style="text-align:center">

QUANTIFICATIONAL RULE

$$\frac{\varphi}{(\forall r)\varphi} \text{ (Gen)}$$

</div>

Table 1. Quantificational axioms and rule

As an illustration of the syntax, consider the sentence: 'John believes that 2+2=4 on the basis of his calculations. We would represent this as: $B(r\!:\!p) \wedge Br$, letting p stand for '2+2=4' and r refer to John's calculating evidence. $B(r\!:\!p)$ says that John believes that r supports φ, that is John is inclined to believe p on the basis of r, and Br says that John actually endorses his reason or justification. We use the paraphrase "John is inclined to believe p on the basis of r" as another paraphrase for $B(r\!:\!p)$, since John may believe that r supports p without thinking of r as being an adequate reason.

Consider an agent, John, who concludes that 2+2=4 on the basis of his calculations, but who does not quite trust his calculating capacities; nevertheless, John would trust his teacher telling him that his calculations are correct. We may represent this by: $B(r\!:\!p) \wedge \neg Br \wedge \neg B\neg r$: John is inclined to believe that 2+2=4 on the basis of his calculations, but is not sure whether his calculations are correct. However, if s were the evidence of his friend confirming, John would believe that s supports that his calculations are adequate. We would then have: $B(s\!:\!r) \wedge Bs$, to mean that John believes that his teacher's testimony supports the adequacy of his calculations, and furthermore John believes that s is adequate, as a result of which John would believe r to be adequate.

Our account is close in spirit to Dutant's account of knowledge in terms of method-based belief (Dutant 2010, 2015). Dutant conceives of justification primarily as methods of belief formation, and we likewise think of reasons not fundamentally as propositions but as processes or experiences by which those propositions come to be believed.[1] There is, however, a noteworthy difference: semantically, we treat $(r\!:\!p)$ as a proposition while Dutant does not; instead, Dutant proposes to handle belief as a binary operator, involving a method-argument and a propositional argument. An objection that can be made to our approach is that it fails to separate the content from the method. We acknowledge this limitation of our system, though we believe that it is possible to impose an interpretation of our syntax compatible with Dutant's idea, namely such that the first reason-argument appearing in the scope of a belief operator is to refer to the method by which the beliefs are produced.

---

[1] We are indebted to J. Dutant for bringing us to endorse that view. In a preliminary version of this work, we left open the possibility that reason symbols could denote propositions directly.

2.3 The system RBBS

The first system we propose is a quantifier-free system of 'subjective' reason-based belief. We call it subjective because to say that a reason supports a proposition is to say that it is *ipso facto* a reason to believe that proposition, and conversely (see the axiom (IS) below). Axiom (IS) is very strong in ensuring that all reasons are accessible, and that if a reason is believed to support a proposition, it is thereby supportive. In this system, $r : \varphi$, which we said is shorthand for 'r supports φ', may therefore also be read as: 'r is a reason to believe φ'.

AXIOM SCHEMES

(CL)   Axiom Schemes of Classical Propositional Logic

(RK) $r : (\varphi \rightarrow \psi) \rightarrow (r : \varphi \rightarrow r : \psi)$
(A) $r : \varphi \rightarrow (r \rightarrow \varphi)$

(RB) $r : \varphi \rightarrow (Br \rightarrow B\varphi)$
(IS) $B(r : \varphi) \leftrightarrow (r : \varphi)$

(D) $B\varphi \rightarrow \neg B\neg\varphi$

RULES

$$\dfrac{\varphi \rightarrow \psi \quad \varphi}{\psi} \ (\text{MP}) \qquad \dfrac{\varphi}{r : \varphi} \ (\text{RN}) \qquad \dfrac{\varphi \leftrightarrow \psi}{B\varphi \leftrightarrow B\psi} \ (\text{E})$$

Table 2. The theory RBBS

As for mnemonics, (CL) is 'Classical Logic', (MP) is 'Modus Ponens', (RK) is Kripke's axiom K of modal logic (used here for reasons), (RN) is 'Reason Necessitation', (IS) is 'Internal Support', (A) is 'Adequacy,' (RB) is 'Reasons to Believe', (D) is a consistency requirement on belief, and (E) is a well-known rule from minimal modal logic (Chellas 1980).

(CL) and (MP) say that RBBS is an extension of classical propositional logic. (D) says that the agent's beliefs are consistent: the agent cannot have contradictory beliefs (i.e., believe both φ and $\neg\varphi$ for some φ). (E) says that the agent's beliefs do not distinguish between provably equivalent formulas. (RK) says that reasons are closed under material implication, and (RN) says that reasons support all derivable formulas.

(A) says that if r is a reason to believe φ and r is an adequate reason, then φ is true. (RB) says that if r is a reason to believe φ and the agent believes that r is an adequate reason, then the agent believes φ. (IS) says that an agent believes that r is a reason to believe φ if and only if r is a reason to believe φ. The following two closure properties are noteworthy consequences of (IS) and (RB) in our system:

(RS) $B(r : \varphi) \rightarrow (Br \rightarrow B\varphi)$
(RB+) $Br \rightarrow (r : \varphi \rightarrow B(r : \varphi))$

Technically, only the left-to-right direction of (IS) is needed to derive (RS) from (RB), but the biconditional version of (IS) also allows us to get (RB+) (directly, by weakening). (RB+) may appear implausibly strong, but so is (RB), and it appears natural to have both as soon as one of them is accepted.

2.4 The system RBB

<table>
<tr><td colspan="2" align="center">AXIOM SCHEMES</td></tr>
<tr><td>(CL)</td><td>Axiom Schemes of Classical Propositional Logic</td></tr>
<tr><td>(RK)</td><td>$r : (\varphi \to \psi) \to (r : \varphi \to r : \psi)$</td></tr>
<tr><td>(A)</td><td>$r : \varphi \to (r \to \varphi)$</td></tr>
<tr><td>(BRK)</td><td>$B(r : \varphi) \to (B(r : (\varphi \to \psi)) \to B(r : \psi))$</td></tr>
<tr><td>(BA)</td><td>$B(r : \varphi) \to (Br \to B\varphi)$</td></tr>
<tr><td>(AS)</td><td>$B(r : \varphi) \to (r \to (r : \varphi))$</td></tr>
<tr><td>(D)</td><td>$B\varphi \to \neg B\neg\varphi$</td></tr>
</table>

<div align="center">

RULES

$$\frac{\varphi \to \psi \quad \varphi}{\psi} \text{ (MP)} \qquad \frac{\varphi}{r : \varphi} \text{ (RN)} \qquad \frac{\varphi \leftrightarrow \psi}{B\varphi \leftrightarrow B\psi} \text{ (E)}$$

</div>

Table 3. The theory RBB

The system RBB is almost exactly like RBBS except for the three axioms (BRK), (BA), and (AS), which we have instead of (RB) and (IS). In RBBS, there is a single axiom constraining the notion of adequacy for reasons, namely axiom (A), and the two axioms (RB) and (IS) only concern the interaction between belief and support for reasons. In RBB, we have an additional axiom (AS) about adequacy, and the two axioms (BRK) and (BA) concern the interaction between belief and support for reasons.

RBB is not a 'subjective' system of reason-based belief since we give up both directions of (IS): a reason r can be a supporting reason for a proposition φ without being believed to support φ, and conversely, one can believe a reason r to support a proposition φ without there being support (we are thinking, here, of cases of delusion or hallucination). This means that in RBB the support relation is allowed to be independent of one's belief. An agent can incorrectly believe a reason to support a proposition. On the other hand, (AS) says that when an agent believes a reason to support a proposition, the reason can only be adequate provided the support relation holds indeed.

Another important difference between RBB and RBBS is that RBBS predicts stronger closure conditions on belief. In both RBB and RBBS, reasons are strong, in virtue of the closure axiom (RK) and of rule (RN): in particular, every reason supports every logical truth. Consider an agent who holds a belief that some reason is adequate, namely for whom Br holds. Then, because r supports every logical truth, it follows by (RB) that the agent believes every logical truth. An agent who believes a single reason to be adequate is thereby logically omniscient. This result is not welcome. In that regard, (BA) is a natural weakening of (RB).

For those various reasons, we think RBB is a better system than RBBS: it predicts fewer closure properties on belief, and it separates out a weak notion of belief from a strong notion of reason more neatly. In what follows, we therefore call QRBB the system that results from the combination of RBB with the quantificational axioms and rule of Table 3, and likewise, we call QRBBS the system resulting from RBBS. Irrespective of the choice between those systems, axiom (A) remains in a sense the central characterization of adequacy for reasons in our approach.

2.5 Semantics

We refer to Egré et al. (2015) for a systematic exploration of the system QRBB and of its model theory. Here, we only point out some essential facts about the interpretation of formulae. The models are structures $M = (W, [\cdot], N, V)$ where W is a nonempty set of possible worlds, N is a neighborhood function associating sets of worlds to each world, and $[\cdot]$ is a function mapping each reason $r \in R$ to a binary relation $[r] \subseteq W \times W$ on the set of possible worlds. Let $[r](w) := \{v \in W; w[r]v\}$, that is $[r](w)$ is the set of r-accessible worlds from w. The main clauses of the semantics are:

- $M, w \models B\varphi$ iff the set $[\![\varphi]\!]$ of φ-worlds belongs to $N(w)$.

- $M, w \models r$ iff $w \in [r](w)$.

- $M, w \models r : \varphi$ iff $[r](w) \subseteq [\![\varphi]\!]$.

- $M, w \models \forall r \varphi$ iff $M, w \models \varphi[s/r]$ for each s free for r in φ.

By imposing appropriate constraints on the function N and on the cardinality of R, the logic can be shown to be sound and complete for the semantics.

One aspect of the semantics worth pointing out is that reasons are handled by means of accessibility relations. Basically, a reason is adequate at a world w provided that w is accessible from itself via the relation, that is if the relation is reflexive at w. Thus, we model adequacy for reasons in a way that is similar to the way in which factivity is usually handled for knowledge when knowledge is taken as a primitive operator. In what follows, however, we propose to define knowledge as a special type of justified true belief, namely as true belief supported by adequate reasons.

3 Two notions of JTB

Whether in QRBB or QRBBS, there are (at least) two natural ways to define JTB, which we call 'external' vs. 'internal'.[2]

- $\mathsf{JTB}_r^e(\varphi) := B(r : \varphi) \wedge Br \wedge r$ (external JTB)

- $\mathsf{JTB}_r^i(\varphi) := B(r : \varphi) \wedge Br \wedge \varphi$ (internal JTB)

Both notions of JTB are factive, but the notions do not have the same properties. To illustrate the difference between the two cases, consider Gettier's case II: Smith wrongly believes Jones owns a Ford (p) on the basis of various plausible inductive evidence (represented by the symbol r), and 'realizes the entailment' (Gettier 1963) that either Jones owns a Ford or Brown is in Barcelona. As it turns out, Brown is in Barcelona (q). This is a case in which: $B(r : p) \wedge Br \wedge B(r : (p \rightarrow p \vee q)) \wedge q \wedge \neg p$. It follows in QRBB that the agent has $\mathsf{JTB}_r^i(p \vee q)$ without $\mathsf{JTB}_r^e(p \vee q)$, for by (A) the falsity of p implies $\neg r$.

Our analysis can be extended to cover more complex cases, such as Goldman-Ginet cases (Goldman 1976), in which an agent has a correct belief in a proposition, but in which the justification is arguably not adequate (because lucky). For instance, consider an agent traveling in the country of fake barns, and thinking of the only true barn that it is a barn (p), based on his visual experience. We can describe the situation as: $B(r : p) \wedge (r : p) \wedge Br \wedge p$, without assuming

---

[2]In QRBBS the definitions are the same, except that we can write $r : \phi$ instead of $B(r : \phi)$, due to axiom (IS).

r to hold, namely the reason to be adequate. This means that the agent's belief would be true, but not adequately true, given that it is lucky. We do not capture this connection between adequacy and luck in our axioms, however. This implies that axioms (A) and (AS) are not jointly sufficient for a reason to be adequate, but are only necessary conditions. This comports with our externalist inspiration, for our semantics basically treats adequacy as a property of a reason and a world: two worlds could be exactly alike relative to an agent's belief, without the reason being adequate in both, or inadequate in both (see Williamson 2000 on *good* vs. *bad* cases).

4 Knowledge and inadequate reasons

In view of the preceding, we surmise that knowledge may be viewed as a form of justified true belief, provided the justification is adequate. That is, we propose to defend that

$$K\varphi \text{ iff } \exists r(\mathsf{JTB}_r^e\varphi)$$

This definition raises two main objections. The first is that the definition is potentially too weak. It does not rule out cases in which an agent believes one and the same proposition based on at least two different reasons, one adequate, the other inadequate. We may require for agents to have knowledge only if every reason they have for a proposition is adequate, i.e. $\forall r(\mathsf{JTB}_r^i\varphi \to \varphi)$. We reject that option: such cases are better described as cases in which an agent knows a proposition, but whose knowledge is *confused* and of *lesser quality* than that of an agent having equally many reasons, all of them adequate (we refer to Egré et al. 2015 for more details). In that sense, our account of knowledge only commits us to a weak form of infallibilism about knowledge: an agent who knows p can still have misconceived beliefs pertaining to p.

A symmetric objection to our account is that it is potentially too strong: several authors (Fitelson 2010; Sorensen 2015; Warfield 2005) consider that there are cases in which we can get knowledge from false lemmas. I may know that I am not late for the meeting if I believe that it is currently 2:58pm, when in fact it is 2:56pm, assuming the meeting is at 7pm. On the present account, my reason to believe that it is currently less than 7pm is inadequate, simply because it also supports the false proposition that it is 2:58pm. This is a case in which I have JTB^i that it is less than 7pm, without having JTB^e that it is less than 7pm.

One option in the face of such examples is to bite the bullet and to resist the intuition that I know I am not late for the meeting. We are not sure that it is the best response. We think the problem concerns how much approximation is tolerated in forming beliefs based on one's evidence. My reading '2:58pm' is obviously wrong regarding the actual time, but still *close enough* to the actual time to be relevantly used. It would be different if the agent's watch indicated 6pm when it is 2:56pm, or even 9am. For the latter cases, our intuition is that I merely have a luckily true belief. If, when I see '2.58pm' (r) on my watch, I form the belief 'it is around 2:58pm' (p), and from that proposition I infer 'it is less than 7pm' (q), then my reason r now is veridical for both p and q.

A way out, therefore, might be to relativize the adequacy of a reason to the selection of an appropriate domain of propositions supported by that reason. This nevertheless puts pressure on us to clarify the relation of support between a reason and a proposition. In our statement of the axiom (A), we include no restriction on the support relation. We think it is better to be normative, and not to include any such restriction in the definition of knowledge in terms of JTB^e. On the other hand, we are ready to accept that in actual ascriptions of knowledge, the adequacy of reasons is referred to a set of relevant propositions that is contextually determined.

5 Perspectives

The two systems presented here do not offer ways of combining reasons, as each axiom or rule always involves a single reason. To that extent, both systems are mostly simplifications of more realistic systems of the sort discussed by Artemov (2008), Artemov and Fitting (2011), Baltag et al. (2014), and Dutant (2015). In Egré et al. (2015), we present variants of QRBB in which reasons can be combined. Some further issues are considered there, in particular regarding the implications of the current approach for higher-order knowledge (see Williamson 2000), and the question of the admission of basic beliefs, namely beliefs based on no further reason.

One may also wonder how our approach relates to the externalist view that knowledge is not definable in terms of belief plus other conditions, a view defended by Zagzebski (1994), and Williamson (2000). *Prima facie*, we may appear to be at odds with that conception, since we explicitly endorse a logical analysis of knowledge in terms of other concepts. As our analysis shows, however, we do not give a reductive analysis of the notion of adequacy, but only bring to light some substantive constraints on this notion.

References

Artemov, S. N. (2008). The logic of justification. *The Review of Symbolic Logic*, 1(4):477–513.

Artemov, S. N. and Fitting, M. (2011). Justification logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.

Baltag, A., Renne, B., and Smets, S. (2014). The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165(1):49–81.

Chellas, B. F. (1980). *Modal Logic: An Introduction*. Cambridge University Press.

Chisholm, R. M. (1977). *Theory of knowledge*. Prentice Hall, Englewood Cliffs, NJ. 2nd edition.

Dretske, F. (1971). Conclusive reasons. *Australasian Journal of Philosophy*, 1(49):1–22.

Dutant, J. (2010). *Knowledge, Methods, and the Impossibility of Error*. PhD, University of Geneva.

Dutant, J. (2015). Methods models for belief and knowledge. Manuscript, London.

Egré, P., Marty, P., and Renne, B. (2015). Knowledge, justification and adequate reasons. Under review.

Fitelson, B. (2010). Strengthening the case for knowledge from falsehood. *Analysis*, 70(4):666–669.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23:121–123.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, 73(20):771–791.

Goldman, A. I. (1979). What is justified belief? In Pappas, G. S., editor, *Justification and Knowledge*, pages 1–23. Dordrecht: Reidel.

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., and Hashimoto, T. (2015). Gettier across cultures. *Noûs*. Forthcoming (DOI: 10.1111/nous.12110).

Sorensen, R. (2015). Fugu for logicians. *Philosophy and Phenomenological Research*. Forthcoming.

Sosa, E. (1974). How do you know? *American Philosophical Quarterly*, pages 113–122.

Sosa, E. (1979). Epistemic presupposition. In Pappas, G., editor, *Justification and Knowledge*, pages 79–92. Springer.

Turri, J. (2012). Is knowledge justified true belief? *Synthese*, 184(3):247–259.

Warfield, T. (2005). Knowledge from falsehood. *Philosophical Perspectives*, pages 405–416.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Zagzebski, L. (1994). The inescapability of Gettier problems. *The Philosophical Quarterly*, 44(74):65–73.

Negation and events as truthmakers

Tim Fernando

Trinity College Dublin, Ireland
`Tim.Fernando@tcd.ie`

Abstract

Kit Fine's conception of truthmakers as exact verifiers and falsifiers provides an approach to negation that can be tested against the notion that events are truthmakers. Temporally locating events presents complications that call for an account of indices at bounded but refinable granularity. A proposal is outlined that links grain with the choice of a finite alphabet, from which strings are formed representing change under forces. Fusing strings is noted to lead to sets of strings, raising the prospects of fusion as a non-deterministic operation against the prospects of truthmakers as string sets.

1 Introduction

As a logical connective, negation is an operation on propositions that are commonly understood as abstract types, in contrast to concrete particulars such as events that according to Davidson 1967, make action sentences true (page 91). Without restricting truthmakers to concrete particulars, Fine 2015 outlines a "highly general and abstract approach" to truthmaking under which a statement φ is interpreted as a pair $(V(\varphi), F(\varphi))$ of sets $V(\varphi)$ of φ's verifiers and $F(\varphi)$ of φ's falsifiers related by negation $\neg$ according to (1), subject to the intuition (2).

(1) $V(\neg\varphi) = F(\varphi)$ and $F(\neg\varphi) = V(\varphi)$

(2) verification is "exact" — i.e., a verifier of φ is "wholly relevant" to φ

What (1) and (2) might mean for events as truthmakers is explored below through examples of the form

$$A(t) \quad := \quad \text{Amundsen flew to the North Pole in } t$$

for different times t (e.g., 1926, May 1926). The basic claim of the present work is that exactness in (2) must be relativized to a bounded granularity that is refinable, and that negation must be viewed more broadly than what (1) suggests, encompassing forces for change. (1) reduces verification and falsification of $\neg\neg\varphi$ to φ

$$V(\neg\neg\varphi) = F(\neg\varphi) = V(\varphi) \quad \text{and} \quad F(\neg\neg\varphi) = V(\neg\varphi) = F(\varphi)$$

supporting the law of double negation described in Horn 1989 as a "betrayal of natural language for logical elegance and simplicity" (page 2), in direct opposition to an "asymmetricalist position" that "negative statements are about the positive statements, while affirmatives are directly about the world" (page 3). This latter position becomes tenable once the interpretation of φ is allowed to be more than the pair $(V(\varphi), F(\varphi))$. Indeed, Fine makes such allowances in expanding a *state space* $(S, \leq)$ to a *modalized state space* $(S, S^{\diamond}, \leq)$, where $V(\varphi)$ and $F(\varphi)$ are defined in $(S, \leq)$ as subsets of a set S of *states*[1] partially ordered by $\leq$, and constraints are encoded as the set $S^{\diamond} \subseteq S$ of *possible* states satisfying them. We start in the next section with a constraint concerning what Fine refers to as "truth-value gluts" which we ban through inexact truthmaking under classical negation.

---

[1] The term "state" is used in Fine 2015 with the understanding that it "is a mere term of art and need not be a state in any intuitive sense of the term." That said, the thrust of the present paper is to flesh out how it might be interpreted in linguistic applications appealing to events.

2 No gluts: inexact truthmaking and falsification

Given a partial order $\leq$ on a set S of states, including verifiers of a statement φ, let the set $V(\varphi_\leq)$ of *inexact verifiers of* φ consist of states with $\leq$-parts that verify φ

$$V(\varphi_\leq) := \{s \in S \mid (\exists s' \in V(\varphi))\ s' \leq s\}.$$

A φ-*glut* is an inexact verifier of both φ and $\neg\varphi$ — i.e., a state in $V(\varphi_\leq) \cap V((\neg\varphi)_\leq)$. To express φ-gluts in terms of $\varphi_\leq \wedge (\neg\varphi)_\leq$, let us recall Fine's interpretation of conjunction $\wedge$, assuming $\leq$ has least upper bounds given by "fusion" $\sqcup$, used also for disjunction $\vee$, (3).

(3) $V(\varphi \wedge \psi) := \{s \sqcup s' \mid (s, s') \in V(\varphi) \times V(\psi)\}$ and $F(\varphi \wedge \psi) := F(\varphi) \cup F(\psi)$
    $V(\varphi \vee \psi) := V(\varphi) \cup V(\psi)$ and $F(\varphi \vee \psi) := \{s \sqcup s' \mid (s, s') \in F(\varphi) \times F(\psi)\}$

(4) $V(\varphi_\leq \wedge \psi_\leq) = V(\varphi_\leq) \cap V(\psi_\leq)$

From (3), we can derive (4) and equate φ-gluts with verifiers of $\varphi_\leq \wedge (\neg\varphi)_\leq$

$$V(\varphi_\leq) \cap V((\neg\varphi)_\leq) = V(\varphi_\leq \wedge (\neg\varphi)_\leq).$$

If we try to capture the the complement

$$S - (V(\varphi_\leq) \cap V((\neg\varphi)_\leq))$$

of φ-gluts by negating $\varphi_\leq \wedge (\neg\varphi)_\leq$, we find

$$V(\neg(\varphi_\leq \wedge (\neg\varphi)_\leq)) = F(\varphi_\leq \wedge (\neg\varphi)_\leq) = F(\varphi_\leq) \cup F((\neg\varphi)_\leq)$$

and are led to asking what the set $F(\varphi_\leq)$ of inexact falsifiers of φ is. Fine 2015 appears to be silent on this question. One possible answer, recorded as (A1) below, is that an inexact falsifier of φ is a state with a $\leq$-part that falsifies φ.

(A1) $F(\varphi_\leq) := \{s \in S \mid (\exists s' \in F(\varphi))\ s' \leq s\} = V((\neg\varphi)_\leq)$

(A1) allows us to drop the parentheses in $\neg(\varphi_\leq)$ and $(\neg\varphi)_\leq$

$$V(\neg(\varphi_\leq)) = F(\varphi_\leq) = V((\neg\varphi)_\leq) \quad \text{and} \quad F(\neg(\varphi_\leq)) = V(\varphi_\leq) = F((\neg\varphi)_\leq)$$

from which it follows that $\neg(\varphi_\leq \wedge (\neg\varphi)_\leq))$ and $\varphi_\leq \vee \neg\varphi_\leq$ have the same verifiers

$$\begin{aligned}
V(\neg(\varphi_\leq \wedge \neg\varphi_\leq)) &= F(\varphi_\leq) \cup F(\neg\varphi_\leq) = F(\varphi_\leq) \cup V(\varphi_\leq) \\
&= V(\neg\varphi_\leq) \cup V(\varphi_\leq) = V(\varphi_\leq) \cup V(\neg\varphi_\leq) = V(\varphi_\leq \vee \neg\varphi_\leq)
\end{aligned}$$

and falsifiers (the φ-gluts)

$$\begin{aligned}
F(\varphi_\leq \vee \neg\varphi_\leq) &= \{s \sqcup s' \mid s \in F(\varphi_\leq) \text{ and } s' \in F(\neg\varphi_\leq)\} \\
&= \{s \sqcup s' \mid s \in V(\neg\varphi_\leq) \text{ and } s' \in V(\varphi_\leq)\} \\
&= V(\neg\varphi_\leq \wedge \varphi_\leq) = F(\neg(\varphi_\leq \wedge \neg\varphi_\leq)).
\end{aligned}$$

The equivalence between $\neg(\varphi_\leq \wedge \neg\varphi_\leq)$ and $\varphi_\leq \vee \neg\varphi_\leq$ conflates "no gluts" with "no gaps" (states verifying or falsifying $\varphi_\leq$). While we might tolerate the silence of gaps (in the interest of partiality), the noise of φ-gluts is another matter.

For an alternative to (A1) that keeps gluts separate from gaps, it is instructive to consider the example of events. Let us suppose for the sake of the argument the sequence of events described in (5), (6).

(5)   Amundsen flew to the North Pole in May 1926.

(6)   Amundsen stayed home in July 1926.

Given (5), we might assert (7), interpreting "in" as "within" (following, for example, Pratt-Hartmann 2005 and Beaver & Condoravdi 2007).

(7)   Amundsen flew to the North Pole in 1926.

The step from (5) to (7) points to a satisfaction relation $\models$ between temporal intervals t and statements φ validating the implication (U) for the subinterval relation $\sqsubseteq$.

(U) $t \models \varphi$ and $t \sqsubseteq t' \implies t' \models \varphi$

A sufficient condition for (U) is that $\models$ be inexact verification, (8), for some relation $\leq$ closed under $\sqsubseteq$ according to (U′).

(8) $t \models \varphi \iff t \in V(\varphi_\leq)$

(U′) $e \leq t$ and $t \sqsubseteq t' \implies e \leq t'$

Adding (6) into the mix would then lead to Amundsen flying to the North Pole and staying home the same year — a somewhat dubious conclusion that only becomes worse when we replace *same year* by *same time*. Talk of *the same time* suggests some notion of temporal extent $\tau(e)$, through which the temporal fit tolerated by (U′) can be tightened as in (E).

(E) $e \leq t \iff \tau(e) = t$

With (E) in place of (U′), (U) no longer falls out of (8). And indeed, the step from (6) to (9) illustrates an inference not upward (U) but downward (D), derivable from a pointwise notion $V_\circ$ of verification according to the account (H) of homogeneity in Taylor 1977 and Dowty 1979.

(9)   Amundsen stayed home in the second week of July 1926.

(D) $t \models \varphi$ and $t' \sqsubseteq t \implies t' \models \varphi$

(H) $t \models \varphi \iff t \subseteq V_\circ(\varphi)$ (where $V_\circ(\varphi)$ is a set of time points)

Together, (U) and (D) have the catastrophic consequence that time collapses under the triviality

$$(\exists t)\ t \models \varphi \iff (\forall t)\ t \models \varphi.$$

This collapse can be avoided by treating (5)/(7) separately from (6)/(9) on the basis of a difference in aspect: (5) describes an event that culminates, while (6) describes a state that holds (e.g., Parsons 1990). This separation is threatened by negation; a small step away from (6) is (10).

(10)   Amundsen did not fly to the North Pole in July 1926.

(10) lends plausibility to the view that the negation of an event is a state. The view, however, is controversial (see, for instance, Condoravdi 2008, Landman 2006). We can sidestep the controversy by avoiding the connective $\neg$ on φ in (8), and instead negating the full statement "$t \models \varphi$" (over time and $\models$) classically.[2] Complementing (8), we introduce a counter-satisfaction

---

[2] This is essentially the move Champollion 2015 makes, applying negation not on an event predicate φ but on a set of event predicates. Champollion's interpretation $\lambda V \lambda f. \neg V(f)$ of *not* in (28), page 47, is approximated here (without the adjustment of his closure, (15), page 40) by $\lambda t \lambda \varphi.$ not$(t \models \varphi)$, where his $V : \langle vt, t \rangle$ is recast as the temporal interval t (understood as the set of event predicates that happen in t), and his event predicate $f : vt$ as φ (understood as the set of events verifying φ). This is in line with Davidson 1967, where (5) above appears as (7), and Reichenbach's use of an existential quantifier is adopted in page 91

> When we were tempted into thinking a sentence like (7) describes a single event we were misled: it does not describe any event at all. But if (7) is true, then there is an event that makes it true. This unrecognized element of generality in action sentences is, I think, of the utmost importance in understanding the relation between actions and desires.

predicate in (11), replacing (A1) by (A2) to (re)define inexact falsifiers of φ as states that fail to inexactly verify φ.

(11) $t \dashv \varphi \iff t \in F(\varphi_\leq)$

(A2) $F(\varphi_\leq) := S - V(\varphi_\leq) = \{s \in S \mid (\forall s' \leq s)\ s' \notin V(\varphi)\}$

(A2) breaks the equivalence between $\neg(\varphi_\leq)$ and $(\neg\varphi)_\leq$, for two distinct negations

  (i) the classical form, $S - V(\varphi_\leq)$, universally quantifying over $\leq$-parts, and

  (ii) the internal form, $V((\neg\varphi)_\leq)$ existentially quantifying over $\leq$-parts.

Under (A2), "no gluts" is expressed by $\neg(\varphi_\leq \wedge (\neg\varphi)_\leq)$

$$
\begin{aligned}
V(\neg(\varphi_\leq \wedge (\neg\varphi)_\leq)) &= F(\varphi_\leq) \cup F((\neg\varphi)_\leq) \\
&= (S - V(\varphi_\leq)) \cup (S - V((\neg\varphi)_\leq)) = S - (V(\varphi_\leq) \cap V((\neg\varphi)_\leq)) \\
F(\neg(\varphi_\leq \wedge (\neg\varphi)_\leq)) &= V(\varphi_\leq \wedge (\neg\varphi)_\leq) = V(\varphi_\leq) \cap V((\neg\varphi)_\leq)
\end{aligned}
$$

which is not to be confused with the expression $\varphi_\leq \vee (\neg\varphi)_\leq$ of "no gaps"

$$
\begin{aligned}
V(\varphi_\leq \vee (\neg\varphi)_\leq) &= V(\varphi_\leq) \cup V((\neg\varphi)_\leq) \\
F(\varphi_\leq \vee (\neg\varphi)_\leq) &= \{s \sqcup s' \mid s \in S - V(\varphi_\leq) \text{ and } s' \in S - V((\neg\varphi)_\leq)\} \\
&= (S - V(\varphi_\leq)) \cap (S - V((\neg\varphi)_\leq)) = S - (V(\varphi_\leq) \cup V((\neg\varphi)_\leq)).
\end{aligned}
$$

The question for events above is what is $\leq$, and, in particular, what are the states to the left and right of $\leq$?

3   Temporal grain: strings fusing into languages

The previous section adopts the common assumption that an event e has a temporal extent $\tau(e)$ that, for instance, determines when $e \leq t$ in line (E). Notorious difficulties in pinning down the precise moment of change (e.g., Kamp 1979) and concerns about "minimalitis" (Fine 2015, page 9) ought, however, to give us pause before fixing an event's temporal extent *once and for all*. As widely accepted a condition as (H) might be on statives φ, note that in practice, "July 1926" in (8) is chosen from a limited set of alternatives (constituting the answers to some question under discussion), sufficient for the purpose at hand (say, to rule out Amundsen from flying to the North Pole within that period), making (8) compatible with (12).

(12)   Amundsen was away from home for a couple of hours in July 1926.

Bounded granularity calls for tolerating temporal imprecision or, when necessary, adjustments to that granularity. Granularity is analyzed below in terms of temporal propositions which express statives as well as the application of forces that bring about changes constituting events. Henceforth, we refer to temporal propositions as *fluents* (for short), which we assume to belong to some fixed infinite set Φ. Against a linear order $(T, <)$ (such as the real line), a fluent a is interpreted as a subset $\tau(a)$ of T subject to an assumption (BV) of bounded variation saying $\tau(a)$ has a finite boundary under the order topology.

(BV)   there are finitely many intervals $I_1, \ldots I_k$ of T such that $\tau(a) = \bigcup_{i=1}^{k} I_i$

Given (BV), we form the function $\tau_a : T \to \mathbb{N}$ from T to the set $\mathbb{N} := \{0, 1, \ldots\}$ of non-negative integers such that for each $t \in T$, $\tau_a(t)$ is the smallest integer in $\mathbb{N}$ for which

$$\tau_a(t) \text{ is odd} \iff t \in \tau(a) \tag{\dagger}$$

and for all $t' \leq t$, $\tau_a(t') \leq \tau_a(t)$. Next, given any finite subset Σ of Φ, we approximate τ by a string s_Σ over the alphabet 2^Σ of subsets of Σ as follows. For any $t \in T$, let τ_t^Σ be the function $\{(a, \tau_a(t)) \mid a \in \Sigma\}$ from Σ to the set $\mathbb{N}$ of non-negative integers. Because Σ is finite, so is the image

$$T_\tau^\Sigma \ := \ \{\tau_t^\Sigma \mid t \in T\}$$

of T under the projection $t \mapsto \tau_t^\Sigma$, and there is a string $f_1 \cdots f_n$ of functions from Σ to $\mathbb{N}$ such that

$$T_\tau^\Sigma = \{f_i \mid 1 \leq i \leq n\} \ \text{ and } \ f_1 < f_2 < \cdots < f_n$$

where $\leq$ is defined componentwise on functions $f, f' : \Sigma \to \mathbb{N}$

$$f \leq f' \ \iff \ (\forall a \in \Sigma) \ f(a) \leq f'(a).$$

Since the function space $\Sigma \to \mathbb{N}$ is a subset of $2^{\Sigma \times \mathbb{N}}$, the string $f_1 \cdots f_n$ belongs to $(2^{\Sigma \times \mathbb{N}})^+$, and it natural to construe a pair $(a, i) \in \Sigma \times \mathbb{N}$ as a fluent. Under the biconditional (†) above, the fluent a is essentially the disjunction $\bigvee\{(a, i) \mid i \in \mathbb{N} \text{ and } i \text{ is odd}\}$. We can turn $f_1 \cdots f_n$ into a string s_Σ over the alphabet 2^Σ by deleting (a, i) when i is even and simplifying all remaining pairs (a, i) to a. It is easy to reconstruct the string $f_1 \cdots f_n$ from s_Σ.

Behind the reduction of T to s_Σ above is the intuition that time advances only through change, which we capture by working with strings $\alpha_1 \alpha_2 \cdots \alpha_j$ that are *stutter-free* in that $\alpha_i \neq \alpha_{i+1}$ for i from 1 to $j - 1$. An equivalent way of characterizing stutter-free strings is through the biconditional

$$s \text{ is stutter-free} \ \iff \ s = bc(s)$$

where the *block compression* $bc(s)$ *of* s compresses blocks α^i of $i > 1$ consecutive occurrences in s of the same symbol α to a single α, leaving s otherwise unchanged

$$bc(s) \ := \ \begin{cases} bc(\alpha s') & \text{if } s = \alpha\alpha s' \\ \alpha \ bc(\beta s') & \text{if } s = \alpha\beta s' \text{ with } \alpha \neq \beta \\ s & \text{otherwise.} \end{cases}$$

For example,

$$bc(\boxed{\ }\,\boxed{\ }\,\boxed{e}\,\boxed{e}\,\boxed{e}\,\boxed{e,y}\,\boxed{e,y}\,\boxed{e}) \ = \ \boxed{\ }\,\boxed{e}\,\boxed{e,y}\,\boxed{e}$$

where we draw boxes (as in Kamp and Reyle 1993) instead of curly braces $\{,\}$ for sets construed as symbols in a string (to be read much like a film/cartoon strip). Apart from applying bc, we can make a string stutter-free by introducing a fresh fluent, say tic, to turn, for instance, $\boxed{a}\,\boxed{a}\,\boxed{a}$ into $\boxed{a}\,\boxed{a,\text{tic}}\,\boxed{a}$. Similarly, to extend the string

$$s_\Sigma \ := \ \boxed{\text{Jan}}\,\boxed{\text{Feb}}\cdots\boxed{\text{Dec}}$$

of length 12 (listing the months in a year), we add days d1,..., d31 to $\Sigma := \{\text{Jan},\ldots,\text{Dec}\}$ for

$$s_{\Sigma \cup \{d1,\ldots,d31\}} \ := \ \boxed{\text{Jan,d1}}\,\boxed{\text{Jan,d2}}\cdots\boxed{\text{Jan,d31}}\,\boxed{\text{Feb,d1}}\cdots\boxed{\text{Dec,d31}}.$$

In general, a finite set $\Sigma \subseteq \Phi$ of fluents fixes a level of granularity, or grain (for short), that gets finer the larger Σ is. Given a string s over the alphabet 2^Φ, its componentwise intersection with Σ yields a string over the alphabet 2^Σ, denoted by $\rho_\Sigma(s)$

$$\rho_\Sigma(\alpha_1 \cdots \alpha_n) \ := \ (\alpha_1 \cap \Sigma) \cdots (\alpha_n \cap \Sigma).$$

In the calendar example above, $\rho_\Sigma(s_{\Sigma \cup \{d1,\ldots,d31\}})$ is $\boxed{\text{Jan}}^{31} \cdots \boxed{\text{Dec}}^{31}$ which bc maps back to s_Σ. Stutter-freeness keeps a string with a particular grain Σ from stretching beyond the distinctions that can be expressed in Σ.

Given a grain Σ, an obvious candidate for a state in Fine's state space is a stutter-free string over the alphabet 2^Σ. As for $\leq$, it is instructive to consider what the fusion $\boxed{\ \boxed{e}\ } \sqcup \boxed{\ \boxed{e'}\ }$ of the strings $\boxed{\ \boxed{e}\ }$ and $\boxed{\ \boxed{e'}\ }$ might be. The string $\boxed{\ \boxed{e,e'}\ }$ comes immediately to mind. But if a stutter-free string s is to represent all strings whose block compression is s, then there is also $\boxed{\ \boxed{e}\boxed{e'}\ }$ and the other strings in the set

$$\{bc(s) \mid s \in (2^{\{e,e'\}})^+,\ bc(\rho_{\{e\}}(s)) = \boxed{\ \boxed{e}\ }\ \text{and}\ bc(\rho_{\{e'\}}(s)) = \boxed{\ \boxed{e'}\ }\}$$

which can be shown to equal

$$\text{Allen}(e,e') \quad := \quad \mathcal{L}(e \bigcirc e') \cup \mathcal{L}(e \prec e') \cup \mathcal{L}(e' \prec e)$$

where $\mathcal{L}(e \bigcirc e')$ is the set of 9 strings $\boxed{s\boxed{e,e'}s'}$, for $s, s' \in \{\epsilon, \boxed{e}, \boxed{e'}\}$, in which e overlaps with e', while $\mathcal{L}(e \prec e')$ consists of the two strings $\boxed{\ \boxed{e}\boxed{e'}\ }$ and $\boxed{\ \boxed{e}\ \boxed{e'}\ }$ in which e precedes e' (putting 13 strings into $\text{Allen}(e,e')$, one for each interval relation in Allen 1983). The set $\text{Allen}(e,e')$ can be obtained from $\boxed{\ \boxed{e}\ }$ and $\boxed{\ \boxed{e'}\ }$ by applying a certain binary operation $\&_{bc}$ defined as follows. The *superposition of* two strings over 2^Φ of the same length is their componentwise union

$$\alpha_1 \cdots \alpha_n \ \& \ \beta_1 \cdots \beta_n \quad := \quad (\alpha_1 \cup \beta_1) \cdots (\alpha_n \cup \beta_n)$$

while the *superposition of* two languages L and L' over 2^Φ is the set of superpositions of strings of the same length from L and L'

$$L \ \& \ L' \quad := \quad \{s \& s' \mid (s, s') \in L \times L' \text{ and length}(s) = \text{length}(s')\}.$$

Next, we collect all strings bc-equivalent to a string in L in

$$L^{bc} \quad := \quad \{s \in (2^\Phi)^* \mid (\exists s' \in L)\ bc(s) = bc(s')\}$$

and take the image under bc of the superposition of L^{bc} and L'^{bc} for

$$L \ \&_{bc} \ L' \quad := \quad \{bc(s) \mid s \in L^{bc} \ \& \ L'^{bc}\}.$$

Conflating strings with their singletons, we have, as promised,

$$\text{Allen}(e,e') \quad = \quad \boxed{\ \boxed{e}\ } \ \&_{bc} \ \boxed{\ \boxed{e'}\ }.$$

Three relations on strings that may or may not be stutter-free are useful in defining a partial order on stutter-free strings:

(i) the *factor* relation, formulated as the set-valued function

$$\text{factor}(s) \quad := \quad \{s' \mid (\exists u, v)\ s = us'v\}$$

returning the factors of s (obtained by removing a prefix u and suffix v from s)

(ii) *subsumption* $\unrhd$, componentwise containment $\supseteq$ between strings over 2^Φ of the same length

$$\alpha_1 \cdots \alpha_n \unrhd \beta_1 \cdots \beta_m \iff n = m \text{ and } \alpha_i \supseteq \beta_i \text{ for } 1 \le i \le n$$
$$\iff \alpha_1 \cdots \alpha_n = \alpha_1 \cdots \alpha_n \ \& \ \beta_1 \cdots \beta_m$$

so that $s \unrhd \rho_\Sigma(s)$ for all $s \in (2^\Phi)^*$, and

(iii) equivalence $=_{bc}$ up to block compression

$$s =_{bc} s' \iff bc(s) = bc(s').$$

Now, we define $s' \le s$ to mean that some factor of s subsumes some string bc-equivalent to s'

$$s' \le s \iff (\exists s_1 \in \text{factor}(s))(\exists s_2 =_{bc} s') \ s_1 \unrhd s_2.$$

One can show $\le$ is a partial order on stutter-free strings, and that for stutter-free s,

$$s' \le s \iff s \in s \ \&_{bc} \ (\boxed{}+\epsilon)s'(\boxed{}+\epsilon) \ . \tag{$\ddagger$}$$

Insofar as $\&_{bc}$ is a well-defined binary operation on languages (rather than strings) over 2^Φ, it is natural to step from states as strings up to states as languages, generalizing ($\ddagger$) to a set L of stutter-free strings for

$$L' \le L \iff L \subseteq L \ \&_{bc} \ (\boxed{}+\epsilon)L'(\boxed{}+\epsilon).$$

Alternatively, keeping stutter-free strings s, s' as states, we might treat $\&_{bc}$ as a non-deterministic operation in which the fusion $s \ \&_{bc} \ s'$ is a set of stutter-free strings, adjusting clause (3) above by taking unions of string sets

$$V(\varphi \wedge \psi) := \bigcup \{s \ \&_{bc} \ s' \mid (s, s') \in V(\varphi) \times V(\psi)\}$$
$$F(\varphi \vee \psi) := \bigcup \{s \ \&_{bc} \ s' \mid (s, s') \in F(\varphi) \times F(\psi)\}.$$

Block compression aside, let us lift subsumption $\unrhd$ to languages L and L' over the alphabet 2^Φ, and agree L *subsumes* L' if L is a subset of its superposition with L'

$$L \unrhd L' \iff L \subseteq L \& L'$$

so that conflating a string s as usual with the singleton $\{s\}$,

$$s \unrhd L \iff (\exists s' \in L) \ s \unrhd s' \ .$$

Now, given a representation of an event e at grain Σ as a string $str_\Sigma(e) \in (2^\Sigma)^+$, we define e to be

- Σ-*telic* if $str_\Sigma(e) \unrhd \boxed{\neg\varphi}^+ \boxed{\varphi}$ for some fluent $\varphi \in \Sigma$ (marking the culmination of e), and

- Σ-*durative* if $bc(str_\Sigma(e))$ has length ≥ 3

(Fernando 2015). For representations $str(e)$ of events that fall into the Vendler classes described in Dowty 1979 and modified in Moens and Steedman 1988, we form not only transitions $\boxed{\neg\varphi \mid \varphi}$ in which a fluent φ representing a stative becomes true, but also transitions $\boxed{ap(f) \mid ef(f)}$ recording an effectual application of a force f, with the intention that

$$ap(f) \quad \text{says} \quad \text{``force } f \text{ is applied''}$$
$$ef(f) \quad \text{says} \quad \text{``a previous application of } f \text{ is effectual.''}$$

The transitions $\boxed{ap(f) \mid ef(f)}$ and $\boxed{\neg\varphi \mid \varphi}$ describe semelfactives and achievements, respectively, together forming the non-durative column in Table 1.

	−durative	+durative	
	semelfactive	activity	
−telic	$\boxed{ap(f)\mid ef(f)}$	$\boxed{ap(f)\mid ap(f),ef(f)\mid ef(f)}$	
	achievement	accomplishment	
+telic	$\boxed{\neg\varphi\mid\varphi}$	$\boxed{\neg\varphi,ap(f)\mid\neg\varphi,ap(f),ef(f)\mid\varphi,ef(f)}$	

Table 1. *A minimal picture of ±durativity and φ-telicity*

Iterating the transitions $\boxed{ap(f)\mid ef(f)}$ yields the language

$$\mathcal{L}(f) \ := \ \boxed{ap(f)\mid ap(f),ef(f)}^{*}\boxed{ef(f)}$$

which we superpose with $\boxed{\ \mid\ \mid\ }^{+}$ and block compress for Table 1's activity entry ($-$telic, $+$durative), and superpose further with $\boxed{\neg\varphi}^{+}\boxed{\varphi}$ for Table 1's accomplishment entry ($+$telic, $+$durative). The four strings in Table 1 can be obtained from $\mathcal{L}(f)$ and $\boxed{\ \mid\ }$ using block compression and the three operators

$$\begin{aligned}
dur(L) \ &:= \ L\ \& \ \boxed{\ \mid\ \mid\ }^{+} \\
non\text{-}dur(L) \ &:= \ L - dur(L) \\
cul(L,\varphi) \ &:= \ L\ \& \ \boxed{\neg\varphi}^{+}\boxed{\varphi}
\end{aligned}$$

that pick out the durative, non-durative and φ-telic strings in L, respectively. These operators bring to mind Dowty's hypothesis that "the different aspectual properties of the various kinds of verbs can be explained by postulating a single homogeneous class of predicates – stative predicates – plus three or four sentential operators and connectives" (Dowty 1979, page 71). The obvious question is: do we need the fluents $ap(f)$ and $ef(f)$, and all the fuss about block compression? Our answer to this question in the next section implicates an event $\boxed{ap(f)\mid ef(f)}$ in the negation $\neg\varphi$ of a stative fluent φ (and in extending stutter-free strings beyond length 1).

4 Change: negation by force

The notion of force mentioned in the fluents $ap(f)$ and $ef(f)$ links up with fluents φ representing statives through a law of inertia decreeing that a stative will continue to hold "unless something happens to change" it (Comrie 1976, page 49). To make this precise, more notation is helpful. Given languages L and L' over the alphabet 2^{Φ}, let us collect all strings over 2^{Φ} whose factors subsuming L also subsume L' in the language

$$L \Rightarrow L' \ := \ \{s \in (2^{\Phi})^{*} \mid (\forall s' \in \text{factor}(s))\ \text{if } s' \trianglerighteq L \text{ then } s' \trianglerighteq L'\}.$$

For example, $\boxed{\varphi\mid\ } \Rightarrow \boxed{\ \mid\varphi}$ is the set of strings over 2^{Φ} such that whenever φ appears at a position, it appears at all later positions. The idea now is to let $force(\varphi)$ be the fluent $ap(f)$ for some force f with $ef(f) := \varphi$ so that the constraint

$$\text{Inr}(\varphi) \ := \ \boxed{\varphi\mid\ } \ \Rightarrow \ \boxed{\ \mid\varphi} + \boxed{force(\neg\varphi)\mid\ }$$

says φ persists (forward) unless a force is applied opposing φ. Similarly, the constraint

$$\mathrm{Inr}_b(\varphi) \;:=\; \boxed{\;\boxed{\varphi}\;} \Rightarrow \boxed{\varphi\;} + \boxed{force(\varphi)\;}$$

says φ persists backward unless a force was applied making it true, while

$$\mathrm{Suo}(\varphi) \;:=\; \boxed{force(\varphi)\;} \Rightarrow \boxed{\varphi\;} + \boxed{force(\neg\varphi)\;}$$

says an application of a force for φ succeeds unless (as in $\mathrm{Inr}(\varphi)$) opposed. We do *not* assume that for every force f, there is a fluent φ with $ef(f) := \varphi$ that is subject to the three constraints above. The constraints $\mathrm{Inr}(\varphi)$ and $\mathrm{Inr}_b(\varphi)$ may fail to apply when the change described is incremental; for example, an increase in the degree $deg[\psi]$ associated with a claim ψ

$$\uparrow\! deg[\psi] \;:=\; \bigvee_{d \in D[\psi]} (d \le deg[\psi]) \wedge Previous(deg[\psi] < d)$$

over some set $D[\psi]$ of ψ-degrees (such as temperatures, for the claim ψ that "the soup is hot"). It is understood above that *Previous* is the obvious temporal operator which, in the case of $\uparrow\! deg[\psi]$, unwinds to the language

$$\boxed{\;\boxed{\uparrow\! deg[\psi]}\;} \;\Leftrightarrow\; \sum_{d \in D[\varphi]} \boxed{deg[\psi] < d \;\big|\; d \le deg[\psi]}$$

where $L \Leftrightarrow L'$ abbreviates the intersection of $L \Rightarrow L'$ and $L' \Rightarrow L$. To keep the alphabet of the language finite, the set $D[\varphi]$ must be assumed finite, and indefinitely refinable, as any finite set chosen for $D[\varphi]$ can be expanded to a larger vocabulary $\Sigma \subset \Phi$.

For fluents φ to which $\mathrm{Inr}(\varphi)$ and $\mathrm{Inr}_b(\varphi)$ apply, let us be clear about what is said about negating φ. Any string over an alphabet 2^Σ in which each $\varphi \in \Sigma$ is subject to $\mathrm{Inr}(\varphi)$ and $\mathrm{Inr}_b(\varphi)$ block compresses to a string of length at most 1

$$(\forall s \in (2^\Sigma)^* \cap \bigcap_{\varphi \in \Sigma} \mathrm{Inr}(\varphi) \cap \mathrm{Inr}_b(\varphi)) \;\; length(bc(s)) \le 1$$

(assuming none of the fluents $force(\varphi)$ belong to Σ). Of course, we can always neutralize $\mathrm{Inr}(\varphi)$, $\mathrm{Inr}_b(\varphi)$ and $\mathrm{Suo}(\varphi)$ by applying forces (with an expanded alphabet), but the challenge is to do so in a principled manner. It is worth quoting from page 52 of Dowty 1986 at length

> This principle of "inertia" in the interpretation of statives in discourse applies to many kinds of statives but of course not to all of them ... there must be a graded hierarchy of the likelihood that various statives will have this kind of implicature, depending on the nature of the state, the agent, and our knowledge of which states are long-lasting and which decay or reappear rapidly. Clearly, an enormous amount of real-world knowledge and expectation must be built into any system which mimics the understanding that humans bring to the temporal interpretations of statives in discourse, so no simple non-pragmatic theory of discourse interpretation is going to handle them very effectively.

There is no "graded hierarchy" of inertia in $\mathrm{Inr}(\varphi)$ and $\mathrm{Inr}_b(\varphi)$. The defeasibility lies in the forces understood to be at play and in the resolution of opposing forces, on which $\mathrm{Inr}(\varphi)$, $\mathrm{Inr}_b(\varphi)$ and $\mathrm{Suo}(\varphi)$ are silent. Inasmuch as evidence for these forces is to be found in discourse (conceived broadly), the negation $\neg\varphi$ of a fluent φ representing an inertial stative is asymmetricalist in the sense of Horn 1989. It is tempting to formulate these forces as attribute value structures or frames (Fillmore 1982, Cooper 2015, Fernando 2015a), but it remains to be seen how much of that theory can be encoded as constraints (like $\mathrm{Inr}(\varphi)$, $\mathrm{Inr}_b(\varphi)$ and $\mathrm{Suo}(\varphi)$) satisfied by some set $S^\diamond$ of possible states in the sense of Fine 2015.

References

James F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM* 26(11):832–843. 1983.

David Beaver and Cleo Condoravdi. On the logic of verbal modification. In M. Aloni, P. Dekker, & F. Roelofsen (eds.), *Proceedings of the 16th Amsterdam Colloquium*, pp. 3–9. University of Amsterdam. 2007.

Lucas Champollion. The interaction of compositional semantics and event semantics. *Linguistics and Philosophy* 38:31–66. 2015.

Bernard Comrie. *Aspect*. Cambridge University Press, 1976.

Cleo Condoravdi. Punctual until as a scalar NPI. In K. Hanson & S. Inkelas (eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky*, pp. 631–654. MIT. 2008.

Robin Cooper. Frames as records. In A. Foret, G. Morrill, R. Muskens & R. Osswald (eds.), *Preproceedings of the 20th Conference on Formal Grammar*, pp. 1–17. Barcelona. 2015.

Donald Davidson. The logical form of action sentences. In N. Rescher (ed.), *The Logic of Decision and Action*, pp. 81–95. University of Pittsburgh Press. 1967.

David R. Dowty. *Word Meaning and Montague Grammar*. Reidel. 1979.

David R. Dowty. The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics? *Linguistics and Philosophy* 9(1):37–61. 1986.

Tim Fernando. The semantics of tense and aspect: a finite-state perspective. In S. Lappin & C. Fox (eds.), *The Handbook of Contemporary Semantic Theory*, second edition, pp. 203–236. Wiley. 2015.

Tim Fernando. Types from frames as finite automata. In A. Foret, G. Morrill, R. Muskens & R. Osswald (eds.), *Preproceedings of the 20th Conference on Formal Grammar*, pp. 18–38. Barcelona. 2015a.

Charles J. Fillmore. Frame semantics. In *Linguistics in the Morning Calm*, pp. 111–137. Hanshin Publishing Co. (Seoul). 1982.

Kit Fine. Truthmaker semantics. Chapter for the Blackwell Philosophy of Language Handbook. 22 page draft available online. 2015.

Laurence R. Horn. *A Natural History of Negation*. University of Chicago Press. 1989.

Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer. 1993.

Hans Kamp. Events, instants and temporal reference. In Rainer Bäuerle, Urs Egli & Arnim von Stechow (eds.), *Semantics from Different Points of View*, pp. 376–417. Springer-Verlag. 1979.

Fred Landman. Stativity operators in 1066. Handout, Tel Aviv University. 2006.

Marc Moens and Mark Steedman. Temporal ontology and temporal reference. *Computational Linguistics* 14(2):15–28. 1988.

Terence Parsons. *Events in the Semantics of English: A Study in Subatomic Semantics*. MIT Press. 1990.

Ian Pratt-Hartmann. Temporal prepositions and their logic. *Artificial Intelligence* 166:1–36. 2005.

Barry Taylor. Tense and continuity. *Linguistics and Philosophy* 1:199–220. 1977.

Quantification and Existence in Natural and Formal Languages

Suki Finn

University of York
`suki.finn@york.ac.uk`

Abstract

Quantification and existence have an intimate yet complicated relationship, which has bearing on important issues for philosophers, linguists, and logicians alike. This paper focuses on the interface between philosophical logic and linguistics regarding the 'particular' quantifier as it is used in natural and formal languages. In classical logic, the 'particular' quantifier is symbolized as ∃ and has now more widely come to be known as the 'existential' quantifier. True to its name, ∃ has been interpreted as the logical notation for existence, drawing the connection between quantification and existence. I challenge this interpretation through a formal study of the semantics of quantification in natural and formal languages, to deny the connection. I put forward a linguistically motivated view of how the semantics of existence works and how it interacts with quantificational expressions, to show that quantification should have nothing to do with existence. I argue that the ontological loading of the quantifier is smuggled in through the restriction of domains of quantification, without which it is clear to see that ∃ is not existential in any way. Once we remove domain restrictions, domains of quantification can include non-existent things, and quantification and existence can be separated once and for all.

1 The Quinean Ontological Criterion

Quine, in his seminal paper 'On What There Is' (1948), puts forward a criterion for how to recognise the ontological commitments of a discourse, manifested via translation into classical first order predicate calculus. Quine believes that we speak in an ontologically committing way in natural language by the use of (what he sees as quantificational) idioms like 'there exists' or 'there are'. Quantification is thus the means by which we display ontological commitment. In stating '3 is a prime number' one is *actually* stating $Na \wedge Pa$ which entails $\exists x(Nx \wedge Px)$, which for Quine is read as '*there exists* something that is a number and is prime'. Quine does not provide any reason for ontologically loading the quantifier ∃, nor argues for his criterion of ontological commitment, claiming that it is "trivial and obvious."[1] I will explore two possible reasons why a Quinean may conclude that the quantifier carries ontological commitment: (1) because ∃ is a regimentation of the ordinary language 'there exists' idiom and this already carries ontological commitment; (2) because ∃ is ontologically loaded by virtue of its semantics. These reasons correspond to the two issues I clarify in this paper: (1) whether quantification in natural language is ontologically committing; and (2) whether quantification in formal language is ontologically committing. I argue that quantification in both English and first order logic are ontologically neutral in section 3 and 4 respectively. In the next section 2, I explore if

---

[1] Quine (1992) p26

there is anything nearing an argument in Quine for ontologically loading quantification, looking to other elements of his philosophical picture for clues or justification. In particular I will look to Quine's set theory, and his slogans about entities and identity.

2   Domain restrictions from SET, NE, and TB

Quine's commitment to set-theoretic model theory (described as 'SET' below) and the following two slogans[2] NE and TB contribute to loading quantification:

SET: Domains are sets
NE: "No entity without identity"
TB: "To be is to be the value of a bound variable"

Quine's slogan TB is intended as a descriptive tool to find out what exists – our ontology will be made up of those things bound by variables in the best scientific theory. 'To be' is for Quine to be an existent entity, and to be a 'value of a bound variable' is to be quantified over in the domain. So TB states that to be existent is to be in a domain of quantification. I reject TB as it entails loaded quantification. The way to evaluate TB is thus to evaluate what it means to be included in a domain, to see whether domains are restricted to existent things. I show how the domain may be restricted using SET and NE in turn, and I reject these in favor of unrestricted domains. With a neutral domain, we get neutral quantification.

2.1   Restriction from SET

For Quine, and in the standard set-theoretic version of model theory, domains are seen as sets. Domains therefore will for Quine be restricted in the same way that sets are restricted. Sets are restricted by identity, since sets are required to have determinate identity conditions. To have determinate identity conditions is for there to be a determinate answer as to whether one set *a* is identical to another set *b*. Set theory also tells us that sets are identified extensionally by their members, and as such their members must also have determinate identity conditions – for every member of the set, there is a determinate answer as to whether it is identical to another member of the set. Since the set-theoretic version of model theory states that domains are sets, domains thus take on these same conditions. Domains, and members of domains, therefore also have determinate identity conditions. This is the restriction from SET on what can go in a domain: *all members must have determinate identity conditions*.

2.2   Restriction from NE

Quine's slogan NE states that there is no entity without identity. So all entities must have determinate identity conditions. This may sound similar to the restriction imposed by SET as having identity, but this restriction posed by NE applies to only certain kinds of thing. An 'entity' for Quine means an existent entity, as there are no other entities for Quine. As such, his NE states that there can be no *existent entity* without determinate identity conditions. Whereas, SET states that there can be no member of the domain (existent or not) without determinate identity conditions. So the restriction from NE on what can go in a domain is: *all the existents must have determinate identity conditions*.

---

[2] Quine (1948) p33

We are trying to find motivation or justification for TB, where the whole domain is restricted to only existent things. So far, from SET and NE we only have the domain restricted to those things with identity. What the Quinean must do to get domain restrictions out of the identity condition requirement, is to hold a biconditional reading of NE, so that the identity restriction selects all *and only* existent things to be possible members of the domain. That way, all things with identity must be existent, and thus restricting the domain to those with identity also restricts to existents. The biconditional is between 'being an entity' and 'having identity', and is read as going in both directions – not only do all existent entities require identity, but all entities with identity require existence. So we read NE as saying both 'no entity without identity' and 'no identity without entity' (where entities exist). These are the two directions for the biconditional:

> **Left-Right**: X cannot exist without having determinate identity conditions as in order to exist it must be determinately distinct from other existents.
>
> **Right-Left**: X cannot have determinate identity conditions without existing as existence is required for completeness or determinacy (which non-existents are said to lack).

From the biconditional NE we bridge the gap between SET and TB – SET provides us with the restriction that domains can only contain things with determinate identity conditions, and the biconditional NE provides us with the restriction that the only things with determinate identity conditions are existents, which brings us to TB which states that to be in a domain is to be an existent entity. Therefore, we derive that all and only existent things can be quantified over in a domain, hence TB and why ∃ is read 'there exists'. For Quine, this is the natural reading of ∃, and being part of the domain is how we use the term 'exists' as this is just what 'exists' means. Quine's identity constraint on domains ensures this reading of ∃, but this constraint is unnecessary. I will go on to reject this constraint by rejecting the restriction that SET imposes (that all members of domains require determinate identity conditions) and by rejecting the restriction that NE imposes (that all things with identity are existent).

2.3 Rejecting TB via SET or NE

To burn the bridge that leads us to TB we can deny the biconditional reading of NE, in particular by denying the direction Right-Left by showing that non-existents can have identity and can go in a domain, and thus we quantify over non-existents, so ∃ is neutral. To do this we need to find non-existents which meet the determinate identity conditions imposed by SET. Or, we can simply reject SET by denying the set-theoretic version of model theory that requires domains to be sets with determinate identity conditions. To do this we need to show that we can quantify over things that lack determinate identity conditions. In the rest of this section I explore these options of rejecting either SET or NE.

Quine's NE is motivated by his issue with possible fat men in doorways.[3] The problem with the possible fat man in the doorway is that there is no determinate answer as to whether he is identical to the possible tall man in the doorway, or the possible smelly man in the doorway etc. Without there being a determinate answer as to whether one is identical with another is for the things to be lacking determinate identity conditions. For Quine, not having determinate identity goes against what it is to be an object or an existent entity. So the possible fat man doesn't qualify. For Quine this may be just a plea to stop talking about possibilia, but it has the effect of restricting domains. The question is whether NE is motivated by the possible fat man being an illegitimate thing to talk about or by such talk problematically introducing him as an object into the domain as existent. If being in the domain

---

[3] Quine (1961) p4

has no ontological significance and only signifies that we talk of that thing then it seems unproblematic to talk of possibilia – it seems only problematic if quantification is loaded to give you existent possible fat men. Yet Quine's identity constraint on domains and its entities is defended as he thinks it affords our resultant theory a degree of clarity and definiteness. But I hope to demonstrate that it is not necessary to impose such a constraint, and so quantification without Quine's add-ons is naturally ontologically neutral.

The biconditional NE ensures that all and only existents have determinate identity conditions, and this is a substantial and controversial claim which makes Quine's logic heavily theory-laden. We needn't accept such a heavy load with our logic though, and in rejecting NE we can reject Quine's ontologically loaded logic. Firstly, it is not clear that *all* existent things meet Quine's identity conditions (and as such the conditions are not necessary), and secondly, some *non*-existent things may meet those identity conditions too (and as such are not sufficient). By not being necessary we deny the direction Left-Right by showing that we can have an entity without identity, and by not being sufficient we deny the direction Right-Left by showing that we can have non-existents with identity. So even if the domain is restricted by SET to include only those things with determinate identity conditions, this set of things need not be a set of existent things, and thus we cannot look to the domain to provide us with an ontology. Determinate identity conditions do not pick out all and only existents, so even if the domain is restricted by SET to have determinate identity conditions this doesn't restrict the domain to all and only existent things. It thus seems that determinate identity is neither necessary nor sufficient for existence. Therefore the biconditional NE cannot be a constraint on domain specification, leaving logic naturally neutral.

As stated before, to have determinate identity conditions means that for all a and all b there must be a definite answer as to whether $a=b$. Benacerraf[4] takes issue with this claim with regard to numbers and sets, by showing how there is no definite answer as to which sets the numbers are. Benacerraf notes there are many potential reductions from numbers to sets but since there is no principled way to choose between them then numbers aren't reducible or identical to sets. If numbers exist then they require determinate identity (according to NE), but without there being a fact of the matter as to which, if any, sets they are identical to, then they do not meet this condition. Many philosophers of mathematics, particularly in the structuralist tradition, take the lesson of this to be that numbers exist but without determinate identity conditions, denying NE. Azzouni[5] denies NE using fictional characters to show that determinate identity is not sufficient for existents as non-existent fictional things may meet the condition by stipulation. Other examples showing that determinate identity is not necessary for existents include things like rainbows or heaps. There are also examples in modern science of existents without determinate identity, such as fermions and bosons in Bose-Einstein statistics.[6] Thus the biconditional NE is too strong: by rejecting it in some direction we break the argument that leads to TB.

But if we feel compelled to allow for the biconditional NE, then in order to prevent the restriction on our domains to only existents we would thus have to reject SET. This would allow for things *without* determinate identity conditions into the domain, and NE would merely state that those things in the domain *with* determinate identity conditions will also be those things in the domain that exist. To reject SET is to deny the set-theoretic version of model theory, and so is to deny that domains are sets. It is standard to take domains as sets however this leads to problems that may motivate its rejection anyway. For example, when domains are sets we cannot have unrestricted universal

---

[4] Benacerraf (1965) p62
[5] Azzouni (2004) p101
[6] This is an example borrowed from Cie and Stoneham (2009) p87-88

quantification. This is because unrestricted quantification requires an unrestricted domain, and if the domain is a set then this requires the set to be unrestricted. Such an unrestricted set is a set of everything, which will therefore contain itself, opening the way to Russell's Paradox. So, treating domains as sets can lead to paradox. If one wants to allow for unrestricted quantification or an unrestricted domain, as Quine seems to (as he answers the question of what exists with 'everything!'), then one needs to deny SET to avoid ending up in Russell's Paradox. This allows for us to quantify over things without determinate identity conditions, and prevents the move from SET to the biconditional NE that leads us to TB which loads ∃ in turn.

2.4   Rejecting TB via quantification

If Quine has an argument for TB it's a poor one, depending on a biconditional reading of NE, a paradoxical acceptance of SET, or an unmotivated statement that quantification being loaded is simply 'trivial and obvious'. We can deny SET or NE as done above to block getting to TB, or we can provide independent reasons for neutral quantification to show that not only is Quine's loaded reading unmotivated but also is not at all trivial or obvious. I will now deny TB by looking at what quantification is in natural and formal languages. As described earlier, there could be two reasons why one may hold that quantification is ontologically loaded: (1) because ∃ is a regimentation of the ordinary language 'there exists' and this is already ontologically loaded; (2) because ∃ is ontologically loaded by virtue of its semantics. These reasons correspond to the two issues I clarify in the next two sections: (1) whether quantification in natural language is ontologically committing; (2) whether quantification in formal language is ontologically committing. I argue that quantification in both English and first order logic are ontologically neutral, and that examples of uses of quantification in natural and formal languages provide evidence against TB and do not support Quine's triviality thesis, whereas neutral quantification is consistent with the evidence.

3   Natural language quantification is neutral

In this section I attack the assumption that quantification in natural language can be ontologically committing. I will explain why it is incorrect to say 'there exists' is synonymous with 'some' in English[7] to show why 'there exists' is not quantificational and how 'some' (along with other quantified idioms) is ontologically neutral. ∃ cannot represent the meaning and logical role of both 'some' and 'there exists' in English (and cognates in other natural languages) since 'exists' is not quantificational (but rather is a predicate). Quantified sentences have nothing to do with existence – they shouldn't require existence for their truth or meaning, and they shouldn't imply ontological commitment.

I now turn to examples. If 'some' is to mean 'at least one existent thing', then there will be no difference between 'some' and 'there exists'. Burgess and Rosen for instance argue it is not easy to understand what the difference can be.[8] Priest responds that they could simply reflect on the sentence 'I thought of something I would like to give you as a Christmas present but I couldn't get it for you as it doesn't exist'.[9] Here, the 'something' cannot mean 'some existent thing' as it would be

---

[7] Though I focus on English, since quantificational logic is meant to be a formalization of idioms in a range of natural languages, my discussion has a more global scope across other languages too.
[8] Burgess and Rosen (1997) p224
[9] Priest (2005) p152

contradictory. However, other quantified 'some' sentences do appear to be ontologically loaded, like 'some beers are in my fridge', which will be true only if there *exists* beer in my fridge. Here however, it is not the 'some' that is giving the appearance of ontological loading, rather the 'in my fridge' is. 'Some' needn't require existence, but to be physically 'in my fridge' does. Furthermore, 'some' *cannot* require existence since that would entail that we cannot talk truly of some non-existent things without contradiction. For example, '*some* mice have American accents' is arguably true due to Mickey Mouse, yet we do not feel that the truth of this commits us to his existence. This is contrasted with '*there do not exist* mice with American accents' to articulate lack of ontological commitment.

Priest's example is a variant of a famous example of Strawson's,[10] who points to a dictionary of legendary and mythical characters and says, with regard to the characters, 'some of these exist and some of them don't exist'. The seemingly loaded word here is 'exist', and 'some' must be considered neutral, to prevent the contradiction in the second disjunct – 'there exist some characters that don't exist'. To account for sentences such as this without contradiction, we must be able to use 'some' in an ontologically neutral way. This points towards the ordinary usage of quantification in natural language to be ontologically neutral. Furthermore, there may be no way of making sense of our fictional practice but to quantify over fictional entities, and as such we must ensure that quantification is neutral to avoid commitment to such fictional entities. Treating the quantifier as ontologically neutral, and distinguishing 'some' as a quantifier and 'exists' as a predicate, will gain expressive resources for sentences which contain both 'some' and 'not exist' (like the examples above) in order to prevent contradictions.

One may protest that 'some' just by definition means 'at least one existent thing' and these examples can thus be dealt with by being not strictly speaking true. They could argue that all such examples are a misuse of language that is parasitic on their use of 'some', and are properly interpreted as involving a cancelling prefix to create a more accurate sentence such as 'in Disney there exists at least one mouse that has an American accent' to make it true. Those who adopt such a reading will argue that all uses of 'some' are loaded until it is cancelled by such a prefix, otherwise the sentence will just be false if it involves non-existent things. However such a strategy will not work for Priest and Strawson's examples, which involve a true sentence and a neutral use of the word 'some', where no prefix will easily fit. These examples give cases when you quantify over a domain of objects some of which are existent and some are not, so you cannot prefix your quantification to explain what is going on. This is since only part of the sentence will pertain to non-existents and another part of the same sentence pertains to existents, and so an overarching cancelling prefix for the whole sentence will not do since only part of the sentence will require the commitment to be cancelled.

So far I have thus argued that, against Quine, ∃ cannot be a regimentation of the ordinary language 'there exists' in virtue of it carrying ontological commitment, since quantificational terms in natural language like 'some' are ontologically un-committing. In the next section I further argue against Quine that ∃ cannot be ontologically loaded in virtue of its semantics either, since the semantics of the quantifier in formal language are ontologically neutral. I show quantification in formal languages like first order predicate logic to be ontologically neutral, and therefore unregimented quantification in natural language is neutral too.

---

[10] Strawson (1967) p13. Here, 'a good proportion' and 'most' mean 'some'.

4 Formal language quantification is neutral

Reading ∃ as 'there exists' is incorrect, as 'there exists' is *not* a quantificational phrase. ∃ properly understood is 'some'. The difference between 'some' and 'there exists' is that 'some' is an ontologically neutral quantificational term, and 'there exists' is not a quantificational term. 'Some' is about the *number* of things (namely only some of them), and so is *quantitative*, whereas 'there exists' describes the *way* things are (namely as existing things), and so is *qualitative*. Therefore, the word 'some' is fit for numerical quantificational use, and 'there exists' is not (as it is a predicate). ∃ cannot be the logical regimentation of the non-quantificational 'there exists'.

The reason 'there exists' is not quantificational can be motivated by looking to Generalized Quantifier Theory (GQT)[11]. According to GQT a quantificational noun phrase is made up of a determiner and noun. Determiners are words like 'some', 'all', 'a', 'most', 'five'. (Determiners, I argue, can be taken as ontologically neutral since we can talk about five unicorns for example). Nouns include words like 'numbers', 'cats', 'objects'. So, it is true that the sentence 'there is a number that is prime between 2 and 4' is a quantified sentence, but it is false that the quantifier is 'there is'. Actually, the quantifier is 'a number', with 'a' being a determiner and 'number' being a noun. The 'there is' is part of the existential construction, and is not part of the quantification, and sometimes is not even existential, for example 'there are many clever detectives, some of which do not exist', where 'there are' and 'some' are both ontologically neutral. The quantification itself is neutral, located in the determiner and noun. Therefore ∃ in logic translates to the neutral quantifier 'some' in English, rather than the non-quantificational 'there exists'.

The argument for quantifiers being ontologically neutral can be strengthened by looking at the logical connection between the two quantifiers ∀ and ∃. Berto asks, "why existential? The dual of 'universal' is not 'existential', but 'particular'."[12] As such, the dual of 'all' should be 'some', and not 'there exists'. This can be demonstrated by considering the inter-translatability between ∀ and ∃ where one quantifier is defined in terms of the other: $\forall x(Cx) = \sim\exists x(\sim Cx)$ and $\exists x(Cx) = \sim\forall x(\sim Cx)$. Furthermore, when we look to the numerical quantities of such words, we can see that ∃ is 0%<n≤100% ('some') and so ∀ as n=100% ('all') is an *instance* of ∃. Therefore, $\forall x(\varphi)x \to \exists x(\varphi)x$ should be a valid inference, since whatever is true of all of the x is true of some of the x. For example, when I have eaten all the cakes it is true that I have eaten some of the cakes. What is true in the universal case ought to carry over to the particular case. However when the particular case is ontologically loaded in virtue of reading ∃ (incorrectly) as 'there exists', then when we infer the particular case from the universal we therefore can prove that something exists. We can thus miraculously derive ontology from logical inferences if we accept $\forall x(\varphi)x \to \exists x(\varphi)x$ as valid and take ∃ to be ontologically loaded.

The above inference $\forall x(\varphi)x \to \exists x(\varphi)x$ is therefore taken as *invalid* when you allow for domains to include non-existent things, or to be empty, and treat ∃ as loaded. Classical logicians have responded by not allowing for empty domains, and Quineans respond by not allowing for non-existent things in domains, in order to retain the validity of the inference and not prove the existence of the things they do not want in their ontology. This is because if we do allow for an empty domain or for domains to include non-existents, whilst we can hypothesize about what all the x would be like in the universal part of the inference, we cannot say anything about a particular x since this requires existence when we read ∃ as loaded. Yet my response is that we should take ∃ to be ontologically neutral and simply

---

[11] Hofweber (2007) p23 and see Gamut (1991) for details.
[12] Berto (2012) p21

to mean >0%, so that the inference is valid, even when the domain contains non-existents (or is empty). This ensures that we cannot derive ontology from logic. We can keep the consistency and inter-translatability between ∀ and ∃ by treating them both as ontologically neutral, which allows them to quantify over domains that contain whatever it is that we speak about. And these domains can be neutrally specified by a meta-language.

Formal languages like first order predicate logic are interpreted with model theory. The model theory for a language is a specification of a model, which consists of a domain and for every 1-place predicate an extension which is a subset of the domain, and for every n-place predicate a set of n-tuples of members of the domain. There are two rules for the quantifiers in our formal language of logic: (∃) when at least one element of the domain is in the extension of the predicate; (∀) when all elements of the domain are in the extension of the predicate. We specify the domain, and specify the extension of the predicates. Thus far there has been no mention of existence or ontology in the meta-language of model theory, and so the model is naturally metaphysically quiet. The metaphysical noise comes through not in the quantification but in the specification of the domain to be quantified over – if the domain is specified in a metaphysical or ontologically loaded way then quantifying over it will also be loaded. Quantification is only committal if the specification of the domain in the model theory is committal. And whether domain specification is committal depends upon whether the meta-language in which the model theory is couched is itself committal. Model theory doesn't require an ontology and ensures that formal languages have no ontological commitments, so that quantification is neutral. Quine's background rules for inclusion in a domain isn't neutral, and this is where ontology is smuggled in, through the back door of domain specification.

In practice, whatever the natural language of English can talk about can go in a domain. Any further restriction (like Quine's) is therefore not part of standard model theory. The point of looking at the model theoretic approach to semantics is to show that it is done in an ontologically neutral way, and that the metaphysics is an addition that is not necessary and may be incorrect. Quine included this addition due to his preconception of what things exist (not including the possible fat man in the doorway). He thus looked to what he thought existed in order to derive his loaded logic which was then used to tell us what exists. So it seems he constructed logic to fit around his premade metaphysical ideas. Quine's method as such is circular (he calls it 'holistic'), as he decides on his ontology and molds identity conditions to fit, then these conditions deliver ontological results. Azzouni makes a similar remark: "One can't read ontological commitments from semantic conditions unless one has already smuggled into those semantic conditions the ontology one would like to read off"[13] and this is precisely what Quine does. It's circular to get ontology from logic given how Quine choses his logic - to fit his ontology. We thus get a criterion for existing (to be in the domain) and a criterion for being in a domain (to exist).

5 Conclusion

Quantification becomes ontologically loaded when the domain that is being quantified over is restricted to include only existent things. In this way, the Quinean can then look to the values of bound variables in scientific theories for their ontology, with the quantifier being the signifier of ontological commitment. ∃ thus becomes ontologically loaded and read as 'there exists', due to this domain restriction. Without such a restriction, quantification ceases to have anything to do with

---

[13] Azzouni (2004) p55

existence, and the quantifier ∃ should be read as 'some' and known as the 'particular' rather than 'existential' quantifier. Quineans may think that quantification in formal languages is ontologically committing because of the model theoretic machinery, the set theory, or the Tarskian semantics. I have shown that these specifications only give us an ontology if domains are not allowed to contain non-existent things, and so a domain restriction is needed. This restriction is not something that has been argued for successfully by Quine, and ∃ certainly is not trivially or obviously loaded, as Quine initially states. The model theoretic, set theoretic, and Tarskian semantics can be adopted just fine without ontological commitment, since there is no good (non-question-begging) argument for why domain membership (even as sets) requires existence. I thus conclude that we can have classical objectual quantifiers without existence, and that ∃ is the 'particular' quantifier as there is nothing existential about it at all.[14]

References

Azzouni, J. (2004) *Deflating Existential Consequence: A case for Nominalism,* OUP

Benaceraff, P. (1965) 'What numbers could not be', *Philosophical Review,* 74, 47-73

Berto, F. (2012) *Existence as a Real Property,* Synthese Library, Springer

Burgess, J. P. & Rosen, G. (1997) *A Subject with No Object,* Clarendon Press

Cei, A. & Stoneham, T. (2009) '"Let the occult quality go": Interpreting Berkeley's Metaphysics of Science', *European Journal of Analytic Philosophy,* 5.1, 73-91

Gamut, L. (1991) *Logic, Language, and Meaning*, volume II, University of Chicago

Hofweber, T. (2007) 'Innocent Statements and Their Metaphysically Loaded Counterparts', *Philosophers' Imprint*, 7.1, 1-33

Priest, G. (2005) *Towards Non-Being. The Logic and Metaphysics of Intentionality,* Oxford: Clarendon Press

Quine, W. V. O. (1948) 'On What There Is', reprinted from the *Review of Metaphysics*, 2.5, in *Proceedings of the Aristotelian Society*, Supplementary Volumes 25, 'Freedom, Language, and Reality' (1951) 217-234

Quine, W. V. O. (1961) *From A Logical Point Of View,* Harvard University Press

Quine, W. V. O. (1992) *Pursuit of Truth,* revised edition, Harvard University Press

Strawson, P. (1967) 'Is Existence Never a Predicate?' *Critica,* 1, 5-15

Prover-Skeptic Games and Logical Pluralism[*]

Rohan French

University of Groningen, Groningen, The Netherlands `rohan.french@gmail.com`

Abstract

Logical Pluralists face an explanatory challenge: how is it that there are a multitude of different, equally correct, accounts of logical consequence. In this paper we will argue that this challenge can be met quite easily if one adopts the 'built-in opponent' conception of logic: a multi-agent, dialogical view of the nature of logical consequence. We introduce Prover-Skeptic games in order to model this view of logical consequence, and use our formal models to make clear how a certain kind of pluralism about explanation leads to an interesting variety of logical pluralism.

1 The Explanatory Problem for Logical Pluralism

Is there a single answer to the question of whether a given argument is (deductively) valid? According to *logical monists* the answer here is an emphatic 'yes', deductive validity being determined by the deliverances of their favoured logic. According to *logical pluralists* such as JC Beall & Greg Restall ([2]) and Stewart Shapiro ([13]) the answer to this question is 'no'; there are a multitude of different logics which determine deductive validity. Immediately the logical pluralist faces a problem, raised quite elegantly by Rosanna Keefe in the following quote:

> A characterisation of a pluralist position needs to explain what it is to endorse all of the various consequence relations the pluralist accepts and how they relate to an intuitive notion of logical consequence. [10, p.1376]

Call this the *explanatory challenge* for logical pluralism. The logical pluralist needs (*i*) to explain in what sense these logics are all correct accounts of deductive validity, i.e. what it means to 'endorse' these logics; and (*ii*) explain how these different logics relate to an intuitive notion of logical consequence. The main concern in [10] (esp. in section 3 therein) is to show that the answer to this challenge given by Beall & Restall in [2] is inadequate, and we shall not detain ourselves any further with it here. Instead we will take a different tack. What we will argue for here is the claim that a particular dialogical conception of logic and deduction, the built-in opponent conception of deduction (outlined in §2) allows quite naturally for a limited but interesting form of logical pluralism. In order to show this we will make use of a formalisation of this dialogical conception of deduction in terms of a novel kind of dialogue game (§3), using this to argue in §5 for a particular kind of logical pluralism which is able to easily answer the explanatory challenge.

2 The Built-In Opponent Conception of Deduction

Here we will adopt the particular multi-agent dialogical conception of logic defended by Catarina Dutilh Novaes in [4, 5]—the *built-in opponent conception of deduction*. According to this approach rules of inference reflect the rules for engaging in a certain kind of discursive practice.

---

[*]Thanks to Leon Geerdink, Catarina Dutilh Novaes, Erik C. Krabbe, and Benjamin Rin for their helpful comments on versions of this material.

The discursive practices involved here are a specialised variety of semi-adversarial dialogues, the participants of which have opposite goals: the *Prover* (or Proponent) seeks to establish that a certain conclusion follows from given premises, while the *Skeptic* (or Opponent) seeks to block the establishment of this conclusion. The adversarial nature of this dialogues is obvious, and flows from the players having opposite goals. At a higher-level, though, there is also a large degree of cooperation between the two players, as Catarina Dutilh Novaes explains:

> Proponent's job is not only to 'beat Opponent'; she also seeks to *persuade* Opponent of the truth of the conclusion, if he has granted the truth of the premises. In fact, the goal is not only to show *that* the conclusion follows from the premises, but also *why* it does; this corresponds to the idea that deductive arguments ought to have explanatory value. In this sense, Proponent and Opponent are cooperating in a common inquiry to establish what follows from the premises, and thus to further investigate the topic in question. [5]

It is not only Prover's role which has a cooperative component, the Skeptic's job also requires a level of higher-order cooperation, requiring them to not simply obstinately refuse that conclusion follows from the premises, but instead to say what would be required to convince them that the conclusion followed from the premises. We will say more about this in §3.

Contemporary logical practice does not, at first glance, resemble the picture we have been sketching above of logic involving a kind of multi-agent interaction. Typically we think of the act of determining whether an argument is deductively valid as being a rather solitary activity, and this thought is at least partially correct. The other important aspect of the present approach is the idea that over time Skeptic has been progressively 'silenced' and idealised, until they are no longer an active participant but instead are part of the deductive method itself—their role being internalised and, in essence, played 'offline' by Prover. To beat such an idealised opponent Prover needs to make sure that there are no counterexamples to the inferential steps which they provide as responses to the Skeptic's challenges. Thus the internalisation of Skeptic provides a bridge between multi-agent dialogical practices and mono-agent inferential practices.

3   Prover-Skeptic Games

In order to formally model the built-in opponent conception of deduction we will use *Prover-Skeptic Games*, a kind of dialogue game introduced in [16, p.89–94] (where they are used to game-theoretically characterise the implicational fragment of intuitionistic logic), although we will largely follow the presentation of such games given in [14] (where they are used to characterise the associative Lambek calculus). In order to get a feel for how these games work consider the following dialogue.

A Dialogue

*In the library two logic students, **Penelope** and **Scott** are arguing over the validity of the argument from $p \to q$ and $q \to r$ to $p \to r$. Penelope thinks that this argument is valid and is trying to convince Scott, who is skeptical.*

- **Penelope (1)**: I reckon that $p \to r$ follows from $p \to q$ and $q \to r$
- **Scott (1)**: Yeah? Well if that's so then suppose I grant you $p \to q$ and $q \to r$ along with p, how are you meant to get r?
- **Penelope (2)**: If you grant me q I can get r from $q \to r$ (which you just granted).
- **Scott (2)**: But why should I grant you q?

Axioms

$$A \succ A$$

Introduction/Elimination Rules

$$\frac{\Gamma, A \succ B}{\Gamma \succ A \to B} \ (\to I) \qquad \frac{\Gamma \succ A \quad \Delta \succ A \to B}{\Gamma, \Delta \succ B} \ (\to E)$$

Structural Rules

$$\frac{\Gamma, A, A \succ B}{\Gamma, A \succ B} \ (W) \qquad \frac{\Gamma \succ B}{\Gamma, A \succ B} \ (K)$$

Figure 1: A Natural Deduction System, in Sequent-to-Sequent style, for the implicational fragment of Intuitionistic Logic, where the structural rules *contraction* $((W))$ and *weakening* $((K))$ are explicit. (Note that in $(\to E)$ that $A \to B$ is the major premise, and A the minor premise.)

- **Penelope (3)**: Well if you were to grant me p then I could get q from $p \to q$ which you granted at the start.
- **Scott (3)**: But why should I grant you p?
- **Penelope (4)**: Because you granted it to me at the start!

Penelope leaves the library triumphantly.

The above dialogue has the structure of the kind of dialogical interactions which are at the heart of the built-in opponent conception of deduction, with Penelope (as her name mnemonically suggests) acting as Prover and Scott as Skeptic. This is also the structure of Prover-Skeptic games. In [16] these are defined in austere syntactic terms, but we will find it much more helpful to characterise these games directly in terms of a natural deduction proof system like that given in Figure 1.[1]

Before we go on to describe the games themselves, let us dispense with some notational preliminaries. Throughout we will be concerned with the language $\mathcal{L}$ of implicational logic, the formulas of which are constructed out of a countable supply of propositional variables $p_0, p_1, p_2, \ldots$ (the first three of which we abbreviate as p, q, r) using the binary connective '$\to$' of implication. Throughout we will use uppercase roman letters as schematic letters for formulas from $\mathcal{L}$, and uppercase Greek letters for multisets of formulas from $\mathcal{L}$. Recall that a *multiset* is a set in which elements can occur multiple times, making $[A, A, B]$ and $[A, B]$ distinct multisets (we use $[A_1, \ldots, A_n]$ to denote the multiset consisting of the formulas $A_1, \ldots A_n$).[2] A *sequent* is a pair $\langle \Gamma, A \rangle$ of a multiset of formulas Γ and a formula A, which we will write throughout as $\Gamma \succ A$.

---

[1]Natural deduction systems are rarely given in this 'structurally explicit' sequent-to-sequent style, the most notable exception being the treatment of intuitionistic logic in [3, p.88f] which has an explicit rule of weakening, or the systen *Nat* in [9, p.114], in which it is shown that (K) (there called $\mathbb{M}$) is derivable given the rules for conjunction. In both of these cases the need for an explicit rule of contraction is avoided by working with sets rather than multisets. Usually structural variation is captured in natural deduction systems in terms of different policies regarding how assumptions can be discharged, but we will find it particularly helpful to be fully explicit about the applications of structural rules throughout.

[2]For more information on multisets consult [15]

3.1   Defining the Game

Given a multiset of sequents $\mathfrak{S}$ and sequents β, γ let us say that $\beta \Rightarrow^*_{(I)} \gamma$ iff we can derive γ from β via zero or more applications of our introduction rules (in our case, $(\to I)$), and $\mathfrak{S} \Rightarrow^*_{(E)} \beta$ iff we can derive β from the sequents in $\mathfrak{S}$ via zero or more application of our elimination rules and structural rules (in our case, $(\to E)$, (W) and (K)). Finally say that a sequent $\Gamma \succ A$ is *atom-focused* just when A is a propositional atom.

Definition 3.1 (Dialogue Game). *A dialogue over (Γ, A) is a (possibly infinite) sequence $\mathfrak{S}_1, \beta_1, \mathfrak{S}_2, \beta_2, \ldots$ where each $\mathfrak{S}_i$ is a multiset of sequents, and each β_i is a sequent where:*

1. *$\mathfrak{S}_1 = [\Gamma \succ A]$*

2. *$\beta_i \Rightarrow^*_{(I)} \sigma$ for some non-axiomatic $\sigma \in \mathfrak{S}_i$ and some atom-focused sequent β_i*

3. *$\mathfrak{S}_{i+1} \Rightarrow^*_{(E)} \beta_i$*

In this setting it is helpful to think of Prover and Skeptic moves in the following terms. First, read a sequent $\Gamma \succ A$ in our dialogues as claiming that 'you cannot grant me Γ without also granting me A', then:

- A *Skeptic move* β where $\beta \Rightarrow^*_{(I)} \sigma$ can be interpreted as questioning σ by asking Prover to show β. That is to say, such a move involves questioning Prover's claim that if you grant me Γ you must also grant me A (where $\sigma = \Gamma \succ A$) by asking Prover to demonstrate that if they grant Γ' that they must also grant p_β, a sequent from which we can derive σ.

- A *Prover move* $\mathfrak{S}$ in responce to a Skeptic challenge β can be interpreted as providing grounds for the claim that if they are granted Γ' then they must also be granted p_β, as if they are granted $\Gamma_1, \ldots, \Gamma_n$ then they must also be granted $A_1, \ldots, A_n$ (where $\mathfrak{S} = [\Gamma_1 \succ A_1, \ldots, \Gamma_n \succ A_n]$). Note that given the structure of our elimination rules the multiset union of the Γ_is will be a sub-multiset of Γ', and so the collection of sequents in $\mathfrak{S}$ can be seen as recording how it is that if they are granted Γ' they must also be granted p_β.

Moves in Prover-Skeptic games correspond quite naturally to the kinds of actions taken by Prover/Proponent and Skeptic/Opponent in the kind of semi-adversarial dialogues which are at the heart of the built-in opponent conception of deduction. What is just as important, though, is that these dialogue games also make clear how it is that Skeptic can be internalised to the method itself. In particular, we can give a completely deterministic, and syntactic, characterisation of the structure of a Skeptic challenge.

Lemma 3.2. *Suppose that $\sigma = \Gamma \succ A$, and:*

- *$A = B_1 \to (\ldots (B_{n-1} \to (B_n \to p_i)) \ldots)$*

- *$\Gamma' = [B_1, \ldots, B_n]$.*

Then $\beta_i \Rightarrow^*_{(I)} \sigma$ iff $\beta_i = \Gamma', \Gamma \succ p_i$.

This mean that Skeptic moves do not require any creativity to perform, and are thus the kind of moves which are apt to be internalised and 'simulated offline' by Prover. This fits quite well with the central idea of the built-in opponent conception that the Skeptic is not an active participant in these dialogues, as their role can simply be played by Prover (with the aid of a randomising device to choose the sequent to be challenged).

How do these games connect up to logical consequence? The answer here is the standard one in dialogical logic: A is a logical consequence of Γ when Prover has a winning strategy in the appropriate dialogue game. That is to say, that no matter what moves Skeptic makes,

Prover can always win any dialogue over (Γ, A), where Prover wins a dialogue if it reaches a point at which Skeptic can make no further move. Similarly, we can say that Skeptic wins if Prover can make no further move.[3] This does not necessarily cover all the cases, though, as we could end up in a situation where both players always have moves available to them. What is important for present purposes, though, are the dialogues for which Prover has a winning strategy.

Definition 3.3 (Winning Strategy). *A winning strategy (for Prover) for the dialogue over* (Γ, A) *is a labelled tree where*

- *The root note of the tree is a P-node* $([\Gamma \succ A])$.
- *Each branch is a dialogue over* (Γ, A)
- *Every P-node* $(\mathfrak{S}_i)$ *has* $|\mathfrak{S}_i|$ *S-node descendants.*
- *Every S-node has a single P-node descendant.*

In [16] it is shown that Prover has a winning strategy in the dialogue over (Γ, A) iff $\Gamma \succ A$ is valid in the implicational fragment of intuitionistic logic. We will discuss similar adequacy results below in §5.

3.2 An Example

To make clear how this works, let us now formalize the dialogue we gave at the opening of this section. This is a dialogue over $([p \rightarrow q, q \rightarrow r], p \rightarrow r)$, and proceeds as follows.

P(1)	$[p \rightarrow q, q \rightarrow r \succ p \rightarrow r]$	*Prover Starts*
S(1)	$p \rightarrow q, q \rightarrow r, p \succ r$	*Skeptic challenges the sole Prover assertion.*
P(2)	$[q \rightarrow r \succ q \rightarrow r, \quad p \rightarrow q, p \succ q]$	*Prover replies, offering sequents from which the Skeptic's challenge can be derived using* $(\rightarrow E)$.
S(2)	$p \rightarrow q, p \succ q$	*Skeptic challenges the only non-axiomatic sequent.*
P(3)	$[p \rightarrow q \succ p \rightarrow q, p \succ p]$	*Prover replies, winning the dialogue as Skeptic has no further moves available to make.*

Moreover, given that at no stage could Skeptic have made any other move, the above dialogue corresponds quite directly to a winning strategy for Prover.

4 Proofs and Explanations

The particular discursive practices which are at the heart of the built-in opponent conception of deduction are clearly norm governed. Some of these norms are quite simple and obvious: for example, the players are normatively required to take turns, and Skeptic is obliged to only doubt or query claims which are not obviously correct. In fact, the two norms mentioned are partially constitutive of the practice involved (and are mirrored in Prover-Skeptic games by the turn taking requirement, and the ban on Skeptic challenging axiomatic sequents). There

---

[3]As presented here Skeptic is rather more 'tolerant' than is ideal—allowing Prover to offer any collection of sequents they want, so long as they derive the challenge sequent. In part this means that there are no finite winning conditions for Skeptic, as Prover will always have a response open to them. This is not that important for present purposes, but does represent one degree in which the current presentation does not quite match up with the built-in opponent conception of deduction.

is another normative aspect of this practice which is worth singling out, though. Namely that both sides must agree on what counts as an adequate response to a challenge.

Recall that Prover moves are meant to do more than merely show that the premises follow from the conclusion, they're meant to be *explanatory*—answers to the question 'why does this follow'. As is pointed out in [7] what counts as an answer to a why question varies with various features, most importantly here with the interests of the person asking the question. This means that what counts as an adequate response to a challenge will be relative to the interests of the audience for which the proof is intended to be explanatory—i.e. to the interests of a community of inquiry.

We can see an example of this in work on the foundations of geometry at the beginning of the 20th century. There mathematicians appear to be concerned not just with what axioms are used in proving a theorems, but also in the *number of times* such axioms are, or indeed must, be appealed to in proving it. For example, G. Hessenberg in 1905 proves the Desargues axiom follows from a collection of axioms for plane projective geometry along with a threefold use of the Pappus axiom. In [11] it is also noted that all known proofs both require, and explicitly mention, this threefold use. One way (and I will readily admit, not the only way) of understanding this situation is in terms of a change in the norms governing deductive dialogues. Rather than it simply being sufficient that you show that the challenged claim follows 'of necessity' from the claims provided, the Prover must also indicate the extent to which different claims are used in showing this—resulting in a change in the norms of explanation in this context. One way to register this 'resource consciousness' is by working in a system in which we can draw a distinction between when A follows from Γ and two copies of B, and when A follows from Γ and a single copy of B—i.e. work in a system in which we do not have the rule (W) of contraction.

This gives us an example, in mathematics of all places, where the norms which govern deductive dialogues appear to be (contra [2, p.88]) relative to communities of inquiry, or perhaps better, relative to the particular interests of those communities. So, in this case, a concern over the extent to which a given axiom must be used in a proof (perhaps sparked by foundational concerns about the truth of such axioms) leads to a particular norm governing the practice of proving—namely that you should note the number of times various axioms must be used in proofs.

Similarly, one might argue that requests for explanatory proofs point towards a kind of relevance criterion, perhaps one in the style of the relevant logic á la [1] where one drops the structural rule of weakening $((K))$. This might suggest a more general norm which one might appeal to in deductive dialogues—of only providing the evidence that is *required* for deriving the challenge, rather than merely providing *enough* information to derive the challenge.

The resulting kind of pluralism has a very pleasant shape. On the dialogical conception described above, logic is normative for a certain kinds of semi-adversarial dialogue games (deductive dialogues). Given that such dialogues can be governed by different norms which can effect what counts as an acceptable responce in such games, we then get a kind of logical pluralism (a 'local pluralism' in the terminology of [8])—which logic is correct depends on the norms which govern that kind of deductive dialogue. The kind of pluralism which emerges here is similar to that argued for by Hartry Field in [6]. Field argues that the normativity of logic for thought combined with pluralism about epistemic norms gives rise to a form of logical pluralism, while here we have argued that the normativity of logic for certain kinds of dialogical interactions coupled with pluralism about the norms governing those dialogical interactions gives rise to logical pluralism. In both cases pluralism about logic arises out of a pluralism concerning what it is over which logic holds normative sway, good reasoning in Field's case and certain kinds of dialogues in ours.

5 A Route to Pluralism

This puts us in a comfortable position to deal with the explanatory challenge for logical pluralism. We have an intuitive notion of logical consequence characterised by the built-in opponent conception of deduction formalised using Prover-Skeptic games. In the previous section we argued that, in this framework, what counts as a response to a challenge is interest relative, and that this interest relativity can sometimes result in the inadmissibility of certain structural rules in Prover moves. What remains to be done is to show that this in turn gives rise to a plurality of *logics*. To do this we will find it helpful to define S-dialogues, generalising our previous definition of a dialogue. If $S \subseteq \{(W), (K)\}$ let us write $\mathfrak{S} \Rightarrow^*_{(E)S} \beta$ iff we can derive β from the sequents in $\mathfrak{S}$ via zero or more applications of our elimination rules and structural rules from S.

Definition 5.1 (*S*-Dialogue). *Let $S \subseteq \{(W), (K)\}$. Then an S-dialogue over (Γ, A) is a (possibly infinite) sequence $\mathfrak{S}_1, \beta_1, \mathfrak{S}_2, \beta_2, \ldots$ where each $\mathfrak{S}_i$ is a set of sequents, and each β_i is a sequent where:*

1. *$\mathfrak{S}_1 = \{\Gamma \succ A\}$*
2. *$\beta_i \Rightarrow^*_{(I)} \sigma$ for some non-axiomatic $\sigma \in \mathfrak{S}_i$*
3. *$\mathfrak{S}_{i+1} \Rightarrow^*_{(E)S} \beta_i$*

If we also write $\vdash_S \Gamma \succ A$ to mean that there is a proof of $\Gamma \succ A$ which uses only structural rules in S then we have the following adequacy result.

Theorem 5.2 (Adequacy). *Prover has a winning strategy in an S-dialogue over (Γ, A) iff $\vdash_S \Gamma \succ A$.*

The 'only if' direction of the above Theorem is relatively straightforward, being proved by relatively straightforward induction on the number of P-nodes in the strategy. This direction is vastly simplified by the fact that we have defined our dialogue games directly in terms of a natural deduction proof system. The 'if' direction is significantly more revealing. For reasons on space we will simply sketch how it goes here. Firstly note that any provable sequent $\Gamma \succ A$ has a proof which is in normal form in the sense of [12]. One of the distinctive features of normal form proofs is that they can be split up into 'tracks' ([17, p.143]), sequences of sequents which can be split into three parts: an E-part in which each sequent follows from the previous one via an E-rule or structural rules; a minimal sequent; and finally an I-part in which each sequent follows from the previous one using I-rules or structural rules. Given any proof in normal form we can then transform it into one in which: (1) the minimal sequent in each track is of the form $\Gamma \succ p_i$ for some Γ and p_i (putting it into $\beta\eta^-$-long form), and (2) structural rules are not used in the I-part of the track (by permuting the structural rules over the introduction rules). Call such a proof one in *top-heavy long normal form*. As it turns out, it is the fact that we can transform every proof into one in top-heavy long normal form which allows us to 'delay' all use of structural rules until Prover moves. Treating structural differences via different discharge policies (i.e. restricting multiple, or vacuous discharge) complicates matters here, as this would most naturally cause the presence or absence of structural rules to effect Skeptic moves, destroying the ability to internalize Skeptic in the natural way we are able to in the present setting.

Given an S-proof in top-heavy long normal form of the sequent $\Gamma \succ A$ it is a simple matter to read off a winning strategy for Prover in the S-dialogue over (Γ, A). Rather than explain the procedure in detail here we will present an illustrative example.
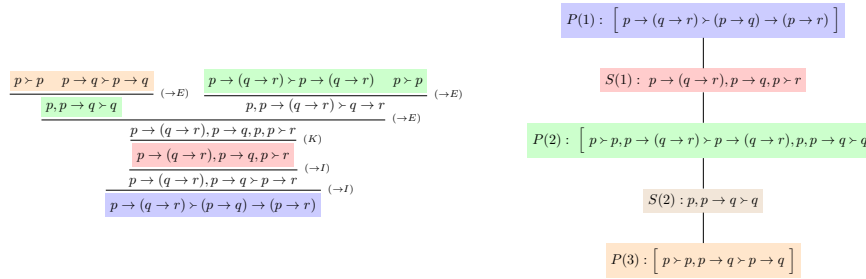
Figure 2: K-proof of $p \to (q \to r) \succ (p \to q) \to (p \to r)$ and a winning strategy in the K-dialogue over $(p \to (q \to r), (p \to q) \to (p \to r))$.

In Figure 2 we have a proof in top-heavy long normal form of the sequent $p \to (q \to r) \succ (p \to q) \to (p \to r)$. We can read a Prover winning strategy off this proof as follows. In essence Skeptic's challenge is the minimal formula on each track, and Prover responds by offering the sequent at the end of the track which the challenge is on along with the minor premises to any elimination rules used in the E-part of that track.

6   Conclusion

Logical pluralists face an explanatory challenge: they must explain what it is for a logic to be correct, and why there are many things which fit that bill. Here we have shown that a species of logical pluralism is forthcoming if we adopt the built-in opponent conception of deduction. On this approach a logic is correct if it is normative for the kind of dialogical interaction which one is currently engaged in, with different normative contraints on these interactions arising out of the interests of the participants. In many ways, though, the presentation is more suggestive than it is decisive. In particular:

- We have only dealt with a relatively limited logical vocabulary—pure implicational logic. In order to fully substantiate the claims made above we would need analogues of the above results for a more substantial selection of logical constants—at least $\{\wedge, \vee, \to, \neg\}$ in the propositional case.
- One might object that the above case we have made for the logical ramifications of the interest relativity of answers to questions about what follows from what could be stronger.

Future work is needed to bring this to fruition, but the above results do show that looking at logic from a dialogical perspective using Prover-Skeptic games affords a new and interesting view of traditional (and new) problems in philosophical logic.

References

[1] A. R. Anderson and N. D. Belnap. *Entailment: The Logic of Relevance and Necessity, Vol. I.* Princeton University Press, 1975.

[2] JC Beall and Greg Restall. *Logical Pluralism.* Oxford University Press,, 2006.

[3] M. Dummett. *Elements of Intuitionism.* Oxford University Press, Oxford, (second edition, 2000) edition, 1977.

[4] C. Dutilh Novaes. A Dialogical Account of Deductive Reasoning as a Case Study for how Culture Shapes Cognition. *Journal of Cognition and Culture*, 13:459–482, 2013.

[5] C, Dutilh Novaes. A dialogical, multi-agent account of the normativity of logic. *Dialectica*, Forthcoming.

[6] H. Field. Pluralism in Logic. *The Review of Symbolic Logic*, 2:342–359, 2009.

[7] B. Van Fraassen. The pragmatic theory of explanation. In Joseph C. Pitt, editor, *Theories of Explanation*, pages 136–155. Oxford University Press, 1988.

[8] O. Hjortland. Logical Pluralism, Meaning-Variance, and Verbal Disputes. *Australasian Journal of Philosophy*, 91:355–373, 2013.

[9] L. Humberstone. *The Connectives*. MIT Press, Cambridge, Massachusetts, 2011.

[10] R. Keefe. What Logical Pluralism Cannot Be. *Synthese*, 191:1375–1390, 2014.

[11] V. Pambuccian. Early Examples of Resource-Consciousness. *Studia Logica*, 77:81–86, 2004.

[12] D. Prawitz. *Natural Deduction: A Proof Theoretic Study*. Almqvist & Wiksell, Stockholm, reprint New York: Dover, 2006. edition, 1965.

[13] S. Shaprio. *The Varieties of Logic*. Oxford University Press, Oxford, 2014.

[14] S. Shoham and N. Francez. Game Semantics for the Lambek-Calculus: Capturing Directionality and the Absence of Structural Rules. *Studia Logica*, 90:161–188, 2008.

[15] D Singh, A.M. Ibrahim, T Yohanna, and J.N. Singh. A Systematization of Fundamentals of Multisets. *Lecturas Matemáticas*, 29:33–48, 2008.

[16] M.H. Sørensen and P. Urzyczyn. *Lectures on the Curry-Howard Isomorphism*. Elsevier Science, 2006.

[17] A Troelstra and H Schwichtenberg. *Basic Proof Theory*. Cambridge University Press, Cambridge, 2000.

Understanding Laughter

Jonathan Ginzburg[1]*, Ellen Breitholtz[2], Robin Cooper[2]
Julian Hough[3], and Ye Tian[1]

[1] Université Paris-Diderot
yonatan.ginzburg@univ-paris-diderot.fr,tiany.03@googlemail.com
[2] University of Gothenburg
cooper,ellen@ling.gu.se
[3] Bielefeld University
julian.hough@uni-bielefeld.de

1 Introduction

Laughter is pervasive in human conversation (more than 30k tokens in the dialogue part of the British National Corpus (BNC) (approx. 1 of every 14 turns (n = 430k).)). Most laughter is not triggered by humorous stimuli[1] and close to half of all laughter events are reactions to one's own speech (e.g., [20]). In this paper, we demonstrate that laughter needs to be integrated with lexically and phrasally produced import—arguing against the common assumption (see e.g., [10]) that laughter has no propositional content. Following this we develop a semantic-pragmatic analysis: we show how relatively general meanings, aligned with contextually–driven reasoning lead to a variety of disparate inferences.

2 Why assign content to laughter acts?

(1a–c) illustrates that laughter can occur as a stand alone utterance. In (1a,b) it conveys[2] that the question does not need addressing (a common use of laughter in political interviews [21], exemplified in (1b)), whereas in (1c) it conveys that the descriptive content of the assertion is false. (1d-e) illustrate cases of laughter with a quite different force: (1d) exemplifies laughter not intended to weaken the speaker's assertion; (1e) from a doctor's interaction with a patient, involves an initial laughter exchange concerning the patient's unpleasant condition, but ends with the doctor's sympathetic laughter, which in no way undermines his statement concerning her situation. (1f) exemplifies intra-utterance laughter, where the laughter has the effect of scare-quoting ([19]) the sub-utterance it precedes. (1g) illustrates that a laughter act can give rise to an intended meaning clarification request ([8]), hence having a content on a par with linguistic speech, whereas (1h) illustrates a non-linguistic trigger for laughter:

(1) a. Isaac: H, How long have you been here? Tracy: We were talking about you. Isaac: That's hilarious. Wh... What... Were you walking around behind us or what? Yale: (laughs) (No response provided ⇝ this question doesn't warrant answering) `https://www.youtube.com/watch?v=FBn28iNcpBA`

[1][In a report about a football match:] *The Chelsea man adopts that 'mirthless laugh at the dreadful awfulness and unfairness of it all' that his manager [Jose Mourinho] enjoys so much.* [The Guardian, November 29, 2015]

[2]For the moment, we use this in a way non-committal between semantic denotation and implicature to mean 'allows one to conclude'. We will return to this crucial issue below.

b. David Gregory; interviewing Chuck Schumer:
   DG: (1) Is Sarah Palin the future of the Republican party?
   CS: (2) **.hh hh=W(h)well(h)heh heh heh .hhuh** (From [21])[3]

c. Frank: She was actually erm phonin the doctor to see if she could come in and see him that morning about her gastroenteritis. Emily: Oh. Frank: She'll love me for telling you that. (laughs) ($\rightsquigarrow$ Frank doesn't think his wife will love him for telling Emily . . . (From [12])

d. 'Dave is someone who stands up for what he believes in, the sort of guy, now I think about it, you would want as your prime minister, if you had your choice.' He laughs, but he's not joking. (The Guardian, 30 Oct, 2015)

e. Anon 1: you know you're not going to be (laughing):[able to move it any] Anon 2: (laughs) Anon 1: and it's going to stay that way. Anon 2: (laughing):[No, Anna, you're alright.] Anon 1: (laughs) Anon 2: (laughing)::[You're alright.] (BNC, 89-92, G4D)

f. A : well I I'm interested 'in it in a (. laughs) ((comfortably)) re:laxed way, you know, I mean I . I do keep, I have kept up with it (London Lund Corpus)

g. Bonnie: (laughs) Cassie: Why are you laughing?

h. (At a doctor's surgery): Anon1: No wonder you're getting a buzzing in your ears there's a big lump of concrete in there. Anon 2: Is there? Anon 1: Let's have another look at this. Oh my. (laughs) (BNC)

3   Laughter as an event anaphor

Classifying laughter is not straightforward, neither in terms of form[4] nor function, and *a fortiori* with respect to the mapping between the two, which remains still very much an open question.[5] There is no consensual functional taxonomy—existing functional taxonomies of laughter focus on what laughter does as an emotion rather than via propositional content.[6] [25] investigate the sounds of human laughter in a very large corpus of naturally occurring conversational speech in Japanese and propose that the two largest classes by far are either polite or genuinely mirthful categories. They achieve better than 70% accuracy in automatic classification using machine learning.

We will offer a proposal that offers a basic account for, a fairly wide range of cases, as exemplified in (1), without claiming to be exhaustive. We propose to analyse laughter as a kind of eventive/situational anaphor, which can involve at least two distinct meanings:

1. A phylo- and onto-genetically prior meaning that conveys **enjoyment** of a situation and which can be related to the earliest laughter by children ([24]).

---

[3]https://www.youtube.com/watch?v=Zq_o8T8qOq0

[4][26] show huge variance within and across subjects. The number of phones used per laugher ranges from 2 to 59, with a mean (and median) of 32 (std : 14.4).

[5][3, 17] have found using both explicit ratings and online methods that voiced laughter elicited much more positive evaluation than did unvoiced laughter among subjects.

[6]For instance, [18] identifies eight social functions of laughter, including the expression of affiliation, aggression, social anxiety, fear, joy, comicality, amusement and as a self-directed comment.

2. A meaning we dub **incongruity** laughter—recognition that a situation or an event is in some ways incongruous. Indeed one might view this as a common feature of various theories of humour, ranging from Hobbs's superiority based account [13] to recent theories by Raskin, Attardo, Hurley, Dennett and Adam (e.g., [2, 14]). We assume that the variety of inferences associated with laughter arise from the combination of these meanings with contextually driven reasoning, which has a significant pragmatic component.

Proposing that laughter is ambiguous is defensible for a variety of reasons. First, the existence of distinct production and control mechanisms [22]. Second, there is copious evidence for misunderstanding the force of laughter; some cases can be analyzed in terms of referential uncertainty as to the laughable event, others seem better analyzed as misidentification of the intended use (which in some cases, might not be under the conscious control of the laugher).

(2) a. How could he explain a 7,000-mile drift at sea with stick figures? His impatience simmered. . . . . The native couple smiled and kindly shook their heads. 'Even though we did not understand each other, I began to talk and talk,' Alvarenga told me. 'The more I talked, the more we all roared with laughter. I am not sure why they were laughing. I was laughing at being saved.'(The Guardian, 9 Nov, 2015)

 b. Yoga Teacher: [Explains how to place a folded blanket above the groin and under the stomach and then bend forward folding the stomach over the blanket.] Student: [laughs] Teacher: Was that funny? (attested example)

At the same time, we propose, somewhat tentatively, that at least on some occasions incongruity laughter involves enjoyment as well. It seems clear, nonetheless, that not all laughter is associated with enjoyment, as with instances of embarrassment and/or nervous laughter, exemplified in (3), which one might wish to argue should be assimilable to incongruity; to the extent, one decides not to make such an assimilation, we nonetheless have an example of an enjoyment-less type of laughter:

(3) INT And then the demands at home are from your husband on one side and your children on the other. And basically the only time that I hear that you have for yourself is once a month on a Thursday night when you go to church. PAT Right. [Hhhh uh huh] INT [That doesn't] sound like very much. PAT It's not much. [hhhh heh hhh] INT [O k a y]. Tell me about depression. Has that been an issue for you . . . (Example from [11]

4   Formal Framework

In the rest of the paper, we show how the analysis of laughter as an eventive anaphor can be formalized within the dialogue framework KoS [8, 9]. KoS is formulated using the type theoretic formalism TTR [6] and has recently been extended to underpin defeasible reasoning [5].

4.1   Dialogue Gameboards and enthymemes

On the approach developed in KoS, there is actually no single context—instead of a single context, analysis is formulated at a level of information states, one per conversational participant. The dialogue gameboard (DGB) represents the publicized information in a given information state. Its structure is given in (4)—the *spkr,addr* fields allow one to track turn ownership, *Facts* represents conversationally shared assumptions, *VisInf* represents the dialogue participant's (view of) the visual situation and attended entities, *Pending* and *Moves* represent respectively

moves that are in the process of being grounded or have been grounded, QUD tracks the questions currently under discussion.

(4) $\quad$ DGBType $=_{def}$ $\begin{bmatrix} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{utt-time : Time} \\ \text{c-utt : addressing(spkr,addr,utt-time)} \\ \text{Facts : Set(Proposition)} \\ \text{VisInf :} \begin{bmatrix} \text{VisSit : RecType} \\ \text{InAttention : Ind} \\ \text{c1 : member(InAttention,VisSit)} \end{bmatrix} \\ \text{Pending : list(locutionary Proposition)} \\ \text{Moves : list(locutionary Proposition)} \\ \text{QUD : poset(Infostruc)} \end{bmatrix}$

In recent work,[5, 4] proposed that the dialogue gameboard also tracks **topoi** and **enthymemes** that conversational participants exploit during an interaction (e.g., in reasoning about rhetorical relations.). Enthymemes are defeasible arguments accounted for in rhetorical theory, but also found in conversational data [15]. Topoi represent general inferential patterns which may be used to underpin the enthymemes (e.g., *given two routes choose the shortest one*). Following [4] we formalise topoi and enthymemes as *dependent types*, more specifically functions from records to record types. The topos just mentioned regarding routes would be represented in TTR as the function in (5), which intuitively should be interpreted as a rule of thumb saying that if we have a situation of the type where we have two routes to choose from and one of these is shorter than the other, we may predict a situation where we choose the shortest route.

(5) $\quad \lambda r:$ $\begin{bmatrix} \text{x:}Ind \\ \text{y:}Ind \\ c_{\text{route}}\text{:route(x)} \\ c_{\text{route}_1}\text{:route(y)} \\ c_{\text{shorter\_than}}\text{:shorter\_than(x, y)} \end{bmatrix}$ $.$ $\begin{bmatrix} c_{\text{choose}}\text{:choose}(r.\text{x}) \end{bmatrix}$

The actual arguments conveyed in dialogue or other discourse which are drawing on topoi are referred to as *enthymemes*. They are applications of topoi in particular cases, e.g., *given that the route via Walnut street is shorter than the route via Alma, choose Walnut street*. Formally, an enthymeme belonging to a topos is a specification of the topos in the sense that the domain type of the enthymeme is a subtype of the domain type of the topos, and for anything, e, in its domain the result of applying the enthymeme to e is a subtype of the result of applying the topos to e. A specification of our topos concerning routes in (5) is shown in (6). It says that in a situation where we have a choice between Walnut Street and Alma and Walnut Street is shorter, we should choose Walnut Street. The domain type of this enthymeme is clearly a subtype of the domain type of the topos since the labels 'x' and 'y' are associated with particular entities, namely Walnut Street and Alma, not just with a type Ind as in the topos.

(6) $\quad \lambda r:$ $\begin{bmatrix} \text{x=Walnut Street:}Ind \\ \text{y=Alma:}Ind \\ c_{\text{route}}\text{:route(x)} \\ c_{\text{route}_1}\text{:route(y)} \\ c_{\text{shorter\_than}}\text{:shorter\_than(x, y)} \end{bmatrix}$ $.$ $\begin{bmatrix} c_{\text{choose}}\text{:choose}(r.\text{x}) \end{bmatrix}$

The analysis of enthymemes and topoi as dependent types in various degrees of specification exploits the possibilities of subtyping in TTR ([4, 6]) and enables us to formally represent how we employ topoi in different enthymemes through operations like restriction, generalisation and composition.

Topoi and enthymemes get added into the dialogue gameboard in a number of different ways. [4] discusses some of these cases. One case is where it is not clear which rhetorical relation is being introduced between two propositions in the dialogue. If the subject matter of the latest move is associated with a topos in the agent's resources, the topos will be added to the gameboard and an enthymeme under discussion (EUD) will also be integrated through accommodation. The other typical case is where the enthymematic structure is clear. In this case the latest move – possibly combined with beliefs already integrated in the discourse model – causes an enthymeme under discussion to be added to the gameboard. This enthymeme may then be matched to the resources of the agent. If there is a topos that validates the EUD, this topos is added to the gameboard. If there is no relevant topos the agent might still accommodate a topos based on the content of the EUD. For example, if agent A says "let's take Walnut Street, it's shorter" and agent B comes from a cultural and social context where efficiency and time is not important, he may not have access to a topos saying that "shorter routes are preferable". However, he may tentatively accommodate a topos which is a generalisation of the EUD. Mismatches between topoi and EUD on the gameboard of a dialogue participant may give rise to clarification requests (e.g., (1g)).

The relevance of enthymemes in dialogue and the topoi that underpin them becomes particularly apparent in cases when our individual takes on an ongoing interaction do not match. Consider an agent involved in dialogue who suggests to another dialogue participant that they choose to go somewhere via a particular street because it is longer (rather than shorter). Dialogue participant A may have a topos in mind about long walks being beneficial to one's health or the like, but this topos might not be available to dialogue participant B, who will then not understand the intention of A's utterance and possibly make a clarification request. It is also possible that some topos is indeed accommodated on the DGB of a dialogue participant in the process of interpreting an utterance, but that this topos does not agree with the topos and enthymeme intended by the speaker.

In (7), we have an example of such a situation. In this interview with Swedish rap artist Petter, the journalist's utterance suggests that she has formed a hypothesis about the kind of argument Petter is making based on his utterance and the topos in the journalist's resources which she associates with the type of situation described in that utterance.

(7) a. Petter: Metal was actually the reason I started doing Hip Hop ...[pause]
         Petter:...Because I hated metal.
         Journalist: Oh, I thought you were going to say something completely different!

(7) provides evidence that we start reasoning before an argument is fully spelled out and the way we process rhetorical structure is analogous to the way we process sentential and non-sentential utterances [7].

4.2   Interfacing with the grammar

We use HPSG$_{TTR}$ [8], a variant of the grammatical formalism Head–driven Phrase Structure Grammar ([23]). In HPSG$_{TTR}$ speech events are identified with records and grammatical types ('signs') are identified with record types.

We exemplify this with a lexical entry for a greeting word such as 'hi', as in (8), whose context—specified via the field 'dgb-params'—is supposed to be the initial state of a conversation:

(8)

$$
\begin{bmatrix}
\text{phon : hi} \\
\text{cat.head = interj : syncat} \\
\text{dgb-params :} \begin{bmatrix}
\text{spkr : IND} \\
\text{addr : IND} \\
\text{utt-time : TIME} \\
\text{Moves} = \langle \rangle \text{: list(LocProp)} \\
\text{qud} = \{\} \text{: set(Question)}
\end{bmatrix} \\
\text{cont = Greet(spkr,ind,utt-time) : IllocProp}
\end{bmatrix}
$$

More generally, the context, represented within the field *dgb-params*, plays a crucial role via QUD, VisSit or Pending, providing the main predicate and/or the conversational move type. Thus, as we have seen above the antecedents of laughter come from varied sources. In this, they resemble nominal pronouns in dialogue. (9a) is an example of anaphora from an ungrounded utterance, whereas (9b) is an example of anaphora from a disfluent utterance:

(9) a. A: Did John phone? B: Is he someone with a booming bass voice?

     b. Peter was, well he was fired.

This motivates the notion of an *active* move, as the source of an antecedent for a pronoun. What is an active move? Clearly this is an intricate notion, but drawing on a dialogue oriented conceptualization of the right frontier constraint (e.g., [1])— the "right frontier" is constituted by elements of QUD, answers to such (TOPICAL facts), and utterances under grounding/correction:

(10)   For a given DGB dgb0, an ActiveMove is an element of dgb0.Moves or dgb0.Pending such that either (a) *qud-update-contribution($m^{content}$)* is in dgb0.QUD or (b) $m^{content}$ is TOPICAL or (c) m is in dgb0.Pending.

4.3   Monitoring and Appraisal

Metacommunicative interaction is handled in KoS by assuming that in the aftermath of an utterance u it is initially represented in the DGB by means of a *locutionary proposition* individuated by u and a grammatical type T_u associated with u. If T_u fully classifies u, u gets grounded, otherwise clarification interaction ensues regulated by a question inferable from u and T_u. If this interaction is successful, this leads to a new, more detailed (or corrected) representation of either u or T_u. [9] develop their account in KoS of disfluencies, or phenomena of Own Communication Management (OCMs), by extending the account just mentioned of the coherence and realization of clarification requests: as the utterance unfolds incrementally there potentially arise questions about what has happened so far (e.g. *what did the speaker mean with sub-utterance u1?*) or what is still to come (e.g. *what word does the speaker mean to utter after sub-utterance u2?*). These can be accommodated into the context if either uncertainty about the correctness of a sub-utterance arises or the speaker has planning or realizational problems.

Thus, the monitoring and update/clarification cycle is modified to happen at the end of each word utterance event, and in case of the need for repair, a repair question gets accommodated into QUD.

This ubiquitous self-monitoring ties in with commonly accepted assumptions in cognitive psychology work on emotion (see [16]). Although there exist a variety of approaches to modelling emotions, there is a basic consensus that emotions are caused by *appraising* events in relation to concerns. In terms of a time course of appraisal, there exists an initial automatic appraisal that does not require conscious processing, and a secondary appraisal that often includes conscious reflection on the meaning of the emotion and that can lead to new intentions. One could hypothesize that *enjoyment* laughter is associated with a first appraisal, while *incongruity* laughter is associated with a secondary appraisal.

5   Enjoyment laughter

In this section we also offer a sketchy explication of enjoyment laughter. (11) associates an enjoyment laugh with the laugher's judgement of an eventuality l as enjoyable; more specifically l with respect to being classified by the type L (an *austinian proposition*) as enjoyable; with respect to form we underspecify this, appealing to a type `laughterphontype` compatible with the apparent large range of possible realizations. We make no special assumptions about the construal of enjoyment beyond those assumed by various cognitive theories of emotion on this score [16]. On most such theories this involves appraisal of the active situation in a positive way (the nature of the appraisal varies with the theory.).

(11)
$$
\begin{bmatrix}
\text{phon : } \texttt{laughterphontype} \\
\text{dgb-params : } \begin{bmatrix}
\text{spkr : Ind} \\
\text{addr : Ind} \\
\text{t : TIME} \\
\text{c1 : addressing(spkr,addr,t)} \\
\text{p} = \begin{bmatrix} \text{sit} = l \\ \text{sit-type} = L \end{bmatrix} : \text{Prop} \\
\text{c2 : ActiveSit(l)}
\end{bmatrix} \\
\text{content} = \text{Enjoy(spkr,p) : RecType}
\end{bmatrix}
$$

The notion of active situation in dialogue pertains to the accessible situational antecedents of a laughter act, which involve a generalisation of Asher's characterisation of eventive antecedents via the right frontier constraint for text (e.g., [1]). As we have seen above, they can be exophoric (1h), whereas when their antecedent is an utterance, they can be the ongoing utterance (1f), the most recent (1a–c), or a situation described by a sequence of moves, including the most recent move (*the punchline*):

(12)   Given l : Rec and d : DGBType, Active(l,d) if (i) =(l,d.VisSit)
        or (ii) =(l,d.MaxPending)
        or (iii) =(l,d.MaxQUD.sit)

Given this meaning and the topos *If I'm enjoying that I/you said that p, then I agree that p*, we can obtain as a consequence that enjoyment laughter can be used as a positive feedbak signal, as in (1d) and the doctor's final laugh in (1e).

6 Incongruous laughter

We can describe the meaning of incongruous laughter, to a first approximation, as in (13): the laugh marks a proposition whose situational component l is *active* as *incongruous*, relative to the currently maximal enthymeme under discussion:

$$(13) \quad \begin{bmatrix} \text{phon : laughterphontype} \\[4pt] \text{dgb-params :} \begin{bmatrix} \text{spkr : Ind} \\ \text{addr : Ind} \\ \text{t : TIME} \\ \text{c1 : addressing(spkr,addr,t)} \\ \text{MaxEud = e : (Rec)RecType} \\ \text{p} = \begin{bmatrix} \text{sit = l} \\ \text{sit-type = L} \end{bmatrix} : \text{prop} \\ \text{c2 : ActiveSit(l)} \end{bmatrix} \\[4pt] \text{content = Incongr(p,e,}\tau\text{) : RecType} \end{bmatrix}$$

We explicate *incongruity* in terms of a clash between the enthymeme triggered by the laughable and a topos which the enthymeme is supposed to instantiate. That is, the laughable l satisfies the domain type of the enthymeme, but there is a clash between the range of the enthymeme and that of the topos that the enthymeme is supposed to instantiate. Specifically, in (14), p is a proposition comprised of l, the laughable event, and L a type that classifies l, E is the triggered enthymeme, and τ is the clashing topos—E's domain is a subtype of τ, but its range (P(unch)L(ine)) is incompatible with τ's range:

$$(14) \quad \text{Incongruous } (p, E, \tau) \text{ iff } p = \begin{bmatrix} \text{sit = l} \\ \text{sit-type = L} \end{bmatrix} : TrueProp, \tau = \lambda r\!:\!T_1 \,.\, T_2 : (Rec{\rightarrow}RecType),$$
$$E = \lambda r\!:\!L \,.\, PL : (Rec{\rightarrow}RecType) \; L \sqsubseteq T_1 \text{ and } PL \perp T_2$$

Assertion cancellation Consider again the example in (1b)—Frank tells Emily that his wife has gastroenteritis, then he says that his wife will love him for telling Emily this. One simple view of the function of laughter here is as (mock) self-repair. Frank relies on the enthymeme 'If I'm saying she will love me because I mentioned her gastroenteritis, then I don't mean it.' This clashes with the sincerity topos 'If A says p, then A means p'.

Laughter as scare quoting Here we have an interaction between laughter and disfluency. Here the laughable is A's upcoming utterance u classified by a sign T_u. u is maximally Pending, assuming an incremental view of processing, as motivated by the treatment of disfluencies in [9]. The laughter can mark it as incongruous. In what way incongruous? We assume this often relies on a clash with the topos 'If Utter(A,u1), u1 represents A's choice to refer to u1's referent.'

Question deflection How does laughing enable questions to be deflected? A poses q; for B not to address q, B has to accommodate the issue ?Wish(B,q) [*whether B wishes to discuss q*] into QUD and provide an utterance about this issue (see[8], Chapter 8, section 8.3.1). A possible active situation accessible to B is the maximally pending utterance—A's query. Hence,

B's laughter can convey that this utterance is incongruous. In what way? We assume this relies on the violation of the topos 'If Ask(A,B,q), then q is a good/serious question'.

7   Conclusions and Future Work

Laughter is a frequent phenomenon in conversations with a wide range of meanings—it can both reinforce and cancel an assertion, deflect a question, or be used as scare quoting. We propose that laughter is an event anaphor that can convey at least two distinct meanings: the **enjoyment** of an event and the recognition of an **incongruous** event. We formalize these two meanings in the KoS framework, which captures the high degree of context dependency of laughter. The treatment explains how laughter can be the source for clarification requests due to ambiguity and difficulty in resolving the laughable. We are currently working on a more detailed classification of laughter based on the enjoyment/incongruity distinction, and linking form with function. We plan to address cross-linguistic differences, to experimentally test the enjoyment/incongruity distinction and its relation to the appraisal mechanism.

References

[1] Nicholas Asher. *Reference to Abstract Objects in English: a Philosophical Semantics for Natural Language Metaphysics.* Studies in Linguistics and Philosophy. Kluwer, Dordrecht, 1993.

[2] S. Attardo. *Linguistic theories of humor*, volume 1. Walter de Gruyter, 1994.

[3] Jo-Anne Bachorowski, Moria J Smoski, and Michael J Owren. The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, 110(3):1581–1597, 2001.

[4] Ellen Breitholtz. *Enthymemes in Dialogue: A micro-rhetorical approach.* PhD thesis, University of Gothenburg, 2014.

[5] Ellen Breitholtz. Reasoning with topoi–towards a rhetorical approach to non-monotonicity. In *Proceedings of the 50th anniversary convention of the AISB, 1st–4th April 2014, Goldsmiths, University of London*, 2014.

[6] Robin Cooper. Type theory and semantics in flux. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics. Elsevier, Amsterdam, 2012.

[7] Arash Eshghi, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matthew Purver. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 261–271, London, UK, 2015. ACL.

[8] Jonathan Ginzburg. *The Interactive Stance: Meaning for Conversation.* Oxford University Press, Oxford, 2012.

[9] Jonathan Ginzburg, Raquel Fernández, and David Schlangen. Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, 7(9):1–64, 2014.

[10] Philip Glenn and Elizabeth Holt. Introduction. In Philip Glenn and Elizabeth Holt, editors, *Studies of Laughter in Interaction*. Bloomsbury, 2013.

[11] Markku Haakana. Laughing matters: A conversation analytical study of laughter in doctor-patient interaction. 1999.

[12] Alexa Hepburn and Scott Varney. Beyond ((laughter)): some notes on transcription. In Philip Glenn and Elizabeth Holt, editors, *Studies of Laughter in Interaction*. Bloomsbury, 2013.

[13] Thomas Hobbes. *Leviathan, 1651.* Scolar Press, 1969.

[14] M.M. Hurley, D.C. Dennett, and R.B. Adams. *Inside Jokes: Using Humor to Reverse-Engineer the Mind.* The MIT Press, 2011.

[15] Sally Jackson and Scott Jacobs. Structure of conversational argument: Pragmatic bases for the enthymeme. *Quarterly Journal of Speech*, 66(3):251–265, 1980.

[16] Keith Oatley and PN Johnson-Laird. Cognitive approaches to emotions. *Trends in cognitive sciences*, 18(3):134–140, 2014.

[17] MJ Owren, N Trivedi, AS Schulman, and JA Bachorowski. Explicit and implicit evaluation of voiced versus unvoiced laughter. *Manuscript in preparation*, 2004.

[18] Fernando Poyatos. The many voices of laughter: A new audible-visual paralinguistic approach. *Semiotica*, 93(1-2):61–82, 1993.

[19] Stefano Predelli. Scare Quotes and Their Relation to Other Semantic Issues. *Linguistics and Philosophy*, 26:1–28, 2003.

[20] R. R. Provine. Laughter punctuates speech: Linguistic, social and gender contexts of laughter. *Ethology*, 95(4):291–298, 1993.

[21] T. Romaniuk. Interviewee laughter and disaffiliation in broadcast news interviews. In Philip Glenn and Elizabeth Holt, editors, *Studies of Laughter in Interaction*. Bloomsbury, 2013.

[22] Willibald Ruch and Paul Ekman. The expressive pattern of laughter. *Emotion, qualia, and consciousness*, pages 426–443, 2001.

[23] Ivan A. Sag, Thomas Wasow, and Emily Bender. *Syntactic Theory: A Formal Introduction (second edition)*. CSLI, Stanford, 2003.

[24] L. A. Sroufe and J. P. Wunsch. The development of laughter in the first year of life. *Child Development*, pages 1326–1344, 1972.

[25] Hiroki Tanaka and Nick Campbell. Classification of social laughter in natural conversational speech. *Computer Speech & Language*, 28(1):314–325, 2014.

[26] Jérôme Urbain and Thierry Dutoit. A phonetic analysis of natural laughter, for use in automatic laughter processing systems. In *Affective Computing and Intelligent Interaction*, pages 397–406. Springer, 2011.

Even, Comparative Likelihood and Gradability[*]

Yael Greenberg

Bar Ilan University, Ramat Gan, Israel
yaelgree@gmail.com

Abstract

A popular view of the semantics of *even* takes it to presuppose that its prejacent, p, is less likely than all its contextually relevant focus alternatives, q. In this paper I point out three novel problems for this 'comparative likelihood' view, having to do with (a) cases where *even p* is felicitous though p cannot be considered less likely than q, (b) cases where *even p* is infelicitous though p asymmetrically entails and is less likely than q, and (c) cases where *even* interacts with gradable predicates, indicating that merely requiring p to be higher on the scale than q is not enough to make *even p* felicitous. Instead, both p and q must also yield degrees which are at least as high as the standard of comparison.

In response to these problems I develop a revised scalar presupposition for *even* which resembles the semantics of comparative conditionals, and which requires that for a salient x, retrieved from p, and a salient gradable property G, (i) x's degree on G is higher in all accessible p worlds than in all accessible *q-and-not-p* worlds and that (ii) in the latter worlds this degree is at least as high as the standard on G. I show how this presupposition accounts for both traditional observations concerning *even*, as well as for the novel data and propose that the common presence of 'less likely' inferences with *even* can be indirectly derived from the common use of 'distributional' standards of comparison with gradable properties. A general contribution of the proposal, then, is in attempting to apply tools from research of gradability-based phenomena for a better understanding of scalarity-based phenomena.

1 A Brief Background on the Standard Semantics for *Even*:

The scalar presupposition of *even* is usually taken to be based on comparative likelihood, as in 1:[1]

(1) $||even||^{g,c} : \lambda C.\lambda p.\lambda w : \forall q \in C \ q \neq p \rightarrow p <_{likely} q \ . \ p(w) = 1,$
Where $C \subseteq ||p||^F \wedge |p||^O \in C \wedge \exists q \ q \neq p \wedge q \in C$

Given (1), $even(C)(p)(w)$ presupposes that p is **less likely** than all its distinct contextually supplied focus alternatives in C, and asserts that p is true in w.[2] An immediate advantage of this kind of entry is in explaining felicity contrasts as in (2) (where p is underlined in the respective C sets). Given the natural assumption that *winning gold* $<_{likely}$ *winning silver* $<_{likely}$ *winning bronze*, p in the felicitous (2a), but not in the infelicitous (2b) is indeed less likely than its focus alternative q in C:

---

[1]See e.g. [13],[20], [25], [16], [10], [7] and [4].
[2]For simplicity I ignore here the debated presence of an additive presupposition, in addition to the 'scalar' one, the debate about the universal quantification over alternatives, and the 'scope' / 'ambiguity' debate.

(2)   a. *John won silver. Bill even won [gold]$_F$*   (C: {*Bill won silver, Bill won gold*})

      b. *John won gold. Bill (#even) won [silver]$_F$* (C: {*Bill won gold, Bill won silver*})

In addition, the comparative likelihood view has other well-known achievements: It was shown to account for the interaction of overt *even* with various operators like DE operators (e.g. [16]), modals, covert *exh* ([5], [7]), non-monotone operators (e.g. [7]), and questions (e.g. [10]). In addition, it was also extended to account for the behavior of (some) Negative Polarity Items (NPIs), by assuming that their semantic structure involves a **covert** variant of *even*. (e.g. [16], [18], [4]).

   Despite these points of strength, however, various theories pointed out problems for this approach. E.g. in (3), *even* is felicitous though given the information about John's non-conformity, his reading a censored book does not seem less likely than reading other, alternative books:

(3)   *John is a political non-conformist. He even read [Manufacturing Consent]$_F$ although it has been banned by the censorship committee.*[21]

Such data led some theories to suggest that the scale for *even* should be based on "pragmatic entailment" "better informativeness" [14], "noteworthiness" [11], etc. instead of on (un)likelihood. However, these suggestions were usually not integrated into analyses of *even*. First, they remained intuitive and were not formally developed. Second, it was not clear that such data is really problematic for the "likelihood" view. For example, in (3) we might evaluate likelihood from the point of view of the average person (ignoring the information about John's non-conformity). Historically, then, the comparative likelihood view remained the strongest and most prominent theory of *even* on the market.

   In this paper I first consider (in section 2) three novel challenges for the comparative likelihood view, indicating that this view should be more seriously re-considered. In section 3 I offer a revised version of the scalar presupposition of *even*, which relies on contextually supplied gradability, and which makes reference to standards of comparison. Section 4 shows how the revised presupposition can handle the novel data. Section 5 summarizes.

2   Three Challenges for the Comparative Likelihood View of *Even*

2.1   First Challenge: More Cases where *Even p* is Felicitous though *p* is <u>not</u> Less Likely than *q*

The felicity of *even* in (4) ([9]) and in both (5a) and (5b) is harder to explain than in (3) using the comparative likelihood approach:

(4)   <u>*Client: I need a strong tool for this work. What materials are these two tools made of?*</u>
      <u>*Seller: Both are strong enough for what you need. The red one is made of strong aluminum and the blue one is* even *[made of steel]$_F$* [9].</u>

(5)   *My hat got stuck on a branch. I wonder whether John or Bill can help me fetch it.*

      a. (Context (a): the branch is 2.50m high) *Neither John nor Bill can fetch the hat. John is 1.70m. tall. He is definitely too short for that. And Bill is even [shorter]$_F$*

      b. (Context (b): the branch is 1.50m high) *Both John and Bill can fetch the hat. John is 1.70m tall. He is definitely tall enough for that. And Bill is even [taller]$_F$.*

Given the general perspective on what working tools are usually made of, p in (4) (*The blue tool is made of steel*) does not seem less likely, but MORE likely than q (*The blue tool is made of strong aluminum*). Nonetheless *even* is perfectly felicitous in (4). Regarding (5), to capture the felicity of *even* in both (5a) and (5b) using the comparative likelihood view one has to assume for (5a) *that Bill is shorter than 170 $<_{likely}$ that Bill is 1.70*, and for (5b) the opposite likelihood judgment holds, namely *that Bill is taller than 1.70 $<_{likely}$ that Bill is 1.70*. This seems unmotivated, given that the only difference between the two respective contexts concerns the height of the branch.

Our first interim conclusion, then, is that comparative unlikelihood of p is not a necessary condition for the felicity of *even p*.

2.2 Second Challenge: *Even p* with Entailed Alternatives [9]

A prediction of the comparative likelihood view ([16], [4], [7]) is that *even p* will be systematically felicitous whenever p asymmetrically entails q (unless p and q are contextually equivalent and hence equi-probable). This is because likelihood respects entailment: if p asymmetrically entails q, then unless they are contextually equivalent, p is less likely than q (true in fewer situations). But this prediction is not borne out. First, in many cases the felicity of *even p* with entailed alternatives varies. In (6a), for example, both giving birth to a boy and giving birth to a girl entail and are less likely than giving birth. Similarly, in (6b) both drinking whisky and drinking beer entail and are less likely than drinking alcohol. Nonetheless, in the indicated contexts case only one of them is fine:

(6) a. Context: Any princess who gives birth can stay in the palace. If she gives birth to a boy she also becomes a queen (i.e. on average 50% of those who give birth get to be queens):
   A: *What's happening with Princess Jane?*
   B: *She gave birth. She (even) gave birth to [a boy]_F / #[a girl]_F* ($\rightarrow$ **varied felicity**)

   b. Context: We were at a party where only two alcoholic drinks (beer and whisky) were served.
   A: *John drank alcohol in the party. He better not drive now*
   B: *Yea. He even drank [whisky]_F/#[beer]_F* ($\rightarrow$ **varied felicity**)

Moreover, with an entailed disjunction *(p or q)*, *even* is <u>systematically infelicitous</u>, as can be seen in (7):

(7) Context: We were at a party where only two alcoholic drinks (beer and whisky) were served.
   A: *John drank beer or whisky in the party. He better not drive now*
   B: *Yea. He even drank #[whisky]_F / #[beer]_F* ($\rightarrow$ **systematic infelicity**)

Our second interim conclusion, then, is that Comparative likelihood is not only not a necessary condition, but also not a sufficient condition for the felicity of *even p*.

2.3 Third Challenge: *Even* and Standards of Comparisons

Consider the felicity contrasts in (8). Crucially, while seller (a)'s utterance is felicitous given the indicated context and assuming a scale of physical strength, seller (b)'s and seller (c)'s utterances are not:

(8) (Context: plastic–<–aluminum–<–**standard**–<–iron–<–steel–– → *Physical strength*)
Client: I need a strong tool. What about the red and blue tools over there?

a. *Seller (a): Well, the red one is made of iron and the blue one is (even) made of [steel]$_F$*

b. *Seller(b): Well, the red one is made of plastic and the blue one is (??even) made of [aluminum]$_F$*

c. *Seller (c): Well, the red one is made of plastic and the blue one is (??even) made of [steel]$_F$*

I take these felicity contrasts to indicate that for *even p* to be felicitous, comparison between *p* and its alternatives *q* is not enough. More specifically, it is not enough that we end up with a higher degree with *p* than with *q* (on the relevant scale). Instead, with both *p* and *q* we should also end up with degrees which are at least as high as the **standard** on the scale. This intuition is supported by the similar felicity contrasts in (9):

(9) (John and Bill want to join our basketball team, where the standard height is 1.90m)
A: Well, what about John and Bill? Should we take them?

a. Agent (a): *Well, John is 1.95m tall. Bill is (even) 2.10. (We can take both).*

b. Agent (b): *Well, John is 1.65m tall. Bill is (??even) 1.75. (We shouldn't take them).*

c. Agent (c): *Well, John is 1.75m tall. Bill is (??even) 1.95. (We can take Bill).*

Further support for this intuition comes from another novel observation concerning a surprising interaction between *even* and comparatives: Unlike what usually happens,[3] when *even* associates with comparatives (based on relative adjectives) we get entailment to their positive forms, as seen in (10):

(10) a. *The blue tool is (even) [stronger than the red tool]$_F$.*
<u>Without *even*</u>: No inference: *The blue tool is strong / The red tool is strong (both can be weak)*
<u>With *even*</u>: Entailment: *The blue tool is strong / The red tool is strong (#both can be weak)*

b. *Bill is (even) [taller than John]$_F$.*
<u>Without *even*</u>: No inference: *Bill is tall / John is tall (...both can be short)*
<u>With *even*</u>: Entailment: *Bill is tall / John is tall (#...both can be short)*

Thus, whereas the data in section 2.1 and 2.2 seems to challenge the 'likelihood' component in the 'comparative likelihood' approach to *even*, the data in (8)-(10) seems problematic for the 'comparative' component in this approach, and in fact for any 'comparative' approach to *even*, requiring that *p* is merely 'higher' than *q* on a scale. Instead, we can see that both *p* and *q* must also 'lead to' (in a sense to be made precise below) degrees which are at least as high as the standard on the relevant scale.

3   A Revised Scalar Presupposition for Overt *Even*

Given the challenges considered above for the comparative likelihood approach, let me try to develop a revised presupposition for *even*. This presupposition is based on intuitive ideas in [22],

---

[3]E.g. [15].

according to which *even* ranks the alternatives by "correlating them with a graded property which is salient in the context" (p. 11).

[22] suggests that "alternatives $p_1...p_n$ are correlated with some graded property $q...(when)$ ...p_1 is the strongest argument for q and p_n the weakest" (p. 13).[4] This 'stronger argument' component by itself (in which [22] follows [14]) does not seem to work for *even*. Trying to make this component more precise, one can require that *even p* presupposes that p is a stronger argument for some salient goal H than q, and more formally that the probability of H given p is higher than its probability given q (cf. [2], [17], [24], [26]). But crucially, *even* is also felicitous when p and q are equally strong arguments for H, and more formally where the conditional probability of H given p and given q is identical. For example, *even* is perfectly felicitous in (5b) above, though the conditional probability of H (*Bill is suitable to get your hat*) is 1, both given q (Bill is 1.70 m tall and can definitely get the hat), and p (Bill is taller than 1.70).

In contrast, [22]'s intuitive 'correlation with a graded property' component seems more promising. To make this intuition more precise / workable, I will follow the spirit of [3]'s analysis of **comparative correlatives** like *The better Otto is prepared, the better his talk is* as **comparative conditionals**, as paraphrased in (11):

(11)   *In all accessible worlds w1, w2 where Otto's maximal degree of preparation in w2 > his maximal degree of preparation in w1, his degree of success in w2 > his degree of success in w1.*

In light of this line of thought, I propose that *even* $(C)(p)(w)$ is defined iff, given an accommodated salient gradable property G, determined by a salient goal in the discourse and / or the QUD (e.g. *How successful Bill is? How strong / suitable for this work this tool is?*) and an entity x, denoted by some non-focused / contrastive topic constituent in p (e.g. *Bill* in (2a) or *this tool* in (4)), the following holds:

(12) $\forall q \in C \; q \neq p \rightarrow \forall w1, w2 \; [w1Rw \wedge w2Rw \wedge w2 \in p \wedge w1 \in [q \wedge \neg p]] \rightarrow [the \; max \, (\lambda d2.G(d2)(x)(w2)) > the \; max \, (\lambda d1.G(d1)(x)(w1)) \wedge the \; max(\lambda d1.G(d1)(x)(w1)) \geq \mathbf{stand}_G]$

In prose, *Even* $(C)(p)(w)$ is defined iff for all distinct alternatives q in C the following two conjuncts are met: (a) x is more G in all accessible p worlds than in all accessible *q-and-not-p* worlds (the worlds where the exhaustified alternative to p holds), and (b) in the *q-and-not-p* worlds x is considered to have G (i.e. x's degree of G is at least as high the standard of G).

The revised presupposition, then, differs from the 'comparative likelihood' presupposition of *even* in three main features, seen in the representation in (13). First, unlike the latter presupposition, the revised one does not directly compare propositions p and q, but rather entities x in worlds where these propositions hold. Second, the dimension of the scale is not likelihood, but a variable dimension, based on the accommodated salient property G. Finally, the relationship between the measured entities is not merely comparative. Instead, their degree is also compared to the standard of comparison:

(13)   A schematic comparison between the two presuppositions:

    a. **The 'comparative likelihood' presupposition:**
      $- - - - - - - - q - - - < - - - p - - - - - - - - - - \rightarrow$ unlikelihood

    b. **The revised presupposition:** $q \wedge \neg p$ p
                                       $\downarrow$ $\downarrow$
      $- - - - - - - - - \mathbf{stand} - - - - x's \; degree - - - < - - - x's \; degree - - \rightarrow G$

_____

[4]Cf. [8] for a similar suggestion, using 'pragmatic strength'.

4 Illustration and Accounting for the Data

4.1 Basic Felicity Differences

Let us start with the felicitous (2a). Assuming x=Bill and G=successful we get the presupposition in (14):

(14) $\forall w1, w2$ [$w1Rw$ $\wedge$ $w2Rw$ $\wedge$ $w2 \in$ *Bill got gold* $\wedge$ $w1 \in$ [*Bill got silver* $\wedge$ $\neg Bill$ *got gold*]] $\rightarrow$ [*the max* $(\lambda d2.SUCCESSFUL(d2)(Bill)(w2) > the\ max\ (\lambda d1.$ $SUCCESSFUL(d1)(Bill)(w1)) \wedge the\ max\ (\lambda d1.SUCCESSFUL(d1)(Bill)(w1)) \geq$ **STAND**$_{SUCCESSFUL}$]

In prose: (a) Bill is more successful in the accessible worlds where he got gold, than in those where he got silver-and-not-gold, and (b) his degree of success in the accessible worlds where he won silver-and-not-gold is at least as high as the standard (i.e. he is still considered successful). Since both conjuncts are easily met, the presupposition is correctly predicted to be met as well, so the felicity of (2a) follows. In contrast, assuming again **x=Bill, G=successful** as above, the first conjunct in the presupposition of (2b), requiring that Bill is more successful in the accessible worlds where he got silver, than in those where he got gold-and-not-silver, is not met, thus correctly predicting its infelicity:

(15) $\forall w1, w2$ [$w1Rw0$ $\wedge$ $w2Rw0$ $\wedge$ $w2 \in$ *Bill got silver* $\wedge$ $w1 \in$ [*Bill got gold* $\wedge$ $\neg Bill$ *got silver*]] $\rightarrow$ [*the max $d2$* $(\lambda d2.SUCCESSFUL(d2)(Bill)(w2) > the\ max\ d1(\lambda$ $d1.SUCCESSFUL(d1)(Bill)(w1)) \wedge the\ max\ d1\ (\lambda d1.SUCCESSFUL(d1)(Bill)(w1))$ $\geq$ **STAND**$_{SUCCESSFUL}$]

However, one can wonder whether accommodating any G is not too flexible. For example, wouldn't accommodating G with a reversed scale, measuring degrees of UNscuccessfulness, wrongly predict the infelicitous (2b) to be felicitous? The answer seems negative, as can be seen in(16):

(16) $\forall w1, w2$ [$w1Rw0$ $\wedge$ $w2Rw0$ $\wedge$ $w2 \in$ *Bill got silver* $\wedge$ $w1 \in$ [*Bill got gold* $\wedge$ $\neg Bill$ *got silver*]] $\rightarrow$ [*the max $d2$* $(\lambda d2.UNSUCCESSFUL(d2)(Bill)(w2)) > the\ max$ $d1(\lambda d1.UNSUCCESSFUL(d1)(Bill)(w1)) \wedge the\ max\ d1\ (\lambda d1.UNBSUCCESSFUL$ $(d1)(Bill)(w1)) \geq$ **STAND**$_{SUCCESSFUL}$]

While in this case the first conjunct is met (Bill is indeed more unsuccessful in the accessible worlds where he got silver, than in those where he got gold-and-not-silver), the second one, requiring that Bill's degree of being 'unsuccessful' in the accessible worlds where he got gold-and-not-silver is at least as high as the standard, fails. This is because winning gold is being maximally successful. The presupposition for (2b), then, fails both with a non-reversed and a reversed scale, since in both cases one of the conjuncts is false.

Moreover, we can now correctly predict that unlike (2b), (17) will be better:

(17) *I think Bill will win silver, or even [bronze]$_F$* (i.e. he is not that good)

Notice that the felicity contrast between (2a) and (17) pose another problem for the likelihood view, since winning bronze is more likely than winning silver, just as winning silver is more likely than winning gold. Hence, <u>both</u> sentences are predicted to be equally infelicitous. In contrast, while, as seen above, the revised presupposition fails for (2a) it can be met for (17), accommodating a new standard of unsuccessfulness where getting anything below gold is considered disappointing / unsuccessful.

Finally, the revised presupposition can also account for the felicity differences in e.g. (8) above. Assume that for (8a)-(8c) we accommodate x=*the blue tool*, G=**physically strong**, then in all these cases the blue tool is considered stronger in the p worlds than in the *q-and-not-p* worlds. However, only for (8a) the blue tool's degree of strength is at least as high as the standard in the *q-and-not-p* worlds (and of course also in the p worlds).

4.2 *Even* with Entailed Alternatives (i.e. alternatives which are entailed by p)

Remember that the 'comparative likelihood' approach predicts **systematic felicity** of *even p* with asymmetrically entailed alternatives.[5] In reality, though, we saw above that this prediction fails in two cases: First, with some entailed alternatives, as in (6), we actually get **varied felicity**. Second, with entailed disjunctive alternatives, as in (7), we get **systematic infelicity**.

The revised presupposition can rather easily account for cases of **varied felicity**. Assume, for example, that for (6a) above we accommodate x=Jane, G=***important***. Then the sentence is felicitous with 'a boy' since given the indicated context, Jane's degree of importance is **higher** in the accessible worlds where she gave birth to a boy than in those where she gave birth to a child who is not a boy (i.e. to a girl). In contrast, the sentence is infelicitous with 'a girl' since Jane's degree of importance is **LOWER** in the worlds where she gave birth to a girl than in those where she gave birth to a child who is not a girl (i.e. to a boy).

However, with the systematically infelicitous cases (with disjunctive alternatives) there seems to be a problem. Compare again (6b), with varied felicity and (7) with systematic infelicity. There is no problem explaining why the version with 'beer' is bad for both cases. But on the surface the version with 'whisky' should be felicitous for both cases. For example, if we assume that in both cases with 'whisky' we accommodate *x=Bill*, **G=unsuitable for driving**, then (6a) will presuppose that John's degree of unsuitability for driving in the accessible whisky worlds is higher than in the alcohol-but-not-whisky worlds (i.e. than in the beer worlds), and in (7) we will presuppose that John's degree of unsuitability for driving in the accessible whisky worlds is higher than in the beer-or-whisky- but-not-whisky worlds (i.e. again than in the beer worlds). More intuitively, in both cases it seems that the presupposition ends up as requiring a higher degree of unsuitability for driving in the accessible worlds where John drank whisky than in those in which he only drank beer, and in both cases this presupposition is supposed to be equally met. What is then the problem with this presupposition in (7)? What is so special about disjunctive alternatives in this case?

The answer seems to lie in a problem, independent of *even*, of quantifying over accessible worlds where John drank beer or whisky and not whisky, or, more generally over worlds where $[p\ or\ q] \wedge \neg p$ hold. That quantifying over such worlds is problematic is both empirically and theoretically supported. Empirically, notice that unlike conditionals like (18a), ones like (18b) (with no *even*) are odd:

(18) a. *If John drank alcohol but not whisky, the situation is not hopeless*

   b. *#If John drank beer or whisky but not whisky, the situation is not hopeless.*

Theoretically, conditionals (e.g. counterfactuals) with disjunctive antecedents like: *If p or q were true, then r*, are known to be interpreted as conjunctions of conditionals like *If p were true then r, and if q were true then r*. [1] interprets such equivalences to be due to the alternative-based (Hamblin style) semantics of *or*, according to which $p\ or\ q$ introduces a set of alternatives $\{p,q\}$, in which each disjunct is equally 'visible' in the interpretive process. This is done by

---

[5]Again, unless these alternatives are contextually equivalent to p.

using pointwise functional application, where functions are applied to each input in the set of alternatives.

Turning back to *even*, the crucial point to note is that the revised presupposition developed above is conditional in nature. Specifically, when the alternative to p is q, it has the schematic form: $\forall w1, \forall w2[w2 \in p \ \wedge \ w1 \in [q \wedge \neg p]] \rightarrow [\ldots\ldots]$. Now, when the alternative to p is a disjunction *q or p* (instead of a simple q), the quantification over worlds w1 gives us the substructure $\forall w1[w1 \in [[q \ or \ p] \ \wedge \ \neg p]] \rightarrow [\ldots\ldots]$. But given the behavior of disjunctive antecedents, this ends up as the conjunction of conditionals $[\forall w1[w1 \in [q \wedge \neg p] \rightarrow [\ldots\ldots]] \wedge [\forall w1[\ldots w1 \in [p \wedge \neg p]] \rightarrow [\ldots\ldots]]$. For example, quantifying over accessible worlds where *John drank beer or whisky but not whisky* will result in quantifying over accessible worlds where *John drank beer and not whisky* and accessible worlds where *John drank whisky and not whisky*. While the first set of accessible worlds can be constructed with no problem, the second is problematic, as we actually end up with an empty set of worlds, and hence with vacuous quantification over worlds in the presupposition. This is what seems to lead to infelicity.

The systematic infelicity of *even* with disjunctive alternatives, then, can be derived from the combination of the 'conditional'-like nature of the revised presupposition of *even* and the independently motivated behavior of disjunctions in the antecedent of conditionals.

4.3 Why do Sentences with *Even* so Often (but not always) Lead to a 'Less Likely' Inference?

We saw before that the 'less likely' inference is not always present with *even* ("Comparative unlikelihood is not a necessary condition for the felicity of *even p*"). However, such inferences are still very common with *even*. In (19), for example, we naturally infer that getting accepted to S-U is less likely than to R-U:

(19)   *John got accepted to R-University. He even got accepted to [S-University]$_F$*

Using the revised presupposition, the presence of this inference can be derived in two ways. One, direct, way is to accommodate a gradable property G which measures degrees to which x (John in this case) is surprising us. But there is an indirect, and perhaps more promising way, namely to rely on the fact that in many cases[6] standards of comparison are <u>distributional</u>, i.e. determined by some central point, e.g. the median / average point in the comparison class.

Assume for example, that we accommodate **G=successful** for (19), so its presupposition is as in (20):

(20) $\forall w1, w2[w1Rw0 \wedge w2Rw0 \wedge w2 \in John \ got \ accepted \ to \ S - university \wedge w1 \in [John \ got \ accepted \ to \ R - university \wedge \neg John \ got \ accepted \ to \ S - university]] \rightarrow [the \ max(\lambda d2.SUCCESSFUL(d2)(John)(w2)) > the \ max(\lambda d1.SUCCESSFUL(d1)(John)(w1)) \wedge the \ max(\lambda d1.SUCCESSFUL(d1)(John)(w1)) \geq \mathbf{STAND}_{SUCCESSFUL}]$

Suppose further that the standard of success is distributional, i.e. determined by the median point. Then the farther your degree of success is from the standard, the farther it is from the median. Intuitively, in this case in order for the presupposition to be met, less people are accepted to S-U than to R-U. Hence John's getting accepted to S-U is understood as less likely than his getting accepted to R-U.

But crucially, standards are not always 'distributional'[7]. In particular, standards can be 'functional', i.e. determined by the requirements of a given situation. Two examples of using such standards are (21) and (22), from [12] and [23], respectively:

---

[6]At least with open scale adjectives
[7]Not even with open scale adjectives

(21) *Let's buy this book for $7? No, that's expensive. We only have $6.*

(22) *Three of the boards were cut to exactly the right length, but the fourth one was long*

Our prediction, then, is that in contexts where functional standards are being used, and when they do not correlate with distributional standards, the revised presupposition of *even* can be met with no necessary 'less likely' inference. This prediction seems to be borne out. For example, in (4) above, the standard of physical strength is determined by the requirements of 'this work', and does not correspond to the median degree in the comparison class (of tools). Hence, being farther away from this standard does NOT indicate being father away from the median, so *even* is perfectly felicitous here although there is no implication that 'being made of steel' is less likely than 'being made of strong aluminum'. Similar considerations hold for (5a)-(5b), where the standard is determined by the (un)suitability to fetch the hat.

If this suggestion is on the right track, the question is why 'less likely' inferences are so common / found in default contexts with *even*. A potential answer to consider is that this is because distributional standards are so common / found in default contexts with gradable properties. To quote [12], "Evidently, the functional standard is less accessible for the positive form of gradable adjectives, as it seems to require special contexts; the distributional standard is far more salient for it." ([12], p.(6)).

5 Summary

In this paper we examined the 'comparative likelihood' presupposition for *even* and pointed out three novel types of data which pose challenges for both the 'comparative' and the 'likelihood' components in it. We then developed a revised presupposition (inspired by previous intuitions in [22]), which accounts for the novel data, as well as for traditional felicity differences, explains why 'less likely' inferences often (but not always) arise with *even* (especially in default contexts), and more generally - applies tools from research on gradability to research on scalar operators.

Notice, though, that the revised presupposition relies on a limited set of data, namely overt *even* in matrix and UE environments. In future research I intend to examine the application of this presupposition to the behavior for overt *even* in (Strawsson) DE and non-monotone environments, to covert *even* with (some) NPIs, and for other scalar operators, e.g. some of the readings of (overt and covert) exclusives like *only* and *merely* in English (cf. [6]), or *rak* and *stam* in Hebrew (cf. [19]).

References

[1] Luis Alonso-Ovalle. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, 32:207–244, 2009.

[2] Jean-Claude Anscombre and Ducrot Oswald. *L'argumentation dans la langue*. Pierre Mardaga, Liège: Bruxelles, 1983.

[3] Sigrid Beck. On the semantics of comparative conditionals. *Linguistics and Philosophy*, 20:229–271, 1997.

[4] Gennaro Chierchia. *Logic in Grammar. Polarity, Free Choice, and Intervention.* Oxford University Press, 2013.

[5] Gennaro Chierchia, Danny Fox, and Benjamin Spector. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Paul Portner, Claudia Maienborn, and Klaus von Heusinger, editors, *Semantics: an international handbook of contemporary research*. Mouton De Gruyter, 2011.

[6] Elizabeth Coppock and David Beaver. Principles of the exclusive muddle. *Journal of Semantics*, 31(3):371–432, 2013.

[7] Luka Crnič. Against a dogma on npi licensing. In Luka Crnić and Uli Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 1, pages 117–145. MITWPL, 2014.

[8] Volker Gast and Johan van der Auwera. Scalar additive operators in the languages of europe. *Language*, 87(1):2–54, 2011.

[9] Yael Greenberg. A novel problem for the likelihood-based semantics of *even*. *Semantics & Pragmatics*, in press.

[10] Elena Guerzoni. *Why Even Ask? On the Pragmatics of Questions and the Semantics of Answers.* PhD thesis, MIT, 2003.

[11] Elena Herburger. *What Counts: Focus and Quantification.* MIT Press, Cambridge, MA, 2000.

[12] Olga Kagan and Sascha Alexeyenko. Degree modification in russian morphology: The case of the suffix-ovat. In Yehuda N. Falk, editor, *Proceeding of Israel Association for Theoretical Linguistics*, volume 26. Bar Ilan University, 2010.

[13] Lauri Karttunen and Stanley Peters. Conventional implicature. In D. A. Dinneen and C. K. Oh, editors, *Syntax and Semantics*, volume 11: Presuppositions, pages 1–56. Academic Press, New York, 1979.

[14] Paul Kay. Even. *Linguistics and philosophy*, 13:59–111, 1990.

[15] Chris Kennedy and Louise McNally. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81:345–381, 2005.

[16] Utpal Lahiri. Focus and negative polarity in Hindi. *Natural Language Semantics*, 6:57–123, 1998.

[17] Arthur Merin. Information, relevance and social decision-making. In L. Moss, J. Ginzburg, and de Rijke M., editors, *Logic, Language, and Computation*, volume 2, pages 179–221. CSLI Publications, Stanford, CA, 1999.

[18] Kimiko Nakanishi. Even, only, and negative polarity in japanese. In *Proceedings of SALT 16*, 2006.

[19] Dina Orenstein and Yael Greenberg. Exclusives in Hebrew: A core meaning and varying parameters. In *Proceedings of IATL*, volume 29, 2014.

[20] Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1:75–116, 1992.

[21] Hotze Rullmann. Even, polarity, and scope. In Martha Gibson, Grace Wiebe, and Gary Libben, editors, *Papers in experimental and theoretical linguistics*, volume 4, pages 40–64. Department of Linguistics, University of Alberta, 1997.

[22] Hotze Rullmann. What does *even* even mean? MS. UBC., 2007.

[23] Stephanie Solt. Comparison to arbitrary standards. In Ana Aguilar Guevara, Anna Chernilovskay, and Rick Nouwen, editors, *Proceedings of Sinn und Bedeutung 16*, volume 2, pages 557–570, Cambridge, MA, 2012. MIT Working Papers in Linguistics.

[24] Robert van Rooij. Cooperative versus argumentative communication. *Philosophia Scientia*, 2:195–209, 2004.

[25] Karina Wilkinson. The scope of even. *Natural Language Semantics*, 4:193–215, 1996.

[26] Gregoire Winterstein. What but-sentences argue for: An argumentative analysis of but. *Lingua*, 122:1864–1885.

An argument for vagueness with holes

Timothy W. Grinsell

University of Chicago
Chicago, IL, USA
twg@uchicago.edu

Linguistic vagueness is a consequence of aggregating many judgments into one. For example, whether something is a *heap* depends on judgments along at least two dimensions, height and width. Results from social choice theory—a branch of economics dealing with collective decision making—show that such judgment aggregations face significant limitations. Topologically, these limitations stem from "holes" in the structure of multidimensional domains over which judgment aggregations occur. Semantically, these limitations manifest as vagueness effects like susceptibility to the sorites paradox.

1 Vagueness Distinguished

In philosophical and linguistic studies of vagueness, there is widespread agreement as to the core vagueness phenomena: participation in the sorites paradox, borderline cases, and higher-order vagueness [24, 15]. In one way or another, these phenomena reflect the difficulty of drawing a boundary between a predicates true and false applicationsbetween, for example, the *tall* things and the non-*tall* things.

The sorites paradox, also known as the paradox of the heap, is illustrated in (1).

(1)   Premise 1: A pile of 10,000 grains of sand is a heap.
       Premise 2: A heap of sand minus one grain is a heap.
       Conclusion: A pile of 1 grain is a heap.

This paradox, most important of vagueness phenomena, is traditionally attributed to Eubulides of Miletus in the 4th century BCE. As the structure of the paradox reveals, it is applicable to a wide range of natural language expressions. Common nouns like heap are susceptible, as are gradable adjectives (so-called because they can appear in comparative constructions) like tall and red, as are adverbs (very), quantifiers (many), verbs (start), proper names (Chicago), and definite descriptions (the border between Illinois and Iowa) [23].

Vague predicates also have "borderline cases." Borderline cases reflect an intuition of indecisiveness about the truth of a proposition. For example, even if we know Clarences exact height—5 feet 11 inches—we are not sure if *Clarence is tall* is true. Not only are borders hard to find between the extension and anti-extension of a predicate, but they are also hard to find between, for instance, *tall* and *definitely tall*. This is the essence of higher-order vagueness: boundarylessness in a predicates extension.[1]

---

[1]The problem of boundarylessness separates vagueness from ambiguity or underspecification. An ambiguous word like bank has two (or more) semantically distinct meanings; disambiguation requires solving a mapping problem between sound (or some other representation) and meaning. In contrast, vague predicates involve meanings with unclear boundaries; interpreting a vague predicate requires solving a sorting problem among the relevant entities. Thus, ambiguous words do not display vagueness effects like borderline cases, at least not on account of their distinct meanings. For example, there is generally no borderline case between "financial institution" and "shore of a river" implicated by each use of bank. Nor is vagueness something like generality or underspecification. The utterance *a woman wrote Middlemarch* is perhaps underspecified for the author role, but it is not vague in terms of the sorites or borderline cases (at least not because the proposition uses the phrase *a woman* rather than *George Eliot*) [24].

2   Camping and the Sorites

I follow [23] in adopting a choice functional approach to adjectival semantics. Building on the framework of [16], van Rooij treats gradable adjectives as choice functions that select entities from the comparison class (the set of entities somehow relevant to the interpretation of the gradable adjective).[2] The gradable adjective P can be "thought of as a choice function, selecting the *best* elements of [the comparison class] c" [23, 140].

For each subset S of available options X, a *choice function* assigns to each S some element or elements of S. The rule by which the choice function decides which elements to assign is called the *choice rule*. For my purposes, then, the set of available options X will be the set of entities that constitute the comparison class.

The challenge in accounting for gradable adjectival semantics is figuring out the choice rule. For example, on one prominent theory of adjectival semantics, gradable adjectives like *tall* are measure functions that map an entity to a position on a one-dimensional scale of height [14]. These measure functions combine with a phonologically null morpheme to derive the denotation in (2).

(2) $[\![\text{tall}]\!] = \lambda x.\text{tall}(x) \geq \mathbf{s}(\text{tall})$

Loosely paraphrased, x is tall if x's height stands out relative to some "standard of comparison," which represents the cutoff point between positive and negative extensions. In this case, the relevant choice rule is the mechanism that helps us figure out how to set the standard of comparison—how, that is, to sort the entities that fall into the positive denotation from those that don't. I will adopt this view of the choice rule in what follows.

2.1   Adjectives as Collective Choice Functions

Adjectives like *tall* are not only choice functions, they are *collective* choice functions. These adjectives aggregate multiple potential standards into a single standard.[3]

There have been multiple proposals for how the standard $\mathbf{s}(\text{tall})$ is determined. The heights of the relevant objects are important, of course, but some norm-based cut-off (like the mean) is inadequate to account for vagueness effects [11]. As Fara puts it, "the property expressed context-invariantly by *tall* is a property which is such that whether a thing has it depends not only on heights, but on other things as well" [8]. For Fara, this includes "what our interests are." Others argue that a "stand-out" relation plays a role [15];[4] others, the distribution of

---

[2]For example, a sentence like *Ruth is tall* appears to make little sense unless we are comparing Ruth's height to the height of relevant individuals [16]. The set of relevant individuals is the comparison class.

[3]In some ways, this is the essential insight of supervaluationism [7]. [7] identifies two sources of judgment aggregation: verdict aggregation (as in *tall*) and dimension aggregation (as in *healthy*). One distinction between the two types of aggregators is the comparative: *taller* is not vague, while *healthier* may be. Here, I am treating adjectives like *tall* as (something like) dimension aggregators, though I leave open the possibility that verdict aggregation may be playing a role as well. One important reason for treating *healthy* and *tall* similarly is the fact that, at least on one view, the comparative *healthier* is not vague. If a possesses certain measures on all of its dimensions, and b possesses the same measures on all dimensions save one—on which b measures slightly less than a—then a is healthier than b. In this case, a small change in some value clearly demarcates true uses of *healthier* from false uses.

[4]As [15] shows, implicit comparisons like (1a) are incompatible with "crisp judgments," judgments based on small but noticeable differences in degree. In (1a), the positive form *long* conveys the fact that there is an asymmetric ordering between two objects along some dimension, just like the explicit comparative form *longer* (1b). In particular, both (1a) and (1b) can be used to make a claim about a 100-page book in opposition to a 50-page book. However, implicit comparison is infelicitous in contexts requiring crisp judgments. For example, (1a), but not (1b), is infelicitous when used to make a claim about a 100-page book and a 99-page book.

heights in the comparison class [22]; others, the metric distance from a prototype [13, 6]; and still others, the shape of the probabilistic distribution of heights [17].

Contextualist, many-valued logical, and supervaluationist approaches to adjectival semantics also reflect this ambivalence. Contextualist analyses like [23], for instance, hold that the adjectival standard constantly changes. "The basic idea behind contextualism is that vagueness is a diachronic phenomenon, which only emerges when we consider the semantic state of a language over time (or more generally, over multiple instances of interpretation)" [21, 113]. Contextualist accounts place particular emphasis on the act of interpretation, arguing that the act itself changes the semantic "facts on the ground." Many-valued approaches proliferate truth values rather than standards, but the effect on interpretation mimics contextualist theories [5, 3]. And supervaluationist approaches explicitly rely on an interpretation function that simultaneously takes into account multiple adjectival standards [9, 11, 20].

Instead of choosing among these competing approaches, then, I associate *tall*'s adjectival standard with with a vector of several values—a norm-based potential cutoff, a threshold value, a probability value associated with the distribution of heights, and more. For example, *tall*'s standard may be associated with a vector in two or more dimensions, as in (3).

(3) $\mathbf{s}(\text{tall}) = (\text{height norm, ordinal rank in comparison class}, \dots)$

I will refer to a particular vector of values $(a, b, \dots)$ as a *cutoff.*

I assume the availability of multiple potential cutoffs in any given context. There are multiple ways to rank these cutoffs. For example, one ranking might prioritize distance from a prototype and then ordinal rank in the comparison class, while another might prioritize (positive or negative) ordinal rank in the comparison class alone. I use the term *standards* to refer to these ways of ranking. This recapitulates the central insight of supervaluationism, in which the standard-setting function looks to multiple different ways of "making precise" a vague predicate. Moreover, the availability of multiple different standards in the same context accounts for the felicity of seeming contradictions like (4a). On the present view, (4a) is really something like (4b). Contrast this with (4c), which is a true contradiction.

(4)   a. Clarence is tall and Clarence is not tall.

         b. In one respect, Clarence is tall, and in another respect, Clarence is not tall.

         c. # In one respect, Clarence is tall, and in that same respect, Clarence is not tall.

The semantic task for adjectives like *tall*, then, is to aggregate choices among multiple standards into a single standard. This becomes the "standard of comparison." Gradable adjectives like *tall* are therefore collective choice functions, and they are subject to the limitations of collective choice.

2.2   The Limitations of Collective Choice

Social choice theory, the branch of economics dealing with the aggregation of judgments, places significant limits on whether such aggregations are possible. Arrow's Theorem represents these

---

(1)   a. This book is long compared to that book.

         b. This book is longer than that book.

As Kennedy explains, "Since ... *long* is true of an object if and only if it stands out in length in the context of evaluation, an assertion of [(1a)] involves a commitment to the highly unlikely position that a difference of one page could actually be relevant to whether a book of the size of the two under consideration stands out in length or not" [15, 18].

Figure 1: Aggregation of location preferences along a circular lake.
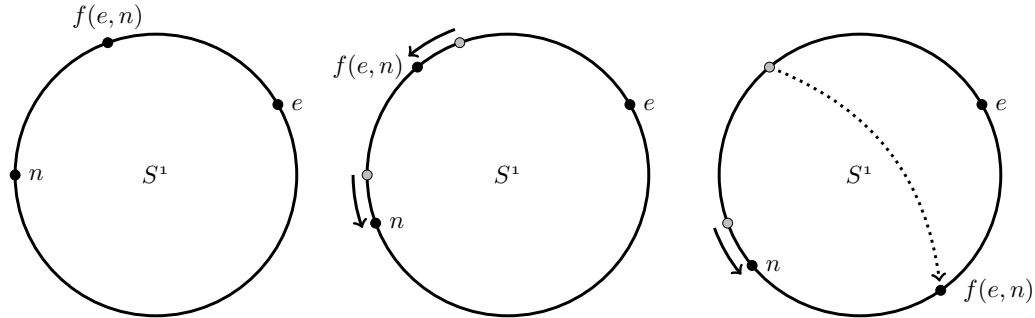
limits. It states that there is no collective decision procedure that respects certain reasonable assumptions and avoids intransitivity [1].

Generalizing Arrow's Theorem, Chichilnisky ([4]) provides a topological proof of the limits of collective choice. This proof is informally presented by [2]. Suppose Nino and Elena want to go camping along the shore of a perfectly circular lake. Nino and Elena may prefer the same geographic location along the shore, or they may not. If they agree on the location, that is where they will camp. Call this feature of the decision rule "unanimity." And the decision will not depend on who chose what: if Elena picks location 1 and Nino picks location 2, the outcome (whatever it is) will be the same as if Nino picked location 1 and Elena picked location 2. Call this feature of the decision rule "anonymity." Finally, the decision rule should be "relatively insensitive to small changes in individual preference" [4, 337]. This last requirement brings a kind of stability into the decision rule [18]. Call this feature "continuity."

Let S^1 denote the shore of the lake, let e be the location Elena picks, n the location Nino picks, and let $f(e, n)$ be the compromise choice, the result of aggregating Elena and Nino's preferences. Then a decision rule that chooses $f(e, n)$ such that $f(e, n)$ represents the unique shortest (arc) distance between e and n is both unanimous and anonymous. It is unanimous because, if $e = n$, then $f(e, n) = e = n$. It is anonymous because $f(e, n) = f(n, e)$.

However, it is not continuous. Holding e fixed, as in Fig. 1, if Nino's preferred location n moves continuously in a counterclockwise direction, so does $f(e, n)$—until, that is, n reaches the antipode of e, in which case $f(e, n)$ abruptly jumps to the other side of the lake. Chichilnisky's theorem generalizes this behavior.

Chichilnisky's theorem There is no continuous aggregation rule $f : S^1 \times S^1 \to S^1$ that satisfies unanimity and anonymity.

In terms of our example, there is no "fair" way to aggregate the location preferences along the shore of the lake. Any way we try to give Nino and Elena an equal say over the outcome, we will be stymied so long as we treat closely located camping spots similarly.

3   Vagueness Effects and the Structure of the Aggregation Domain

To construct an analogy between Chichilnisky's social choice paradox and the sorites paradox, it remains to show that the space of standards is analogous to a circular lake, and that adjectives

obey the relevant analogues to unanimity, anonymity, and continuity.

3.1   The Space of Standards Has a Hole

First, it is possible to interpret adjectival standards as representing preferences in a circular (or spherical) choice space. An aggregation of these standards constitutes a collective choice subject to Chichilnisky's result.

Take the choice space X to be a subset of $\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} = \mathbb{R}^n$. In Chichilnisky's native habitat of consumer theory, the elements of the choice space may be bundles of goods, like 4 bottles of wine and 3 bottles of beer, represented by the ordered pair $(4, 3)$. In our interpretation, the elements will be bundles of information of the type that determines potential cutoffs. Recall (a modified version of) the possible vector for the cutoff of *tall* (3).

(3) $\mathbf{s}$(tall) = (height norm, ordinal rank in comparison class)

An ordered tuple $(4, 3)$ in this context would stand for a height of 4 (in relevant units) and an ordinal rank of 3 in the comparison class. This represents a potential cutoff between the extension and anti-extension of a gradable adjective.

To interpret preferences over X—to answer what it means to prefer one cutoff over another—we may consider any reasonably consistent (i.e., transitive, complete, and continuous) choice rule. This choice rule corresponds to an adjectival standard. And different standards may rank different cutoffs differently. For example, a probability-based standard and a prototype-based standard may assign different values (and thus different relative rankings) to cutoffs $\vec{v_1}$ and $\vec{v_2}$. The diversity of adjectival standards likely means that different standards will prefer different cutoffs.[5]

Moreover, we expect the relevant standards to generate (linear) indifference sets or equivalence classes. A norm-based standard may rank the $(5, 4)$ tuple equivalently to the $(5, 3)$ tuple for the simple reason that the norm-based standard only cares about mean height, not relative position in the comparison class. The "gradient vector" of these indifference sets indicates the direction of greatest increase in along the relevant standard, as $\vec{v_{s_1}}$ and $\vec{v_{s_2}}$ in Figs. 2(a) and 2(b). In other words, the gradient vector points in the direction of cutoffs ranked highest by the relevant standard.

Finally, we consider ordinal preferences among the standards as a simplifying assumption. Therefore, we normalize these gradient vectors to length 1, ignoring preference strength. Now, without loss of generality, we can lift these vectors and place them at the origin of a unit circle S^1 (Fig. 2(c)). This makes it easy to see that each point on S^1 represents a standard—namely, the standard whose unit vector points to that location on S^1 [18].

An aggregation rule is a map F (as in 5) from a tuple of standards to a single standard.

(5) $F : S^1 \times S^1 \times \ldots \times S^1 \longrightarrow S^1$

This recalls the camping problem (Fig. 1). The set of standards involved in the interpretation of an adjective like *tall* can be represented as points along the boundary of a circle, just like camping spots along a lakeshore.

And just like the lakeshore, the space of adjectival standards has a "hole" in it. In the camping problem, the campers' choices were limited to the shore of the lake—they could not decide to camp *in* the lake. This lacuna in the domain of the aggregation function is topologically a hole. Similarly, in the present example, the space of adjectival standards consists of the boundary of the unit circle S^1. As I show in subsection 3.3, it is the hole-y structure of the

---

[5]Indeed, Chichilnisky's result only works for suitably diverse voters.

Figure 2: Gradient vectors from (a) and (b) in a unit circle.

(a)   (b)   (c)
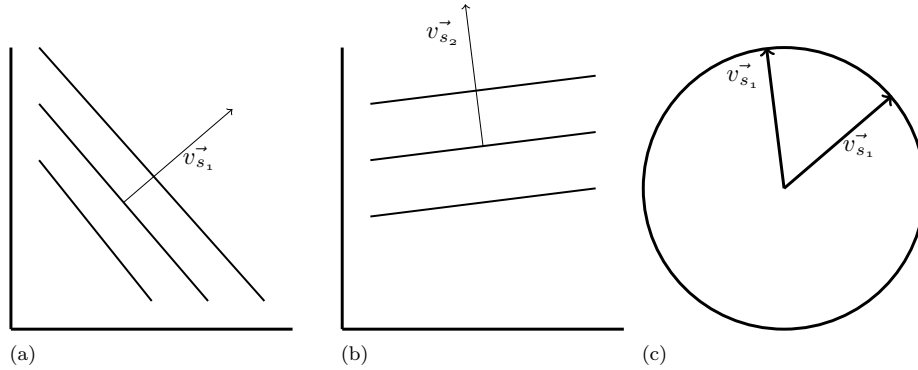
space of standards that is ultimately responsible for Chichilnisky's result and, consequently, vagueness effects.

3.2   Chichilnisky's Assumptions Are Satisfied

Accordingly, if the map F in (5) is continuous, unanimous, and anonymous, then F does not exist. Formally, the three assumptions unanimity, anonymity, and continuity look like (6).

(6)   a. **Unanimity:** for each point $p \in S^1$, $F(p, p, \ldots, p) = p$.

        b. **Anonymity:** for any $p, q \in S^1$, $F(p, q) = F(q, p)$.

        c. **Continuity:** for any map $F : \mathbb{R}^m \to \mathbb{R}^k$, the map F is continuous at $p \in \mathbb{R}^m$ if

           i. for each $\epsilon > 0$, there exists a $\delta > 0$ such that $|F(q) - F(p)| < \epsilon$ as soon as $|q - p| < \delta$.[6]

Unanimity and anonymity appear to be reasonable constraints on the choice of adjectival standard. For example, unanimity implies that if every relevant standard would rank x as tall, then x is tall.[7] Similarly, if every relevant standard would judge x as not tall, then x is not tall. Anonymity reflects the idea that there is no contextually relevant standard that is somehow more important than another. Formally, the aggregation rule is invariant under permutations of the standards in its domain. At least for some uses of adjectives like *tall*, this seems like an uncontroversial assumption.

    Continuity is more controversial. Continuity is sometimes explained as "stability." For example, if Nino suddenly changes his opinion and claims that his favorite camping site is a site next to his previous choice, then the output of the aggregation rule should change at most

---

[6]This is only one possible definition of continuity consistent with the theorem. The definition of continuity is dependent on the particular topology imposed on the space of standards (that is, what constitutes the "open" sets).

(1)   **Continuous map**: Let X and Y be topological spaces. A map $f : X \to Y$ is called continuous if the inverse image of opens sets is always open.

The particular topology need only reduce to the Euclidean topology, and "any reasonable topology, certainly any topology which has been used on preference spaces, satisfies this condition" [10, 3].

[7]This is analagous to "supertruth" in supervaluationism [24].

to a site next to the previous one. As Lauwers ([18, 454]) explains, "Continuity is a form of stability of the social outcome with respect to small changes in the individual preferences."

Tolerance is just this kind of stability. The notion of tolerance (7) has played a prominent role in theories of vagueness.[8]

(7) a. Suppose the objects a and b are observationally indistinguishable in the respects relevant to P; then either both a and b satisfy P or neither of them does.   [12]

b. $P(a) \land a \sim_P b \to P(b)$

Vague predicates are said to be tolerant, as the second premise of the sorites paradox (8) demonstrates.

(8) Premise 1: A 2m tall woman is tall.
Premise 2: Any woman 1mm shorter than a tall woman is tall.
Conclusion: Therefore, a 1m tall woman is tall.

A vague predicate like *tall* is tolerant in the sense that small differences in height seemingly do not change the value of $[\![tall(woman)]\!]$ from 1 to 0. Put another way, as the heights of x and y converge, the values $[\![tall(x)]\!]$ and $[\![tall(y)]\!]$ also converge. This implies that a small change in height should not be the difference between the *tall* entities and the non-*tall* entities. That is, vague predicates are "stable" in the manner demanded by continuity. Thus, the principle of tolerance—a key property of the sorites premise—approximates the work of continuity.[9]

When defined in terms like (6c), continuity emphasizes what Saari ([19, 57]) calls "local behavior": if x_o and x_n are so close together that they are difficult to distinguish, so must be $f(x_o)$ and $f(x_n)$ (where f is a continuous function). But this ignores the global structure of the domain, and when the local structure and the global structure disagree, problems arise. Saari ([19, 57]) provides the following example: "If we lived on a huge circle and used only local information, we would view the circle as being, essentially, a straight line." And the camping problem demonstrates the perils of treating circles as lines.

3.3 Vagueness with Holes

If the domain of the aggregation function contains holes, then unanimity, anonymity, and continuity are mutually incompatible. Vagueness effects result from this incompatibility. For example, the sorites paradox represents a tradeoff between conflicting assumptions. The first premise and the conclusion reflect the assumption of unanimity: every standard agrees that a 2m tall woman is *tall* and that a 1m tall woman is not *tall*. The sorites premise reflects continuity and anonymity. According to continuity, a small change in a tall woman's height should not take her from *tall* to not *tall*. According to anonymity, no contextually relevant standard is somehow more important than another. The sorites paradox is paradoxical because it rests on inconsistent assumptions.[10]

---

[8]See [5] for a recent tolerance-based approach.

[9]Despite this affinity between tolerance and continuity, [21] argues that continuity is too strong a constraint to model tolerance. First, continuity overgenerates the number of vague predicates. Second, sometimes domains are discrete. For example, the sorites premise for *heap* works with grains of sand, not fractional grains of sand. Topologically speaking, every discrete domain could support a continuous function, but we want to distinguish between "discrete" vague predicates like *heap* and "discrete" non-vague predicates like *has 100 hairs or fewer*.

Instead, Smith ([21]) argues for a "local" notion of continuity, which he dubs "closeness." Vague predicates support local continuity (a *small* change in input produces at most a *small* change in the value of the function) but *not* "global" continuity.

[10]Borderline cases result when the aggregation function fails to return a value for non-unanimous aggregations.

Figure 3: Loops on a closed disk and an annulus.

(a)   (b)   (c)
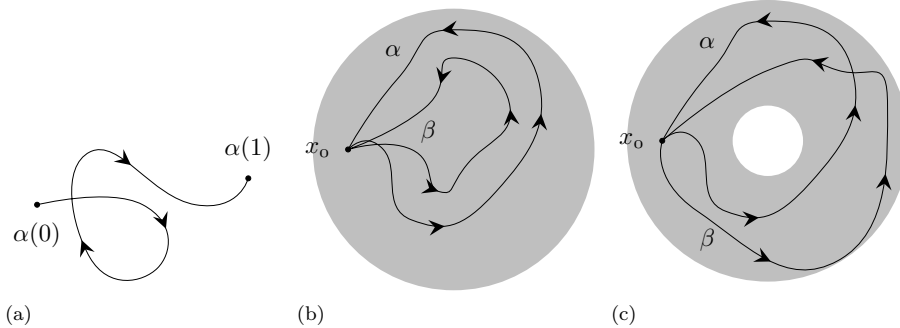
This inconsistency only arises in choice spaces with holes. To see what I mean, we need the tools of algebraic topology, which is handy for studying continuous maps between different kinds of spaces. Topology doesn't care about metrics or distances between points in a space, but rather the overall "shape" of a space. For example, a topologist does not distinguish a circle from a square because it is possible to construct a continuous bijective map (whose inverse is also continuous) from one to the other. This is a fancy way of saying that one can be stretched into the shape of the other.

But there is no way to stretch a space with a hole in it, like a circle, into a space without a hole, like a disk. Topological spaces are equivalent if it is possible to continuously deform one into another. This notion of continuous deformation goes by the name "homotopy." And there is no way to continuously deform a hole-y space into a space without holes.

To prove this, we need the notion of a path.[11] A "path" is a continuous function $\alpha : [0,1] \to X$ from the unit interval to a topological space X, with $\alpha(0)$ mapping to the initial point of the path and $\alpha(1)$ mapping to the final point (Fig. 3(a)). A "loop" is a path in which $\alpha(0) = \alpha(1)$. Usually, this point is identified at x_o, as in Fig. 3(b).

Roughly speaking, two paths are considered to be "homotopic" if they can be continuously deformed—shrunk, stretched, twisted, but not torn—into one another. The homotopy relation is an equivalence relation. It is reflexive (each path is homotopic to itself), symmetric (if α is homotopic to β, then β is homotopic to α) and transitive (if α is homotopic to β, and β is homotopic to γ, then α is homotopic to γ). We can use this relation to create equivalence classes of loops.

In Fig. 3(b), for instance, the two loops α and β are homotopic to one another—we can continuously shrink the closed disk until α and β overlap completely. Moreover, we can keep shrinking the disk until both α and β are equivalent to the "constant loop," the loop that begins and ends at x_o and goes nowhere in between (that is, the constant loop is the point x_o). And we can do this for all loops in the closed disk. Therefore, the set of equivalence classes of loops starting at x_o in the closed disk is just the set of equivalence classes of the constant loop.

However, consider the annulus in Fig. 3(c). Loops in the annulus cannot be shrunken to a point because of the hole in the middle. In this case, loops are homotopic just in case they go around the hole in the same direction the same number of times (like α and β in Fig. 3(c)). One can walk along the path α one time, or two times, or three .... And one can walk along

---

[11]I rely heavily on [18] and [2] in what follows. See these references for a more detailed account of Chichilnisky's proof.

the path α in the opposite direction once, twice, etc.

Imagine that the outer radius of the annulus is fixed, and that the inner radius increases until it equals the outer radius. Then we have a circle S^1 with the same homotopic properties as the annulus. The space of adjectival standards S^1 (or the sphere S^n) is homotopic to the annulus is Fig. 3(c).

We can now prove (a simple version of) Chichilnisky's theorem and its linguistic corollary. We associate the set of integers $\mathbb{Z}$ with the equivalence classes of loops in S^1 (once around is 1, twice around is 2, etc.) and the set of ordered pairs of integers $\mathbb{Z} \times \mathbb{Z}$ with loops in $S^1 \times S^1$. The loop in which one individual standard makes a complete rotation but a second individual standard does not is $(1,0) \in \mathbb{Z} \times \mathbb{Z}$. The opposite loop (in which the second individual standard makes a complete rotation but the first does not) is $(0,1)$. Sequential positive rotations are thus represented by $(1,0) + (0,1)$. Simultaneous positive rotations are represented by $(1,1)$.

In this integer-oriented framework, an aggregation function is a function $F_* : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$. Anonymity implies (9) because it doesn't matter which standard does the rotating (the aggregated standard depends on total rotations, not how the total is distributed among the standards).

(9) $F_*((1,0) + (0,1)) = F_*(2,0) = 2F_*(1,0)$

But $(1,0) + (0,1) = (1,1)$. And since unanimity implies that $F_*(1,1) = 1$, we have (10).

(10) $1 = F_*(1,1) = 2F_*(1,0)$

This means that there must be some integer $F_*(1,0)$ such that 2 times this integer equals 1. Impossible.

Vagueness effects like the sorites paradox are thus a product of constrained choice over spaces with holes. Evidence for this approach comes from non-vague adjectives, so-called "absolute" adjectives like *full*. Absolute adjectives do not easily give rise to the sorites paradox because the second premise is judged to be false (11).

(11)   # Premise 2. Any theater with one fewer occupied seat than a full theater is full.

These adjectives are seemingly sensitive to only one value, like volume in the case of *full* or degree of bend in the case of *straight*.

(12) $\mathbf{s}(\text{full}) = (\text{volume})$

And if the space of standards is one-dimensional, the domain of the aggregation function is more like a line than a circle. Aggregation problems do not arise; it is possible for an aggregation function to be continuous, anonymous, and unanimous [4, 347].

Chichilnisky and Heal 1983. The space $\mathcal{P}$ allows for topological aggregation if and only if each closed path is homotopic to the constant path. (In other words, there are no "holes" in $\mathcal{P}$.)

The property distinguishing absolute from adjectives like *tall* is the dimensionality of their standards. If vagueness is a problem of collective choice, this distinction explains both why vagueness effects do not generally arise in absolute adjectives and why such effects may arise if more dimensions are introduced into the aggregation process. In particular, multidimensionality creates holes.

References

[1] Kenneth J Arrow. A difficulty in the concept of social welfare. *The Journal of Political Economy*, pages 328–346, 1950.

[2] Nicholas Baigent. Topological theories of social choice. In Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura, editors, *Handbook of social choice and welfare*, volume 2, pages 301–334. Elsevier, 2010.

[3] Heather Burnett. A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy*, 37(1):1–39, 2014.

[4] G. Chichilnisky. Social aggregation rules and continuity. *The Quarterly Journal of Economics*, 97(2):337–352, 1982.

[5] Pablo Cobreros, Paul Egré, David Ripley, and Robert van Rooij. Tolerant, classical, strict. *Journal of Philosophical Logic*, 41(2):347–385, 2012.

[6] Igor Douven, Lieven Decock, Richard Dietz, and Paul Égré. Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic*, 42(1):137–160, 2013.

[7] Paul Égré, Christian List, and Franz Dietrich. Supervaluationism: a judgment-aggregation perspective. Presentation at VII Navarra Workshop on Vagueness, December 2013.

[8] Delia Graff Fara. Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, 28(1):45–82, 2002. Originally published under the name Delia Graff.

[9] K. Fine. Vagueness, truth and logic. *Synthese*, 30(3):265–300, 1975.

[10] Geoffrey Heal. Social choice and resource allocation: a topological perspective. In Geoffrey Heal, editor, *Topological social choice*, pages 1–14. Springer, 1997.

[11] Hans Kamp. Two theories about adjectives. In Edward Keenan, editor, *Formal semantics of natural language*, pages 123–155. Cambridge University Press, 1975.

[12] Hans Kamp. A theory of truth and semantic representation. In Paul Portner and Barbara H. Partee, editors, *Formal semantics: the essential readings*, pages 189–222. Blackwell, 1981.

[13] Hans Kamp and Barbara Partee. Prototype theory and compositionality. *Cognition*, 57(2):129–191, 1995.

[14] Christopher Kennedy. *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge, 1999.

[15] Christopher Kennedy. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45, 2007.

[16] Ewan Klein. A semantics for positive and comparative adjectives. *Linguistics and philosophy*, 4(1):1–45, 1980.

[17] Daniel Lassiter and Noah D Goodman. Adjectival vagueness in a bayesian model of interpretation. *Synthese*, pages 1–36, 2015.

[18] Luc Lauwers. The topological approach to the aggregation of preferences. *Social Choice and Welfare*, 33(3):449–476, 2009.

[19] Donald G. Saari. *Disposing dictators, demystifying voting paradoxes*. Cambridge University Press, 2008.

[20] Galit W. Sassoon. *Vagueness, Gradability and Typicality: The Interpretation of Adjectives and Nouns*. Brill, 2013.

[21] Nicholas JJ Smith. *Vagueness and degrees of truth*. Oxford University Press, 2008.

[22] Stephanie Solt and Nicole Gotzner. Experimenting with degree. In Anca Chereches, editor, *Semantics and Linguistic Theory*, volume 22, pages 166–187, 2012.

[23] R. van Rooij. Vagueness and linguistics. In Giuseppina Ronzitti, editor, *Vagueness: A guide*, pages 123–170. Springer, 2011.

[24] Timothy Williamson. *Vagueness*. Routledge, 1994.

Culminations and presuppositions[*]

Zsófia Gyarmathy

Heinrich-Heine-Universität Düsseldorf
zsofia.gyarmathy@gmail.com

Abstract

In this paper, I make some novel observations about presuppositions triggered by a subset of achievements, namely, culminations like *win*, and argue that culminations are not only soft presupposition triggers of the occurrence of the activity generally preceding them, *pace* the received view (cf. e.g., Abusch 2010), but instead their presupposition is of a special kind that could be called an *extra soft presupposition*, which, in contrast to soft and hard presuppositions alike, can be cancelled even when unembedded. Moreover, I argue that there are lexical differences among different presupposition triggers, even among culmination predicates themselves, with respect to the degree of the strength of their presuppositions. In line with recent ideas (e.g., Chemla, 2009; Abusch, 2010; Romoli, 2015) about the computability of presuppositions, I propose that at least the presupposition of culminations can be derived, to which end I exploit abductive reasoning.

1 Introduction: culminations as presupposition triggers

A widely accepted idea (Piñón 1997; Löbner 2002; Heyde-Zybatow 2008; Malink 2008; Abusch 2010; Martin 2011) is that some achievements *presuppose* the occurrence of an eventuality, namely, an activity. Although (as I will discuss later) this claim is assumed to hold in general for several predicates belonging to some subset of the achievements of Vendler (1957), most authors (cf. e.g., Abusch 2010; Romoli 2015) focus on the predicate *win*, which presupposes an activity of participation. That *participate* is presupposed by *win* can be verified with standard tests of projection. Negation and conditional antecedents are two of the typical *holes* for presuppositions, i.e., presuppositions are preserved in their scope (Karttunen, 1973), and this is what we observe in the case of *win*, as shown by (1) and (2) from Zinova and Filip (2014).

(1) a. John didn't win the marathon.
     b. ⇒ John participated in the marathon.

(2) a. If John won the marathon, he will celebrate tonight.
     b. ⇒ John participated in the marathon.

These tests indicate that the implication of participation is not part of the asserted content of *win*; on the other hand, we can establish that it is a presupposition rather than a *conversational implicature* by its cancellation possibilities (Beaver and Geurts, 2013). The received view is that presuppositions — more precisely, so called *soft presuppositions*, cf., e.g., Abusch (2010) — can be cancelled under negation, but not in unembedded contexts, while conversational implicatures

---

can be cancelled when unembedded. As (3) shows, *win* appears to pattern like presuppositions (though see more on this in Section 2.1[1]).

(3) a. John didn't win the race. In fact, he never participated.
     b. #John won the race. In fact, he never participated.

Thus, the received view based on considerations such as those detailed in the foregoing is that the achievement predicate *win* presupposes the occurrence of an activity of participation.

Delineating activity-presupposing achievements   Although, as noted at the outset, it is mostly only the predicate *win* that is discussed in this respect, an activity-presupposition is generally attributed to a wider range of achievements. The set of verbal predicates with an existential presupposition similar to that of *win* is characterised in the literature as "achievements with a preparatory phase" (Abusch, 2010), achievements that denote the "right boundary" of an extended event (Piñón 1997; Heyde-Zybatow 2008; Malink 2008; Martin 2011), or terms that "denote the culmination of a process" (Löbner, 2002). These appear to correspond to the class of predicates called *culminations* by Bach (1986) building on the classification of Carlson (1981), which includes predicates like *arrive, reach, die, win, find*.[2]

The activity presupposed by *find* is *look for* (cf., e.g., Martin 2006; Malink 2008), the activity presupposed by *arrive* and *reach* is moving towards the goal, and the process presupposed by *die* can perhaps be best characterised as a gradual physical event that leads to the death of the theme (e.g., physical deterioration, a gradually decreasing supply in oxygen etc.). However, of these predicates, I know of only *find* that was studied besides *win* as a presupposition trigger (e.g., by Martin 2006, Malink 2008 and Piñón 2008). The tenability of the general claim that culminations (using henceforth Bach's terminology) presuppose the occurrence of an activity preceding them (let us call this the *activity-implication*) is therefore in need of corroboration. Piñón (1997) offers a justification for this general claim, and argues that it is rooted in an ontological presupposition: in particular, culminations describe events that are the right boundary of an extended event and so they require the existence of an extended event of which they are the boundary of. However, as I will argue in Section 2.1, such a view of the activity-implication, which obviously does not allow for exceptions to the rule, runs into problems in the face of culmination predicates being instantiated without a corresponding preceding activity.

2   Culminations as extra soft presupposition triggers

2.1   Cancellability of the activity-implication

An observation I make in Gyarmathy (2015a,b) about presuppositions triggered by culminations is that they raise a worry that has not been noted and addressed so far, namely, that there may be cases where such culminations are true of an event *without* the occurrence of an activity preceding them. Some examples are as follows:

(4) a. The right wing **won** this contest without even participating in it.[3]

---

[1]Native speakers have voiced their concern that (3b) might, in fact, be assertable in a scenario in which John won the race without participation by virtue of the fact that everyone else was disqualified. This and similar considerations will be grounds to argue that culminations are what I will call extra soft presupposition triggers.

[2]Note that *find* is classified as a right-boundary achievement (in Bach's terminology, a culmination) by Malink (2008) and Piñón (2008).

[3]http://links.org.au/node/1210

b. I thought the whole deal of using 115 was that you can curve space in such a way that you "pull" your destination towards you and restore space once you are at the apex of the curvature, having effectively **arrived at your destination** without moving.[4]

c. Mary **found** a penny quite by accident without looking for one. *(based on Martin, 2006)*

d. The spy **died** by shooting herself, and so she wasn't dying before she died. *(based on Piñón, 1997)*

Note that it is a well-known fact that *soft presuppositions*, such as the factive implication of *discover*, are weak and context-dependent and may be obviated. Abusch (2010) includes achievements with a preparatory phase among soft presupposition triggers, along with, e.g., verbs of accompanied motion like *accompany*. Although soft presuppositions can be cancelled, their cancellation is invariably restricted to embedded contexts: e.g., within the scope of negation, a question or the antecedent of a conditional. But what we see in the case of culminations is that their presupposition can be obviated even in unembedded contexts, that is, the sentences in (4) are truthfully assertable.

While some of the examples in (4) (those for *win* and *arrive*) may sound convoluted and involve special cases,[5] there is a distinct contrast with at least some other (soft) presupposition triggers, whose presupposition cannot be cancelled in unembedded contexts, however convoluted the setting is. In fact, such triggers include the same culmination predicates themselves, but the presupposition under scrutiny being the implication to the occurrence of an inverse state as their result state, cf. (5a).

(5) a. #Dorothy **arrived** at the station, even though she was already at the station.
     b. #John **accompanied** Jane across the bridge, even though Jane wasn't crossing the bridge.
     c. #John **stopped smoking**, even though he never smoked.

Given their ability to be cancelled in unembedded contexts, I call the activity-implication of culminations an *extra soft presupposition*, which can be cancelled even when unembedded. Thus, we can extend the hierarchy of presupposition softness as follows:

- **hard presuppositions** cannot be cancelled;

- **soft presuppositions** can be cancelled in embedded contexts;

- **extra soft presuppositions** can be cancelled.

In Section 4, I will, in addition, argue that extra soft presuppositions are also different from conversational implicatures despite their similar cancellation patterns.

---

[4]`http://www.abovetopsecret.com/forum/thread215509/pg1&mem=`

[5]Some further examples from the Internet for *win* are as follows (my emphasis):

(i) as far as i'm concerned sam pretty much **won** every single year without even participating. (*on winners of World's Ugliest Dog title in the past decade and the winner of years 2003-2005*)

(ii) That's what validates the supporters right to say that "**we won the game**" even though they didn't participate in the game it self.

Some further examples from the Internet for *arrive* (my emphasis):

(iii) – at an hour which would make us fashionably late had we **arrived** at that very instant by teleportation –

(iv) She **arrived** instantaneously, almost as if she had been transferred by wire.

2.2   Culminations and degrees of presupposition robustness

Another observation we can make about the activity-implications of culminations based on examples like those in (4) is that the *strength* of this implication differs by predicate. Predicates *win*, *arrive* and *reach* appear to fairly strongly imply the occurrence of a preceding activity: as mentioned above, these predicates necessitate extremely convoluted uses to exemplify presupposition cancellation in unembedded contexts. In contrast, *find* — even though it is sometimes cited (e.g., in Malink, 2008) as presupposing an activity of looking for the Theme — has a presupposition that is more easily cancelled (it is quite possible to find something by accident). As for *die*, even if it does presuppose a process of the condition of the experiencer deteriorating, that is also very easily cancelled (one can die instantaneously, e.g., in an accident). This observation about presuppositions, interestingly, mirrors a recent observation by van Tiel et al. (2015) about *scalar implicatures*: they noted that there are various degrees of the rate at which scalar inferences are drawn from different scalar expressions.

It would be a welcome feature of an account of activity-implications (or of presuppositions, or of implicative contents in general) if it could provide an explanation of why such lexical differences in the degree of presupposition robustness (over and above the well-established soft/ hard, or the proposed extra soft/soft/hard distinction) occur.

3   Former accounts of the activity-implication

Lexical specification   As regards the source of the activity-implication, Piñón (1997, 2008) and Malink (2008) assume that culminations (in their terminology, right-boundary achievements) lexically presuppose a preceding activity, which can be encoded, for instance, with axioms of the kind "For each finding event, there is a searching event of which the finding event is a right boundary" (Piñón, 2008):[6]

$$\forall e \forall t_r [(\mathsf{find}(e) \,\wedge\, \tau(e) \,\sqsubseteq\, t_r) \;\rightarrow\; \exists e'[\mathsf{search}(e') \,\wedge\, t_r \,\sqsubseteq\, \tau(e') \,\wedge\, \mathsf{right\_boundary}(e, e')]] \quad (1)$$

Piñón (1997) argues that the source of this lexical presupposition is an *ontological* presupposition, namely, that a right boundary presupposes the occurrence of something it is the right boundary of.

However, this proposal is obviously not compatible with the observations above about the cancellability of the activity-implication. For instance, not all findings are preceded by searchings. Piñón (2008, p. 164) is, in fact, aware of this problem (and actually appears to be one of only few authors to acknowledge this problem), and proposes in a footnote that (the German counterpart of) the word *find* is ambiguous, and that "[t]here is another sense of *finden* for 'accidental findings' that does not presuppose a searching activity per se [...] [h]owever, for expediency I set aside this use of *finden* here." I am not sure whether positing such a lexical ambiguity in the case of *find* (and *all* culminations, in general) is warranted. Indeed, applying some classical tests of ambiguity from Zwicky and Sadock (1975) suggests that *find* is a single lexical entry.[7]

---

[6]Piñón (2008) uses the variable t_r in the axiom below to quantify over reference times; τ is the temporal trace function from events to their runtime, as usual.

[7]While (ia) below from Zwicky and Sadock (1975) is unproblematic (*dog* is ambiguous between a male dog and the dog species), (ib) appears contradictory. Also, while (iia) from Zwicky and Sadock (1975) only has two readings (both Morton and Oliver threw their lunch to the floor, or both of them ate it) instead of four (i.e., it excludes "crossed understandings"), (iib) does not seem problematic, that is, a "crossed understanding" interpretation (finding after searching and finding by accident) is possible, indicating a lack of ambiguity.

A way to reconcile the lexical axiom of Piñón (2008) in (1) with unpaired culminations (findings, in particular) is to use a generic quantifier instead of a universal one:

$$\text{GEN}[e]\big(\forall t_r \big[(\text{find}(e) \wedge \tau(e) \sqsubseteq t_r) \rightarrow \exists e'[\text{search}(e') \wedge t_r \sqsubseteq \tau(e') \wedge \text{right\_boundary}(e, e')]\big]\big) \quad (2)$$

However, it seems to me that such an axiom still raises some questions. In particular, the precise interpretation of GEN is notoriously elusive; but even given one specific interpretation of GEN, how can such axioms explain the differences we find between different lexical items in the strength of their activity-implication (cf. Section 2.2)? More importantly, though, where does such an axiom come from? We cannot assume with Piñón (1997) anymore that it is an ontological presupposition (i.e., that findings are boundary events and they require something to be a boundary of), because then why are findings without any searching events allowed? My proposal in Section 5 can be seen as an attempt at uncovering the source of such axioms, while allowing for the possibility that different culminations imply the occurrence of an activity with different strength.

Alternatives-based accounts Abusch (2010) and Romoli (2015), in contrast to the purely semantic presupposition account of Piñón and Malink, propose to derive the presupposition of culminations (among others) in analogy to scalar implicatures, arguing that they are associated with lexically specified alternatives: e.g., for *win*, {*win, lose*} (Abusch, 2010) or {*win, participate*} (Romoli, 2015). These alternatives then form the basis of pragmatic reasoning similar to what is traditionally assumed in the case of scalar implicatures.

The question that arises is again what the source of these lexically encoded alternatives is. Why is it exactly these predicates that are specified with alternatives (and not predicates like *explode* that Bach 1986 calls *happenings*, which do not imply the prior occurrence of an activity), and why are these the alternatives that are encoded? And more pertinently to the goals of this paper, how are extra soft presuppositions differentiated from merely soft presuppositions on these accounts? And why do we find different degrees of implication strength?

I do not wish to suggest that these questions cannot be answered in an alternatives-based approach to presuppositions; in fact, differences between triggers in the robustness of their implicative content must have some explanation in such frameworks, given that (as noted above) van Tiel et al. (2015) observed exactly such graded differences in the case of *scalar triggers*, whose semantics is almost invariably assumed to exploit alternatives. I will, however, propose an alternative analysis of activity-implications in Section 5, which might in the long run be actually combined with alternatives-based or axiom-based approaches.

4   Presuppositions versus scalar implicatures

The question arises, though, if we should perhaps conclude based on the possibility of cancellation in a non-embedded context that the activity-implication is a conversational implicature rather than a presupposition (with the observed graded differences in the strength of implication fitting in straightforwardly with the observation of van Tiel et al. 2015 about scalar triggers).

(i)   a.   That dog isn't a dog: it's a bitch.
       b.   ?John found this pen, but didn't *find* it: he wasn't looking for it.

(ii)   a.   Morton and Oliver tossed down their lunches.
       b.   Morton and Oliver found a penny each. Morton was looking for one, but Oliver just found one by accident.

I maintain that that is not the correct conclusion, as, contrary to conversational implicatures, the implication to the occurrence of the activity part does not seem to be calculable solely from the meaning of the culmination and general principles of *cooperative conversation*, but seems to involve a conventional aspect (even the alternatives-based approaches mentioned above use *lexical* alternatives to derive presuppositions, but not scalar implicatures).

In addition, the experimental work by Bill et al. (2014) also indicates that the activity-implication of *win* is different from *scalar implicatures*: they found that the inference from a sentence like *The bear didn't win the race* to the bear participating in the race is much more easily cancelled than the inference from a sentence like *Some of the lions have balloons* to not all of the lions having balloons. Of course, further experimental work is necessary to show that the existential presupposition of *win* and other culminations is different from all other kinds of conversational implicatures besides scalar implicatures. But presuppositions (that of culminations among them) have only been analysed in analogy to scalar implicatures out of conversational implicatures so far, and as for the maxims of Relevance and Manner, it is difficult to see what role they could play in the inference from, say, *win* to *participate*.

Further motivation for regarding the activity-implication of culminations like *win* as a presupposition comes from an observation I put forth in Gyarmathy (2015b), namely that the conjunction of the implied and asserted content (in this order) with *and (in fact)* is admissible for presuppositions, be they hard (as in (6)), soft (as in (7)), or extra soft (as in (8)), but odd for scalar implicatures (as in (9)).

(6)   Jane came and (in fact) John came, **too**.

(7)   a.   Jane walked across the bridge and (in fact) John **accompanied** Jane.
      b.   John was smoking and (in fact) he then **stopped smoking**.

(8)   a.   John participated and (in fact) **won**.
      b.   John walked toward the station and (in fact) **arrived** at the station.

(9)??John didn't eat all of the cookies and (in fact) he ate **some** of the cookies.

Theoretical support for this difference in behaviour comes from the idea put forth by Chemla (2009) and Romoli (2015) that soft presupposition triggers are *strong scalar items* — as opposed to scalar implicature triggers, which occupy an opposite position on Horn-scales, i.e., they are weak scalar items. For instance, according to Romoli (2015), the alternatives associated with *win* are {*win, participate*}, where *win* entails *participate*.[8] In contrast, for the alternatives *some* and *all*, it is *all* that entails *some* (and so due to the Gricean maxims of conversation, *some* introduces the scalar implicature *not all*). Thus, only in the case of presuppositions is it the case that the implied content is *weaker* than the asserted content, and it is pragmatically acceptable to add a stronger piece of information to a weaker one. Note that (as pointed by an anonymous reviewer,[9] as well as Todor Koev, p.c.) (9) can be rendered felicitous by using *but* instead of *and*. This in fact supports the foregoing argumentation, as the difference between *and* and *but* is a pragmatic one, with *but* being often used exactly in cases where the use of

---

[8]As I argue (cf. Section 2.1), this is no strict logical entailment relation, but only a cancellable implication.
[9]The reviewer also suggested that the infelicity of sentence (9) "seems caused by the presence of negation in the first sentence (e.g., *I didn't play tennis and I played football* also sounds odd)". However, this is not the case, as witnessed by (ia) (note that *some* is also a scalar implicature of *not all*) which does not contain negation in the first sentence, and by (ib) showing that negation in the first sentence does not rule out acceptability:

(i)   a.   ??John ate **some** of the cookies and (in fact) he didn't eat all of the cookies.
      b.   I don't play tennis, and (in fact) I only do combat sports.

and would be pragmatically odd (cf., e.g., ?*John is a criminal and an honest man* versus *John is a criminal but an honest man*).

While I do not believe that the activity-implication is a scalar implicature (or any other kind of conversational implicature), I do agree with authors like Chemla (2009) and Abusch (2010) that at least some presuppositions can be *calculated*, similarly as with (though not in the exact same way as) implicatures, and this is what I aim to do for culminations, building on a special kind of inference that is used widely in human reasoning.

5   The source and cancellability of the activity-implication

5.1   Abductive reasoning

In order to derive the activity presupposition of culminations, I propose to exploit *abduction*, i.e., the inference to the best explanation introduced by Charles Sanders Peirce, which is (contrary to deductive reasoning) defeasible. Abductive reasoning has come to be widely employed in AI (cf., e.g., Hobbs et al. 1993; for an overview, see, e.g., McIlraith 1998), and it is also abundantly used in everyday reasoning (cf. Douven, 2011). I propose that abduction is ideally suited for accounting for at least some defeasible inferences—including the activity-implication. Curiously, abduction has not yet been exploited much in formal semantics and pragmatics. An exception is Piñón (2009, 2011), who uses abduction to derive the actuality entailments of ability modals,[10] but the present proposal builds most closely on Varasdi (2010, 2014), who offers an analysis of the imperfective paradox via an exploitation of an inverse reasoning that is basically at the core of abduction (even though he does not use the word "abduction").

Abduction involves: (i) something that is *observed* and is to be explained, (ii) a *theory* (which encodes the non-defeasible rules of reasoning), and (iii) the *explanation* abduced on the basis of (i) and (ii), which together with the theory entails the observation.[11] The abduced explanation needs to be the "best" one among possible explanations, where what counts as "best" depends on the concrete framework of abduction. An explanation is generally regarded better than another one if its simpler (e.g., in a logical representation language, consists of less literals), and/or more probable etc.

Abductive inferences often involve an inference to the antecedent of a conditional on observing the consequent. Thus, if we observe q, and our theory tells us that $p \rightarrow q$, then we abduce p, because together with the theory, this entails what we observe, and is definitely at least among the simplest explanations.

5.2   Culminations and abduction

In deriving the presupposition of culminations, I draw on the exceptionless rule that *whenever* an extended event satisfying certain conditions occurs (for winning, participating and coming in first place, for arriving, movement to the goal), then its culmination can be described with the relevant culmination predicate *without exception*. So we have the (deductive) inference, "if P (an extended event) occurs, then Q (a culmination) occurs". Then, what I propose happens, is an *abductive step*, deriving an inference to the best explanation (which is sufficient, but not necessary for what is observed). So by abduction we get "if Q (a culmination) occurs, then P (an extended event) occurs."

---

[10]Thanks to Fabienne Martin for pointing this out to me.

[11]There are less demanding approaches to abduction, as well, in which case simple consistency with the theory suffices (McIlraith, 1998).

Via this inference, properties common to the (*non-final* parts of) events in P are those presupposed by the culmination denoting Q (e.g., for *win*, *participate*). Because it is not a deductive step, there may be cases where the reasoning fails: finding without searching, winning without participating etc. If we assume that, e.g., there are relatively speaking more findings without searchings than winnings without participatings, and that abduction is sensitive to such ratios (though need not rely solely on these), then we can now also explain lexical differences between culminations relating to the strength of their presupposition given simply differences between the ratio of "unpaired" events belonging to different culmination predicates.

A possible objection to this analysis is that there might be a danger of overgeneralisation, because abduction is an extremely powerful tool. An anonymous reviewer made such an objection: "I'm wondering whether this explanation might not overgeneralise: a situation in which, e.g., some but not all of the apples are red can be described with *some* so if *some* is used listeners may abductively infer that a situation in which all of the apples are red is excluded." But in fact, we do want the defeasible inference from *some* to *not all* to go through (even though it will not go through based on the reasoning put forth by the reviewer, as we do not reason based solely on a single observation), as it *is* a defeasible inference, and defeasibility is exactly the hallmark of abductive inferences. And the standard Gricean maxims-based explanation might be seen as a "meta-level" abduction, which involves not only observations and theories about facts of the world, but also ones about *utterances* about facts of the world: the best explanation for why the speaker used *some* is that it is not true that all apples are red, so this observation (speaker used *some*, but speaker didn't use *all*) can be explained with the theory (Maxim of Quality and Maxim of Quantity) and the abduced explanation (not all apples are red).

What may at first sight seem like a case of overgeneralisation, however, is the following: suppose that, by vast majority, whenever someone eats some apples, they eat all (of a bunch). In this case, from the observation that x ate some apples, we would abduce that they ate all, quite contrary to the scalar implicature of *some*. However, note that this inference would be based on reasoning solely about facts of the world. Taking into account additional facts, in particular that someone uttered *x ate some apples*, and extending our theory with the cooperative principles of conversation, we would now cancel our original inference (that x ate every apple) and abduce its opposite, that x in fact did *not* eat every apple. This (as Károly Varasdi, p.c., noted) highlights the importance of the *non-monotonicity* of abductive reasoning, that is, that abductive inferences are defeasible when information is added.

In contrast to the case of *some* and *(not) all*, the abductive inference about culminations presented above based on the facts of the world is not retracted when taking into account "metal-level" information (about utterances). This is because, as noted by Chemla (2009) and Romoli (2015), culminations as presupposition triggers occupy a different place among the set of alternatives as scalar triggers, and are *strong* scalar items. This also means, however, that their activity-implication is a weak scalar item, and so at the "metal-level" we abduce *not win* from *participate*, similar to abducing *not all* from *some*: if someone utters *John participated in the race yesterday*, we at least weakly infer that John did not win.

So while the danger of overgeneralisation is a valid worry about abduction in general, an abductive system does have predictive power. And it can also arguably cater for the observed differences between the strengths of the implications of triggers. For instance, in our case, if *find* does presuppose *search*, then *win* must also presuppose *participate*, because our model contains much less cases of winning without participating than findings without searching, so if there is enough grounds for abducing *search* from *find*, then the same reasoning must also make us abduce *participate* from *win*. And, by contrast, we do not predict happenings like *explode* to have an activity-implication, as there is no sufficient grounds to abduce one.

5.3 Presupposition projection and at-issueness

The abductive inferencing process only explains the defeasible activity-implication of a sentence asserting the occurrence of a culmination-event. However, once we have this inference from the positive form, we can explain why it *projects* (from under typical holes like negation, the antecedent of a conditional, etc.) drawing on the idea going back to at least Stalnaker (1974) that presuppositions are inferences that are not about the main point (Martin, 2006; Simons et al., 2010; Abrusán, 2011, a.o.). Simons et al. (2010) claim that "all and only those implications of (embedded) sentences which are not-at-issue relative to the Question Under Discussion [QUD] in the context have the potential to project."

   We can argue, as suggested by Fabienne Martin (2006 and p.c.), that the activity-implication of a culmination (e.g, participation) is not at-issue relative to a QUD relative to which the asserted content of the culmination (the occurrence of the culmination, e.g., winning) is at issue. If we can prove this to be true,[12] then we can argue that the activity-implication projects by virtue of it not being at-issue content.

6   Conclusion

Based on an inspection of the presupposition of culmination predicates, I have argued that i) there are so called extra soft presuppositions, which are unlike soft presuppositions in that they can be cancelled when unembedded, ii) there are lexical differences between presupposition triggers as regards the strength of their presupposition, mirroring similar recent observations on scalar expressions, and iii) by drawing on abduction, we can explain the activity-implication of culminations along with its (different degrees of) cancellability.

References

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy 34*(6), 491–535.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics 27*(1), 37–80.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy 9*, 5–16.

Beaver, D. and B. Geurts (2013). Presupposition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 ed.).

Bill, C., J. Romoli, F. Schwarz, and S. Crain (2014). Indirect scalar implicatures are neither scalar implicatures nor presuppositions (or both). Poster presented at the *27th Annual CUNY Conference on Human Sentence Processing*, Columbus, USA.

Carlson, L. (1981). Aspect and quantification. In P. Tedeschi and A. Zaenen (Eds.), *Tense and Aspect*, Volume 14 of *Syntax and Semantics*, pp. 31–64. New York: Academic Press.

Chemla, E. (2009). Presuppositions of quantified sentences: experimental data. *Natural Language Semantics 17*(4), 299–340.

Douven, I. (2011). Abduction. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011 ed.).

Gyarmathy, Zs. (2015a). *Achievements, durativity and scales*. Logos Verlag Berlin. Published version of the dissertation defended at Heinrich-Heine-Universität Düsseldorf.

---

[12]At-issueness is notoriously difficult to verify, but, e.g., an experiment involving question–answer pairs could be used to verify the claim.

Gyarmathy, Zs. (2015b). Culminations as extra soft presupposition triggers. Talk presented at *Experimental and crosslinguistic evidence for the distinction between implicatures and presuppositions (ImPres)*, Berlin.

Heyde-Zybatow, T. (2008). What kind of events do achievements describe? In A. Steube (Ed.), *The Discourse Potential of Underspecified Structures*, Volume 8 of *Language, Context and Cognition*, pp. 109–144. Berlin: De Gruyter.

Hobbs, J. R., M. E. Stickel, D. E. Appelt, and P. Martin (1993). Interpretation as abduction. *Artificial Intelligence 63*(1-2), 69–142.

Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry 4*(2), 169–193.

Löbner, S. (2002). *Understanding Semantics*. Understanding Language. London: Hodder Education.

Malink, M. (2008). Right boundary achievements under conative negation. In A. Steube (Ed.), *The Discourse Potential of Underspecified Structures*, Volume 8 of *Language, Context and Cognition*, pp. 145–162. Berlin: De Gruyter.

Martin, F. (2006). *Prédicats statifs, causatifs et résultatifs en discours. Sémantique des adjectifs évaluatifs et des prédicats psychologiques*. Ph. D. thesis, Université libre de Bruxelles.

Martin, F. (2011). Revisiting the distinction between accomplishments and achievements. In W. de Mulder, J. Mortelmans, and T. Mortelmans (Eds.), *From Now to Eternity*, Volume 22 of *Cahiers Chronos*, Amsterdam/New York, pp. 43–64. Rodopi.

McIlraith, S. A. (1998, July). Logic-based abductive inference. Technical Report KSL-98-19.

Piñón, C. (1997). Achievements in an event semantics. In *Proceedings of SALT VII*, pp. 273–296.

Piñón, C. (2008). Negating right boundary achievements (comments on Malink). In A. Steube (Ed.), *The Discourse Potential of Underspecified Structures*, Volume 8 of *Language, Context and Cognition*, pp. 163–175. Berlin: De Gruyter.

Piñón, C. (2009). Another look at the actuality entailment of certain modal verbs. Talk presented at the *Genericity: Interpretation and Uses*, Paris.

Piñón, C. (2011). The pragmatics of actuality entailments. Talk presented at the *Aspect and Modality in Lexical Semantics*, Stuttgart.

Romoli, J. (2015). The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of Semantics 32*(2), 173–219.

Simons, M., J. Tonhauser, D. Beaver, and C. Roberts (2010). What projects and why. In *Proceedings of SALT 20*, pp. 309–327.

Stalnaker, R. C. (1974). Pragmatic presuppositions. In M. K. Munitz and P. K. Unger (Eds.), *Semantics and Philosophy*, pp. 197–213. New York University Press.

van Tiel, B., E. van Miltenburg, N. Zevakhina, and B. Geurts (2015). Scalar diversity. *Journal of Semantics Advance access*, 1–39.

Varasdi, K. (2010). A higher order extensional framework for the progressive. *Journal of Semantics 27*(2), 177–206.

Varasdi, K. (2014). Making progressives: Necessary conditions are sufficient. *Journal of Semantics 31*(2), 179–207.

Vendler, Z. (1957). Verbs and times. *The philosophical review 66*(2), 143–160.

Zinova, Y. and H. Filip (2014). Meaning components in the constitution of Russian verbs: Presuppositions or implicatures? In *Semantics and Linguistic Theory (SALT) 24*, pp. 353–372. New York: New York University.

Zwicky, A. and J. M. Sadock (1975). Ambiguity tests and how to fail them. In *Syntax and Semantics*, Volume 4, pp. 1–36. New York: Academic Press.

Alternatives in Cantonese: Disjunctions, Questions and (Un)conditionals

Yurie Hara[*]

Department of Linguistics and Trnalsation
City University of Hong Kong
`y.hara@cityu.edu.hk`

1   Introduction

Cantonese has at least four lexical entries that can be translated as English 'or', *waak6ze2*, *ding6(hai6) jat1hai6* and *jik1waak6*.[1] This paper focuses on the first two, *waak6ze2* and *ding6*. Pedagogically, *waak6ze2* and *ding6* are described as 'or' in a statement and 'or' in a question, respectively.

The goal of this paper is to describe the properties of the two disjunctions and provide a compositional analysis that explains their distributions and interpretations in the framework of Suppositional Inquisitive Semantics (InqS; Groenendijk & Roelofsen, 2014).[2] Specifically, both *ding6* and *waak6ze2* denote an inquisitive disjunction which forms a union of a set of propositions, but they differ in that *ding6* has an extra syntactic requirement that the clause containing *ding6* remains inquisitive. Consequently, '*p* waak6ze2 *q*' is a non-inquisitive sentence, while '*p* ding6 *q*' is always inquisitive. Furthermore, the analysis correctly derives the connotations which arise from unconditional sentences.

2   Empirical Data

According to Haspelmath (2007), some languages have two kinds of disjunction, "interrogative disjunction and standard disjunction" (pp. 26–27). For instance, Mandarin Chinese (Li & Thompson, 1981; Yuan, 2015) distinguishes interrogative *háishe* and standard/declarative *huòze*. In Egyptian Arabic (Winans, 2013), a sentence contains *wallaa* is interpreted as an alternative question as it cannot be responded with 'yes' or 'no', while *aw* is a standard disjunction since in a question it is understood as a *yes/no* question.

Cantonese has a comparable pair of lexical entries, *ding6* and *waak6ze2* both of which translate to 'or' in English.[3] The grammaticality judgments of the following examples are given based on the interviews with the consultants and the results of the naturalness rating studies summarized in Figures 1 and 2 taken from Hara (2015).

[1]Numbers indicate lexical tones. 1=high level or high falling; 2= mid rising; 3= mid level; 4= low falling; 5= low rising; 6= low level. *Hai6* in *ding6hai6* can be omitted in casual speech and omitted hereafter.

[2]Winans (2013) also provides Inquisitive Semantics analysis to two kinds of disjunction in Egyptian Arabic.

[3]Despite their similarities, the distribution pattern of Cantonese *ding6* and *waak6ze2* is different from those of Mandarin *háishe* and *huòze* and Egyptian Arabic *wallaa* and *aw*. For instance, Mandarin *háishe* and *huòze* are interchangeable under conditional antecedents and modals as reported in Huang (2010). Egyptian Arabic *wallaa* and *aw* are in complementary distribution while Cantonese *ding6* and *waak6ze2* are not. See Hara (2015) for comparison.

Figure 1 shows that in declarative constructions including embedding under modals and conditional antecedents, *ding6* is ungrammatical while *waak6ze2* is grammatical:

(1) Declaratives
Lisa sik6 zuk1   *ding6/waak6ze2   faan6
Lisa eat congee DING6/WAAK6ZE2 rice
'Lisa eats congee or rice'

(2) Modals
Lisa ho2ji5/jiu3/ho2nang4/jat1ding6-hai6 sik6 zuk1   *ding6/waak6ze2   faan6
Lisa can/must/possibly/definitely-is   eat   congee DING6/WAAK6ZE2 rice
'Lisa can$_{\text{deontic}}$/must$_{\text{deontic}}$/can$_{\text{epistemic}}$/must$_{\text{epistemic}}$ eat congee or rice'

(3) Conditional antecedent
jy4gwo2 Lisa sik6 zuk1   *ding6/waak6ze2   faan6, ceng2 waa6 ngo5 zi1.
if   Lisa eat congee DING6/WAAK6ZE2 rice,   please speak me   know
'If Lisa eats congee or rice, please let me know.'

In unconditional antecedents, both *ding6* and *waak6ze2* are grammatical, as can be seen in Figure 1.

(4) Unconditional antecedent
mou4leon6 Lisa sik6 zuk1   ding6/waak6ze2   faan6, koei5 dou1 wui5 baau2
no.matter Lisa eat congee DING6/WAAK6ZE2 rice,   she   will be   full
'Whether Lisa eats congee or rice, she will be full.'

Figure 2 shows that interrogatives with *ding6* end with a particle *aa4* and are straightforward alternative questions since they have to be answered by one of the choices. Answering 'yes' or 'no' makes the discourse anomalous.

(5) Lisa jiu3 zuk1   ding6 faan6 aa3?
Lisa want congee DING6 rice   PRT
'Does Lisa want congee or rice?'
a. *jiu3 (want) 'Yes, she wants'/*m4 jiu3 (not want) 'No, she doesn't want.'
b. zuk1 (congee) 'Congee.'/faan6 (rice) 'Rice.'

On the other hand, interrogatives with *waak6ze2* are more complicated and need a more careful observation. Interrogatives with *waak6ze2* end with a particle with a different tone, *aa4* and based on the consultants' introspection-based judgments, they should be *yes-no* questions rather than alternative questions. However, the experimental result shows that in fact, none of the answers to *waak6ze2*-questions are perceived as natural as the ones to *ding6*-questions. Furthermore, just like *ding6*-questions, answering just 'yes' or 'no' to *waak6ze2*-questions is judged quite unnatural. Perhaps, this is because the speaker, being maximally cooperative, should provide the actual choice after answering, e.g., 'Yes, I want congee.' Thus, it is judged unnatural due to its uncooperativeness.

(6) Lisa jiu3 zuk1   waak6ze2 faan6 aa4?
Lisa want congee WAAK6ZE2 rice   PRT
'Does Lisa want congee or rice?'
a.???jiu3 (want) 'Yes, she wants'/??m4 jiu3 (not want) 'No, she doesn't want.'
b. ?zuk1 (congee) 'Congee.'/?faan6 (rice) 'Rice.'
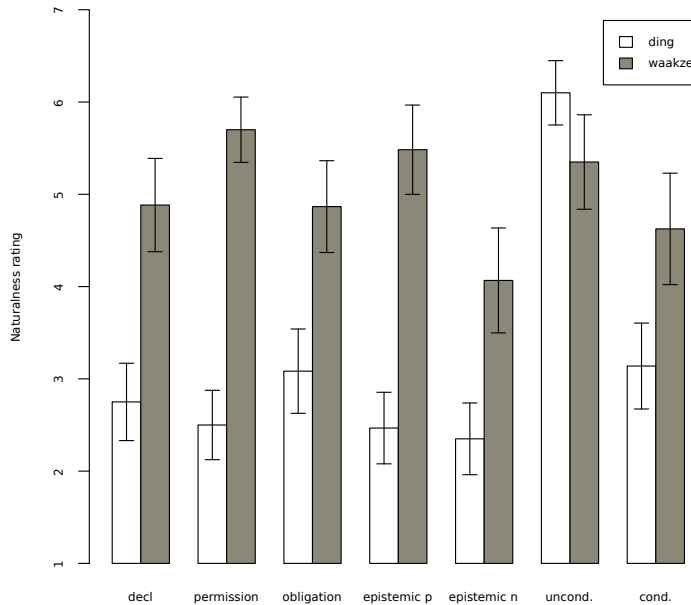
Figure 1: Average naturalness of the constructions (Hara, 2015, 18)

Another possibility is that *waak6ze2*-questions are what Roelofsen & van Gool (2010) call "open intonation" questions. According to Roelofsen & van Gool (2010), when an English disjunctive interrogative has open intonation as indicated in (7), a 'yes' answer is not licensed.

(7) Does <u>Ann</u>↑ or <u>Bill</u>↑ play?
     ✓No. #Yes. ✓Ann does. ✓Bill does.   (Roelofsen & van Gool, 2010)

As for Cantonese *waak6ze2*-questions, indeed, as reported in Hara (2015), the average of 'no'-answers is significantly higher than 'yes'-answers ($t = 3.884, p < 0.001$). Thus, I conclude that *waak6ze2*-questions are interpreted as *yes/no* questions or open questions. The empirical characterization of the distribution of *ding6* and *waak6ze2* is summarized in the following table:

(8)

	declarative	modal	unconditional	conditional	question
ding6	*	*	✓	*	alternative
waak6ze2	✓	✓	✓	✓	yes/no, open

Thus, *waak6ze2* and *ding6* are not in complementary distribution in a strict sense. In particular, *waak6ze2* in an unconditional is also judged quite natural. Thus, Cantonese *waak6ze2* is more like English *or*, which can be treated either as a standard disjunction or an interrogative disjunction in Haspelmath's (2007) terms, while *ding6* exclusively denotes the interrogative
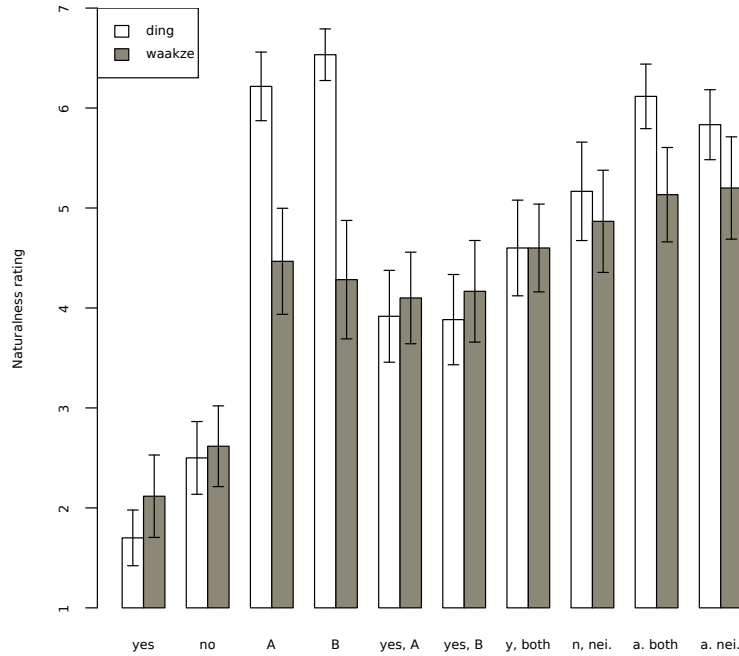
Figure 2: Average naturalness: Question-Answer Pair (Hara, 2015, 16)

disjunction.

3 Proposal

Based on the empirical data obtained in Section 2, I propose that both *ding6* and *waak6ze2* denote an (inquisitive) disjunction which forms a union of set of propositions. The difference between the two items lies in their syntactic specifications. While *α-ding6-β* carries an uninterpretable feature which forces the clause to be inquisitive and resist the declarative operator, *α-waak6ze2-β* lacks such a feature, hence its denotation can be non-inquisitive.

3.1 Suppositional Inquisitive Semantics

Suppose that a possible world is a valuation for atomic sentences and W is the set of all possible worlds. In the standard possible-world semantics, the meaning of a sentence is a set of possible worlds. Thus, $[\![\text{Lisa smiles}]\!] = |\text{smile}(\text{lisa})| = \{w|\ \text{Lisa smiles in } w\}$. In Inquisitive Semantics (Groenendijk & Roelofsen, 2014), possible worlds constitute an information state, σ,

and the meaning of a sentence φ is a set of information states that support φ. In Suppositional Inquisitive Semantics (InqS), there are two more semantic relations besides 'support':[4]

(9)   Atomic sentences

    a. σ *supports* p $(\sigma \vDash^+ p)$ iff $\sigma \neq \emptyset$ and $\forall w \in \sigma.w(p) = 1$.

    b. σ *rejects* p $(\sigma \vDash^- p)$ iff $\sigma \neq \emptyset$ and $\forall w \in \sigma.w(p) = 0$.

    c. σ *dismisses a supposition of* p $(\sigma \vDash^\circ p)$ iff $\sigma = \emptyset$.

A state σ supports $\neg\varphi$ just in case σ rejects φ, and σ supports $\varphi \vee \psi$ just in case σ supports φ or σ supports ψ:

(10)   a. $\sigma \vDash^+ \neg\varphi$ iff $\sigma \vDash^- \varphi$.   (11)   a. $\sigma \vDash^+ \varphi \vee \psi$ iff $\sigma \vDash^+ \varphi$ or $\sigma \vDash^+ \psi$.

        b. $\sigma \vDash^- \neg\varphi$ iff $\sigma \vDash^+ \varphi$.   b. $\sigma \vDash^- \varphi \vee \psi$ iff $\sigma \vDash^- \varphi$ and $\sigma \vDash^- \psi$.

        c. $\sigma \vDash^\circ \neg\varphi$ iff $\sigma \vDash^\circ \varphi$.   c. $\sigma \vDash^\circ \varphi \vee \psi$ iff $\sigma \vDash^\circ \varphi$ or $\sigma \vDash^\circ \psi$.

The semantics for implication is defined as follows:

(12)   a. $\sigma \vDash^+ \varphi \to \psi$ iff $\sigma \cap info(\varphi) \vDash^+ \varphi$ and $\sigma \cap info(\varphi) \vDash^+ \psi$.

        b. $\sigma \vDash^- \varphi \to \psi$ iff $\sigma \cap info(\varphi) \vDash^+ \varphi$ and $\sigma \cap info(\varphi) \vDash^- \psi$.

        c. $\sigma \vDash^\circ \varphi \to \psi$ iff $\sigma \cap info(\varphi) \nvDash^+ \varphi$ and $\sigma \cap info(\varphi) \vDash^\circ \psi$.

The set of all states that support φ, $[\varphi]^+$ is defined as in (13).

(13) $[\varphi]^+ := \{\sigma | \sigma \vDash^+ \varphi\}$

In InqS, this is the meaning of a sentence, i.e., $[\![\varphi]\!] := [\varphi]^+$. Accordingly, the meaning of a sentence $[\![p]\!]$ in InqS becomes the powerset of $|p| = \{w | w(p) = 1\}$ excluding the empty set. Given $\wp^+(S) := \wp^+(S) - \{\emptyset\}$, thus, $[\![\text{Lisa smiles}]\!] = \wp^+(\{w | \text{ Lisa smiles in } w\})$.

The classical meaning of a sentence φ as a set of possible worlds, called the informative content of φ, is retrieved by taking a union of all states that support φ:

(14) $info(\varphi) = \bigcup[\varphi]^+$.

Furthermore, following Groenendijk & Roelofsen (2014), a sentence φ is said to be inquisitive when φ is supported by at least one state and φ is not supported by $info(\varphi)$. That is, if φ is inquisitive, accepting $info(\varphi)$ does not make the supporting state as our updated common ground:

(15) φ is inquisitive iff $[\varphi]^+ \neq \emptyset$ and $info(\varphi) \notin [\varphi]^+$

(Adopted from Groenendijk & Roelofsen, 2014, 9)

3.2   Composition

Inspired by Hamblin's (1973) Alternative Semantics, Ciardelli & Roelofsen (2015); Theiler (2014) provide a framework which enables us to derive a set of information states as the meaning of a sentence without invoking a special composition rule like pointwise functional application.[5] In this framework, the semantic value of a sentence is a powerset of $|p|$, thus in each linguistic expression, all the type t's in the standard Montague grammar are replaced by $\langle\langle s,t\rangle, t\rangle$.

---

[4]The current paper adopts InqS instead of the basic inquisitive semantics framework, InqB, since one of the goals is to analyze unconditional sentences which involve the semantics of implication.

[5]Roelofsen & van Gool (2010) provide a machinery which directly extends Alternative Semantics, which faces several problems as pointed out by Ciardelli & Roelofsen (2015).

(16)   a. $\llbracket \text{Lisa} \rrbracket \in D_e$; $\llbracket \text{Lisa} \rrbracket := \mathsf{lisa}$
        b. $\llbracket \text{smile} \rrbracket \in D_{\langle e, \langle \langle s,t \rangle, t \rangle \rangle}$; $\llbracket \text{smile} \rrbracket := \lambda x. \wp^+(|\mathsf{smile}(x)|)$

In Alternative Semantics, the meaning of a sentence is also considered as a set of propositions, a set of sets of possible worlds. The crucial difference from Alternative Semantics is that in inquisitive semantics, the set is not unconstrained but downward closed, thus if $|p| \in \llbracket \varphi \rrbracket$ and $|q| \subseteq |p|$, $|q| \in \llbracket \varphi \rrbracket$.

In order to give a non-inquisitive semantics to declarative sentences containing a disjunction, I introduce a declarative operator, DECL, which amounts to perform updates of the state in the classical sense. The semantics of DECL is based on Roelofsen & van Gool's (2010) Focus operator F and naturally adapted to the framework of InqS as can be seen in (17).[6]

(17)   a. $\sigma \vDash^+ \text{DECL}\varphi$ iff $\sigma \neq \emptyset$ and $\sigma \subseteq \mathit{info}(\varphi)$
        b. $\sigma \vDash^- \text{DECL}\varphi$ iff $\sigma \neq \emptyset$ and $\sigma \cap \mathit{info}(\varphi) = \emptyset$
        c. $\sigma \vDash^\circ \text{DECL}\varphi$ iff $\sigma = \emptyset$.

DECL renders all sentences into non-inquisitive ones using info defined in (14).[7]

(18) $[\text{DECL}\varphi]^+ = \wp^+(\mathit{info}(\varphi))$

Finally, in formalizing the semantics of alternative and polar questions, I adopt Roelofsen & van Gool's (2010) notion of excluded possibility. Following the standard assumption in the dynamic semantics, the semantics of a sentence φ encapsulate how φ updates the common ground. If a world w is not included in any information state that supports φ, then w *is excluded by* φ. In other words, the set of possibilities excluded by φ is the set of states that support $\neg\text{DECL}\varphi$ and codified as $\llparenthesis \varphi \rrparenthesis$:

(19) $\llparenthesis \varphi \rrparenthesis := [\neg\text{DECL}\varphi]^+ = \wp^+(W - \mathit{info}(\varphi))$

3.3   Analysis

I propose that both *ding6* and *waak6ze2* denote an inquisitive disjunction, which join two sets, but they are different in that only *ding6* has a syntactic-feature-driven requirement which resists the DECL operator.

Let us take a look at *waak6ze2* first, which simply denotes disjunction:

(20)   For any type τ and $\llbracket \alpha \rrbracket, \llbracket \beta \rrbracket \in D_\tau$, $\llbracket \alpha \text{ WAAKZE } \beta \rrbracket := \llbracket \alpha \vee \beta \rrbracket$

I propose that declarative constructions, including modalized sentences and conditional antecedents, involve the DECL operator. In a declarative sentence like (1), thus, 'p waak6ze2 q' first forms a union $\wp^+(|p|) \cup \wp^+(|q|)$ and then the DECL operator (17) renders the sentence non-inquisitive:

(21)   a. $\llbracket p \text{ WAAKZE } q \rrbracket = [p \vee q]^+ = \wp^+(|p|) \cup \wp^+(|q|)$
        b. $\llbracket \text{DECL}(p \text{ WAAKZE } q) \rrbracket = [\text{DECL}(p \vee q)]^+ = \wp^+(\mathit{info}(p \vee q)) = \wp^+(|p \vee q|)$

Turning to *ding6*, 'p ding6 q' has the same semantics as 'p waak6ze2 q' in that it denotes disjunction, $\llbracket \alpha \text{ DING } \beta \rrbracket := \llbracket \alpha \vee \beta \rrbracket$, while it has an additional lexical requirement that a sen-

---

[6]I would like to thank Katsuhiko Sano (personal communication) for his suggestion in the formalization of DECL.

[7]As noted by Roelofsen & van Gool (2010), DECL and F are akin to *non-inquisitive closure* in Groenendijk & Roelofsen (2009) and *existential closure* in Kratzer & Shimoyama (2002).

tence containing *ding6* remains inquisitive. That is, a sentence containing *ding6* cannot be the argument of the DECL operator. Thus, $[\![\text{DECL}(\alpha \text{ DING } \beta)]\!]$ is undefined. I implement this requirement using syntactic feature-checking. *ding6* carries an uninterpretable feature [*u*INQ].

(22)   a.   Lexicon: *ding6*: CONJ, [*u*INQ], $[\![\text{DING}]\!]$
       b.   Semantics: For any type τ and $[\![\alpha]\!], [\![\beta]\!] \in D_\tau$, $[\![\alpha \text{ DING } \beta]\!] := [\![\alpha \vee \beta]\!]$

The [*u*INQ] feature of *ding6* needs to be checked off by an operator $O_{[\text{INQ}]}$ which occupies C position and carries the interpretable feature [INQ]. The operator $O_{[\text{INQ}]}$ requires its complement to be inquisitive.

(23) $[\![O_{[\text{INQ}]}\varphi]\!]$ is defined if $[\![\varphi]\!]$ is inquisitive.

$O_{[\text{INQ}]}$ can be realized as the question particle *aa3* in (25-b) or the head of unconditional antecedent, *mou4leon6* in (29-a) below. Thus, '*p* ding6 *q*' is ungrammatical in a declarative due to its uninterpretable feature. The [*u*INQ] of *ding6* needs to be checked by $O_{[\text{INQ}]}$, but in a declarative sentence, the DECL operator renders the clause non-inquisitive, which conflicts with the lexical requirement of $O_{[\text{INQ}]}$, as in (24-a). If $O_{[\text{INQ}]}$ is not merged to the clause, [*u*INQ] remains unchecked and the derivation does not converge as in (24-b).

feature-checking

(24)   a.   *$O_{[\text{INQ}]}$ DECL [... ding6~~[*u*INQ]~~ ...]
       b.   *DECL [... ding6$_{[u\text{INQ}]}$ ...]

Turning to interrogative sentences like (5) and (6), I propose that *aa4* is a polar interrogative particle analogous to the interrogative complementizer Q in Roelofsen & van Gool, while *aa3* is an exhaustification presupposition particle analogous to the English L tone in Biezma & Rawlins (2012). The semantics of 'φ *aa4*' is the union of the possibilities which support φ and the possibilities which excluded by φ (25-a). In contrast, *aa3* presupposes that its prejacent is inquisitive and exhausts the common ground, thus no possibility is eliminated by the update of the prejacent proposition (25-b).

(25)   a. $[\![\varphi\text{-AA4}]\!] := [\![\varphi]\!] \cup \big[\!\big[\varphi\big]\!\big]$
       b. $[\![\varphi\text{-AA3}]\!]$ is defined if $[\![\varphi]\!]$ is inquisitive and $\big[\!\big[\varphi\big]\!\big] = \emptyset$. If defined, $[\![\varphi\text{-AA3}]\!] := [\![\varphi]\!]$.

There are at leat two ways to interpret *waak6ze2*-interrogatives. In one case, '*p* waak6ze2 *q*' takes the DECL operator. As a result, '*p* waak6ze2 *q* aa4?' is a *yes/no* question which is a set containing two propositions, 'Yes, *p* or *q*' and 'No, $\neg p \wedge \neg q$':

(26) $[\![\text{DECL}(p \text{ WAAKZE } q)\text{-AA4}]\!] = [\![\text{DECL}(p \vee q)]\!] \cup \big[\!\big[\text{DECL}(p \vee q)\big]\!\big]$
       $= \wp^+(|p \vee q|) \cup \wp^+(|\neg p \wedge \neg q|)$

In the other case, '*p* waak6ze2 *q*' does not take the DECL operator, and '*p* waak6ze2 *q* aa4?' is interpreted as an "open intonation" question:

(27) $[\![p \text{ WAAKZE } q\text{-AA4}]\!] = [\![p \vee q]\!] \cup \big[\!\big[p \vee q\big]\!\big] = \wp^+(|p|) \cup \wp^+(|q|) \cup \wp^+(|\neg p \wedge \neg q|)$

This is consistent with the intuition summarized in section 2 and the experimental result reported in Hara (2015). Recall that responding *yes* to *waak6ze2*-interrogatives is lowly rated but responding *no* receives a significantly higher rating. This is because the 'no'-answer can single out one possibility, '$\neg p$ and $\neg q$', while the proposition entailed by the 'yes'-answer remains inquisitive containing two possibilities, '*p*' and '*q*'.

In contrast, 'p ding6 q aa3?' cannot take the DECL operator, so it is always an alternative question which is denoted by a union of the two alternative propositions:

(28) $[\![p \text{ DING } q\text{-AA3}]\!]$ is defined if $|p \text{ DING } q| = \emptyset$.
         If defined, $[\![p \text{ DING } q\text{-AA3}]\!] = [\![p]\!] \cup [\![q]\!] = \wp^+(|p|) \cup \wp^+(|q|)$

3.4   Unconditionals and their connotations

Rawlins (2013) analyzes English unconditionals as universal quantification over a set of conditional sentences. Morpho-syntactically, Rawlins's analysis suits Cantonese unconditionals like (4) as they involves *ding6* or *waak6ze2*, which generates a union of sentences, and a universal quantifier *dou1* 'all'. The "antecedent" of the unconditional is inquisitive, $\wp^+(|p|) \cup \wp^+(|\neg p|)$, i.e., a union of two propositions. The head of the unconditional construction, *mou4leon6* is one of the inquisitive operators $O_{[\text{INQ}]}$, thus it requires its complement to be inquisitive and has the operation of pointwise functional application built in its lexical semantics:

(29)   a. $[\![\text{MOU4LEON6 } \alpha, \beta]\!]$ is defined iff $[\![\alpha]\!]$ is inquisitive.
              If defined, $[\![\text{MOU4LEON6 } \alpha, \beta]\!] := \bigcup\{[p \to q]^+ \,||p| \in [\![\alpha]\!] \text{ and } |q| \in [\![\beta]\!]\}$
         b. $[\![p \text{ DING } \neg p]\!] = [\![p]\!] \cup [\![\neg p]\!] = \wp^+(|p|) \cup \wp^+(|\neg p|)$
         c. $[\![\text{MOU4LEON6 } p \text{ DING } \neg p, q]\!] = [p \to q]^+ \cup [\neg p \to q]^+$

Thus, we obtain a union of conditional sentences, $[p \to q]^+ \cup [\neg p \to q]^+$. Finally, *dou1* defined in (30) renders the union of propositions into a conjoined sentences as in (30-b).

(30)   a. $[\![\text{DOU1}(\varphi)]\!] := \bigcap[\varphi]^+$
         b. $[\![\text{MOU4LEON6 } p \text{ DING } \neg p, \text{DOU1-}q]\!] = [(p \to q) \wedge (\neg p \to q)]^+$

This InqS analysis of unconditonals can also account for the connotations of unconditionals. Intuitively, an unconditional sentence 'whether or not p, q' gives rise to a consequent entailment, e.g., in (4), 'She will be full' is always true, and an independence connotation between two issues p and q. Indeed, it can be shown that the consequent entailment holds:

Proposition 1 (Consequent entailment). $\sigma \vDash^+ (p \to q) \wedge (\neg p \to q)$ imply that $\sigma \vDash^+ q$.

Proof. By assumption, $\sigma \vDash^+ p \to q$ and $\sigma \vDash^+ \neg p \to q$. By the definition of '$\to$' in (12), $\sigma \cap info(p) \vDash^+ p$ and $\sigma \cap info(p) \vDash^+ q$; $\sigma \cap info(\neg p) \vDash^+ \neg p$ and $\sigma \cap info(\neg p) \vDash^+ q$. By the definition of '$\vDash^+$', thus, $\sigma \cap info(p) \neq \emptyset$ and $\sigma \cap info(\neg p) \neq \emptyset$. Since $info(\neg p) = W - info(p)$, $\sigma \cap (W - info(p)) \vDash^+ q$. Thus, $\sigma \cap (W - info(p)) \subseteq info(q)$. Also, since $\sigma \cap info(p) \vDash^+ q$, $\sigma \cap info(p) \subseteq info(q)$. Therefore, $\sigma \subseteq info(q)$. Hence, $\sigma \vDash^+ q$.   □

As for the independence connotation, Franke (2009) defines that two propositions are (epistemically) independent when learning the truth/falsity of one proposition does not affect the truth/falsity of the other proposition.[8] Franke's notion of independence can be carried over to the InqS framework as in (31) on the basis of Aher & Groenendijk (To appear) definition of $\diamond$ in InqS as in (32).

(31) p and q are *independent* in σ if $\sigma \vDash^+ \diamond x$ and $\sigma \vDash^+ \diamond y$ imply $\sigma \vDash^+ \diamond(x \wedge y)$, for all
         $x \in \{p, \neg p\}$ and $y \in \{q, \neg q\}$.   (Adapted from Franke, 2009, 266)

(32) $\sigma \vDash^+ \diamond\varphi$ iff φ is supposable in σ, i.e., $\sigma \cap info(\varphi) \vDash^+ \varphi$.
                                          (Adapted from Aher & Groenendijk, To appear, 9)

---

[8] See Sano & Hara (2014) for a dynamic extension of independence.

In showing the truth of the unconditional "whether or not p, q" entails the independence between p and q, the following fact from Groenendijk & Roelofsen (2014) is used:

Fact 2 (Persistence modulo inconsistency)**.** If $\sigma \vDash^{+} \varphi$ and $\sigma \supseteq \tau \neq \emptyset$, then $\tau \vDash^{+} \varphi$

(Adapted from Groenendijk & Roelofsen, 2014, 11)

Proposition 3 (Independence)**.** If $\sigma \vDash^{+} (p \rightarrow q) \wedge (\neg p \rightarrow q)$, then p and q are independent in σ.

Proof. Suppose $\sigma \vDash^{+} (p \rightarrow q) \wedge (\neg p \rightarrow q)$. By Proposition 1, $\sigma \vDash^{+} q$. We need to show the following four implications: 1. If $\sigma \vDash^{+} \Diamond p$ and $\sigma \vDash^{+} \Diamond q$, then $\sigma \vDash^{+} \Diamond(p \wedge q)$. 2. If $\sigma \vDash^{+} \Diamond \neg p$ and $\sigma \vDash^{+} \Diamond q$, then $\sigma \vDash^{+} \Diamond(\neg p \wedge q)$ 3. If $\sigma \vDash^{+} \Diamond p$ and $\sigma \vDash^{+} \Diamond \neg q$, then $\sigma \vDash^{+} \Diamond(p \wedge \neg q)$. 4. If $\sigma \vDash^{+} \Diamond \neg p$ and $\sigma \vDash^{+} \Diamond \neg q$, then $\sigma \vDash^{+} \Diamond(\neg p \wedge \neg q)$. Case 1: Suppose $\sigma \vDash^{+} \Diamond p$ and $\sigma \vDash^{+} \Diamond q$. By the assumption, p is supposable in σ. Thus, $\sigma \cap info(p) \vDash^{+} p$ and $\sigma \cap info(p) \neq \emptyset$. Also, since $\sigma \vDash^{+} q$ and Fact 2, $\sigma \cap info(p) \vDash^{+} q$, which implies $\emptyset \neq \sigma \cap info(p) \cap info(p) = \sigma \cap info(p \wedge q)$. Since $info(p \wedge q) \subseteq info(p)$, $\sigma \cap info(p \wedge q) \subseteq \sigma \cap info(p)$. By Fact 2, $\sigma \cap info(p \wedge q) \vDash^{+} p \wedge q$. Therefore, $p \wedge q$ is supposable in σ, thus $\sigma \vDash^{+} \Diamond(p \wedge q)$. Case 2: Similar to Case 1. Case 3: Suppose $\sigma \vDash^{+} \Diamond p$ and $\sigma \vDash^{+} \Diamond \neg q$. By the assumption, $\sigma \cap info(\neg p) \vDash^{+} \neg q$ and $\sigma \cap info(\neg p) \neq \emptyset$. By the definition of '$\neg$', $\sigma \cap info(\neg p) \vDash^{-} q$. By the assumption and Fact 2, $\sigma \cap info(\neg p) \vDash^{+} q$. Since the assumptions contradict each other, this case never happens. Case 4: Similar to Case 3. $\qquad\qquad\square$

4   Conclusion

This paper provided a compositional analysis that explains their distributions and interpretations of the two kinds of Cantonese disjunction, *waak6ze2* and *ding6* in InqS. Both *ding6* and *waak6ze2* denote an inquisitive disjunction which contains a set of alternative propositions, but they differ in that *ding6* has an extra syntactic-feature-driven requirement that the clause remains to be inquisitive. The analysis is consistent with the result of the experiments reported in Hara (2015) and correctly derives the interpretations of the interrogative constructions and the connotations which arise from unconditional sentences.

As mentioned in section 2, Haspelmath (2007) states that there are a lot of languages which have two kinds of disjunction, "interrogative disjunction and standard disjunction". However, Cantonese disjunction system is different from Egyptian Arabic system (Winans, 2013, See). Even the distribution of the two disjunctions in Mandarin Chinese, which is quite similar to Cantonese in a number of respects, is different from that in Cantonese. More specifically, Mandarin "interrogative" disjunction, *háishe*, is available under modals and conditional antecedents (Huang, 2010, See), while Cantonese *ding6* is ungrammatical under those constructions. It would be interesting to investigate whether this distributional difference arises from the difference in the disjunction system or the properties of modals and conditionals.

References

Aher, Martin & Jeroen Groenendijk. To appear. Deontic and epistemic modals in suppositional [inquisitive] semantics. In *Proceedings of the Nineteenth Sinn und Bedeutung conference*, .

Ciardelli, Ivano & Floris Roelofsen. 2015. Alternatives in Montague grammar. In Eva Csipak & Hedde Zeijlstra (eds.), *Proceedings of Sinn und Bedeutung 19*, 161–178.

Franke, Michael. 2009. *Signal to Act: Game Theory in Pragmatics*: Universiteit van Amsterdam dissertation.

Groenendijk, Jeroen & Floris Roelofsen. 2009. Inquisitive semantics and pragmatics. In Jesus M. Larrazabal & Larraitz Zubeldia (eds.), *Meaning, Content, and Argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric*, 41–72. University of the Basque Country Publication Service. `www.illc.uva.nl/inquisitive-semantics`.

Groenendijk, Jeroen & Floris Roelofsen. 2014. Toward a suppositional inquisitive semantics. In Martin Aher, Daniel Hole, Emil Jerabek & Clemens Kupke (eds.), *Logic, Language, and Computation: revised selected papers from the 10th International Tbilisi Symposium on Language, Logic, and Computation*, Springer.

Hamblin, C. L. 1973. Questions in Montague English. *Foundations of Language* 10. 41–53.

Hara, Yurie. 2015. Two kinds of disjunction in Cantonese: A rating study. In *Proceedings of Logic and Engineering of Natural Language Semantics 12 (LENLS12)*, 11–24.

Haspelmath, Martin. 2007. Coordination. In Timothy Shopen (ed.), *Language typology and syntactic description, volume II: complex constructions*, 1–51. Cambridge University Press.

Huang, Rui-heng Ray. 2010. *Disjunction, Coordination, and Question: a comparative study*: National Taiwan Normal University dissertation.

Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In Y. Otsu (ed.), *Proceedings of the 3rd Tokyo conference on psycholinguistics*, 1–25. Hitsuji Syobo.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. University of California Press.

Rawlins, Kyle. 2013. (un)conditionals. *Natural Language Semantics* 40. 111–178.

Roelofsen, Floris & Sam van Gool. 2010. Disjunctive questions, intonation and highlighting. In M. Aloni andH. Bastiaanse, T. de Jager & K. Schulz (eds.), *Logic, Language, and Meaning: Selected Papers from the 17th Amsterdam Colloquium*, 384–394. Berlin Heidelberg: Springer.

Sano, Katsuhiko & Yurie Hara. 2014. Conditional independence and biscuit conditional questions in dynamic semantics. In *Proceedings of SALT 24*, 84–101. Ithaca, NY: Cornell University.

Theiler, Nadine. 2014. *A Multitude of Answers: Embedded Questions in Typed Inquisitive Semantics*. University of Amsterdam MA thesis.

Winans, Lauren Danae. 2013. *Alternative Disjunctions in Egyptian Arabic*. University of California Los Angeles MA thesis.

Yuan, Mengxi. 2015. *Chinese Discourse Adverbs as Presupposition Triggers*: City University of Hong Kong dissertation.

Right node raising, scope, and plurality[*]

Aron Hirsch[1] and Michael Wagner[2]

[1] Massachusetts Institute of Technology, Cambridge, MA, USA
aronh@mit.edu
[2] McGill University, Montreal, QC, Canada
chael@mcgill.ca

Abstract

This paper establishes an empirical generalization about right node raising ('RNR'), based on a novel test using interactions with focus operators: the rightmost constituent ('pivot') can take wide scope above the conjunction, as predicted by ex situ theories of RNR, and as previously argued in Sabbagh (2007). Wide scope is possible, however, only if the pivot is not associated with a position within an island (pace Sabbagh, 2007). Our results are compatible with a hybrid approach to RNR: there is (i) an ex situ analysis achieved by rightward ATB-movement which is island-sensitive, and (ii) an analysis with the pivot in situ, which does not require movement.

1 Ex situ and in situ analyses of right node raising

Right node raising refers to coordinate constructions like (1), where a constituent apparently associated with both conjuncts is pronounced once, at the right edge of the sentence.

(1)   John likes and Mary hates **The Fantastic Mr. Fox**.

Two families of RNR analyses have been advanced in the literature. According to **ex situ** approaches, the surface position of the pivot is external to the coordinate structure. A number of authors have proposed a derivation where the pivot originates internal to the conjuncts and undergoes rightward across-the-board ('ATB') movement to adjoin above the conjunction (Ross, 1967; Hankamer, 1979; Postal, 1974; Sabbagh, 2007, i.a.):[1]

(2) $[[_{TP} [_{TP}$ John likes $t_1]$ and $[_{TP}$ Mary hates $t_1]]$ [The Fantastic Mr. Fox]$_1]$

The second family of approaches (**in situ**) take the pivot to stay internal to the conjunction. One in situ approach involves backward ellipsis: separate occurrences of the pivot are present in each conjunct, and the occurrence of the pivot in the left conjunct elides (Wexler and Culicover, 1980; Swingle, 1995; Kayne, 1994; Wilder, 1995; Hartmann, 2001; Ha, 2008, i.a.).

(3)   [John likes ~~The Fantastic Mr. Fox~~] and [Mary hates The Fantastic Mr. Fox]

Another in situ approach involves a multi-dominance structure: there is a single occurrence of the pivot with multiple mothers, one in each conjunct (McCawley, 1982; Wilder, 1999; Bachrach and Katzir, 2007, 2009). In (1), a single occurrence of *The Fantastic Mr. Fox* is both the complement of *likes* in the left conjunct and the complement of *hates* in the right conjunct.[2]

---

[1]Another type of ex situ approach allows for co-ordination of non-standard constituents, and assumes that a surface structure like (1) is base-generated, rather than derived by movement (Steedman, 1985). In the remainder the paper, we will assume that ex situ structures (if possible at all) involve ATB-movement.

[2]While proponents of in situ accounts usually assume either ellipsis or multi-dominance, Barros and Vicente (2011) argue that *both* are necessary, but see Larson (2012) for counterarguments.

Our aim in this paper is to establish whether, and under what circumstances, an ex situ analysis is available. We focus, in particular, on a dissociate scope prediction:[3]

(4) **Dissociative Scope Prediction**
    a.   If an ex situ analysis is available, the pivot can scope above *and*.
    b.   If only an in situ analysis is available, the pivot must scope below *and*.

We introduce a new way to diagnose the scope of the pivot based on focus operators (§2), which supports the following generalizations:

(5) **Empirical generalizations**
    a.   The pivot can scope above *and*.
    b.   When the (base) position of the pivot is within an island, the pivot must scope below *and*.

The result in (5-a) confirms that RNR has an available ex situ analysis, as already argued in Sabbagh (2007). The result in (5-b) is most straightforwardly understood if RNR has *both* an ex situ analysis and an in situ analysis. The ex situ analysis involves rightward ATB-movement which respects islands, contrary to the finding in Sabbagh (2007). In island configurations, the ex situ analysis is blocked, while the in situ analysis is preserved, predicting (5-b).

In the remainder of the paper, we consider other ways of testing for the scope of the pivot involving universal quantifiers (§3, building on Sabbagh, 2007) and the distributive operator *different* (§4, building on Abels, 2004). Although Sabbagh (2007) has argued that data with universal quantifiers indicate that the pivot can take scope above *and* out of an island, we reconcile his data with (5-b). In the final section (§6), we identify a residual puzzle.

2   Focus operators

Our strategy to diagnose the scope of the pivot will be to insert *only* into it. To set up discussion, let us first introduce the analysis of *only* by considering the mono-clausal example:

(6)   John likes **only Hamlet**.

We assume that *only* is a two-place operator (cf. Rooth, 1985; Drubig, 1994; Krifka, 2006; Wagner, 2006). *Only* combines with a focused XP of any type α, and then a constituent of type $<\alpha,st>$. We take it that, in (6), this argument structure obtains with covert movement of *only Hamlet* to derive the LF in (7). *Only* combines with *Hamlet* (type e) and the derived property $\lambda x.\lambda w.$ *John likes x in w* (type $<e,st>$).

(7) $[_{TP} [_{DP}$ only Hamlet$_F] \lambda 1 [_{TP}$ John likes $t_1]]$

Only is defined as in (8) (cf. Rooth, 1985; Wagner, 2006). Only(x)(f) presupposes f(x)(w) and asserts the falsity at w of all alternatives f(a) that are not entailed by f(x).

(8) $[\![$only$]\!] = \lambda x_\alpha$. $\lambda f_{<\alpha,st>}$. $\forall a \in$ ALT(x) $[f(a)(w) \rightarrow (f(x) \Rightarrow f(a))]$
     Presupposition: f(x)(w)

Then, (6) presupposes that John likes Hamlet and asserts that for every alternative a to Hamlet, John does not like a, unless John liking a is already entailed by him liking Hamlet.

Given this analysis, *only* is predicted to scopally interact with other operators. Taglicht (1984), for instance, observes (9), which is ambiguous between (9-a) and (9-b). (9-a) derives from an LF like

---

[3]Note that Bachrach and Katzir (2007) propose a way to make a multi-dominance account compatible with wide scope readings. We will not be able to pursue this possibility further here due to space restrictions.

(10-a) where *only Spanish* QRs to a position above *advised*, while (9-b) derives from an LF like (10-b) where QR targets a position below *advised*.[4]

(9) You were advised to learn **only Spanish**.

    a. Only Spanish is such that you were advised to learn it.   *only > advised; (10-a)*

    b. What you were advised was to learn only Spanish.   *advised > only; (10-b)*

(10) a. $[_{TP}$ $[_{DP}$ only Spanish$_F]$ $\lambda 1$ $[_{TP}$ you were advised to learn $t_1]]$

      b. $[_{TP}$ you were advised $[_{TP}$ $[_{DP}$ only Spanish$_F]$ $\lambda 1$ $[_{TP}$ to learn $t_1]]$

2.1 Diagnosing *only > and*

Let us now consider an RNR example minimal to (1), but with *only* inserted into the pivot, as in (11):

(11) John likes and Mary hates only The Fantastic Mr. Fox.

A first available reading is one with *and > only*, according to which (11) licenses the inference:

(12) ... therefore, John dislikes all the other movies and Mary likes all the other movies.

This inference is expected to be valid if the pivot including *only* is interpreted separately within each conjunct — i.e. *and > only* — as in the baseline:

(13) John likes only The Fantastic Mr. Fox and Mary hates only The Fantastic Mr. Fox.

Critically, (11) also allows for a second reading with *only > and*, which can be paraphrased: only The Fantastic Mr. Fox is such that both John likes it and Mary hates it. This reading is brought out:

(14) John and Mary like and dislike many movies, but their taste is nearly identical. **John likes and Mary hates only the Fantastic Mr. Fox**.

    The observation that *only > and* is possible provides a direct argument that RNR has an available ex situ analysis, per the Dissociative Scope Prediction. On the in situ analysis, the pivot is internal to the conjuncts in the narrow syntax and would have to undergo covert ATB-movement. It is clear from the unavailability of an *only > and* reading in the in situ baseline in (13) that such covert movement is not generally available. An ex situ analysis is required to derive *only > and*.

    Regarding the first reading (*and > only*), it derives on an ex situ analysis with ATB-reconstruction of the pivot. In fact, since we will argue below that RNR has both an ex situ analysis and in situ analysis, there are *two* derivations which each yield *and > only* in (11): *and > only* is the "reconstructed" reading of the ex situ analysis and straightforwardly derives on the in situ analysis.

2.2 The effect of an island

It is well known that RNR is grammatical even when the conjunct-internal base position of the pivot is within an island:

(15) John found a critic who likes and Mary found a critic who hates The Fantastic Mr. Fox.

Yet, we demonstrate that the generalization is more nuanced: although RNR is grammatical in island configurations, the range of available readings is more restricted. In particular, the pivot cannot scope above *and*. Once again, this is diagnosed by inserting *only* into the pivot.

---

[4]The surface distribution of *only* can be accounted for as follows: *only* can adjoin to any constituent, as long as at LF the syntactic configuration [only(x)][y] can be established. Consistent with this, covert movement of *only* and its associate leading to Taglicht ambiguities is subject to familiar LF-movement constraints (for example, it is not possible from a finite clause) (Wagner, 2009).

As a control, the data in (16) demonstrate that a DP containing *only* cannot normally escape a relative clause island. (16-a) involves overt movement with negative inversion, and (16-b) shows that *only* must be interpreted internal to the relative clause and scope below *someone*.

(16)   a.   *Only The Fantastic Mr. Fox did John find a critic who likes.
        b.   John found someone who likes only The Fantastic Mr. Fox. *(some > only, *only > some)*

Turning to our RNR test case, consider the bolded example, first in a context biasing *and > only*:

(17)   It seems some critics either like or dislike almost everything. **John found a critic who likes and Mary found a critic who hates only The Fantastic Mr. Fox.**

In (17), the sentence is felicitous, indicating that it can convey the *and > only* reading where John managed to find a critic who likes a single movie, and Mary one who hates a single movie. But, now consider the following, which biases *only > and*:

(18)   It's hard to find critics that disagree with each other. **#John found a critic who likes and Mary found a critic who hates only only The Fantastic Mr. Fox.**

The incoherence of (18) suggests that *only > and* is not available. This is evidence that for island-violating RNR, the pivot must scope below the conjunction. The interaction between *only* and the conjunction hence provides evidence for the generalizations in (5).[5]

To account for the island facts, we suggest that the ex situ analysis of RNR is island-sensitive, and that RNR has, in addition, an in situ analysis. In island-violating RNR, only the in situ analysis is available, so the pivot must scope below *and*. Given that the ex situ analysis involves ATB-movement, island sensitivity is expected from the behavior of leftward ATB-movement:

(19)   *Which book does John know the man who likes _ and Mary know the woman who hates _?[6]

3   Universal quantifiers

Sabbagh (2007) provides a different way of diagnosing the scope of the pivot, and reports results convergent with our first generalization (that *pivot > and* is available), but contradictory to our second (that *pivot > and* is bled in island configurations). In Sabbagh's test cases, the pivot is universal:

(20)   Some nurse gave a flu shot to and administered a blood test for every patient.

The subject *some nurse* necessarily scopes above *and*; that is, it is necessarily the same nurse that gives the flu shot and administers the blood test.[7] The pivot is *every patient*. *Every patient* is scopally

---

[5] We do not explore the exact generalization of what constrains wide scope reading here. We think that the wide-scope pattern might be similar to that of overt leftward ATB movement. It seems different from LF movement in that wide scope seems possible in double object constructions (i.). And it might be also in allowing extraction from a finite clause (ii.), although the judgments we obtained were conflicting; it is different from rightward movement, in that wide scope readings seem to be possible with preposition stranding:

(i.)   a.   John sent some student and Mary sent some professor every piece of gossip that they heard of around the department.
        b.   I found that Bill likes and you found that Bill hates only The Fantastic Mr. Fox.
        c.   Bill raved about and yet Mary complained about only The Fantastic Mr. Fox.

Another caveat is that we found that examples involving the operator *even* might pattern differently. To explore the exact generalization more carefully is beyond the scope of this paper.

[6] A parallel example in which island-violating RNR feeds *wh*-movement is reported as acceptable in Bachrach and Katzir (2007); however, the judgments we elicited suggest that this is in fact not possible. We think this data point has to be tested more extensively.

[7] The assumptions is *v*Ps are conjoined, and some nurse ATB-moves out of the conjuncts to spec-TP, where it is interpreted as scoping above the conjunction (Moltmann, 1992; Fox, 2000).

commutative with *and*, but interacts with *some nurse*. Because *some nurse* scopes above *and*, if *every patient* can scope above *some nurse*, it must also be able to scope above *and*. Sabbagh observes that this wide scope reading is available in (20).

3.1 Island configurations

Let us now consider Sabbagh's datum with an island configuration:

(21)   John knows someone who speaks and Bill knows someone who wants to learn every Germanic language.

The universal pivot, as above, is commutative with *and*, but scopally interacts with *someone*. Sabbagh observes that (21) is ambiguous between *some > every* and *every > some*. The critical result is the availability of *every > some*, paraphrased: for every Germanic language x, John knows someone who speaks x and Bill knows someone who wants to learn x.

This reading could come about in two ways. On an in situ analysis, *every Germanic language* could QR above *someone* separately in each conjunct:

(22)   **Derivation 1** (after in situ analysis): (*and > every > some*)

a.   [$_{TP}$ [$_{DP}$ every Germanic language] $\lambda 1$ John knows someone who speaks t$_1$]
and [$_{TP}$ [$_{DP}$ every Germanic language] $\lambda 2$ Bill knows someone who wants to learn t$_2$]

On an ex situ analysis, the LF corresponding to the narrow syntactic structure is (23-b), where *every Germanic language* has ATB-moved above the conjunction. If the pivot is above the conjunction, it also must scope above *someone*.

(23)   **Derivation 2** (after ex situ analysis): (*every > and > some*)

a.   [$_{TP}$ John knows someone who speaks t$_1$
and Bill knows someone who wants to learn t$_1$] $\lambda 1$ [$_{DP}$ every Germanic language]

For (21) to be a counter-example to our conclusions, it must be demonstrated that the ex situ derivation is an available one. The premise of the argument is that the example in (24) is unambiguous (*some > every*, **every > some*), where one conjunct is considered in isolation:

(24)   John knows someone who speaks every Germanic language.

This shows that *every Germanic language* cannot QR above *someone* within the conjunct in violation of the island, ruling out Derivation 1. To account for (21), rightward ATB-movement must, then, not be sensitive to islands, allowing *every > some* to derive from an ex situ structure like Derivation 2.

We believe, however, based on the judgments we elicited, that the baseline in (24) may actually allow for *every > some*. We can improve the test if we find an island configuration where QR of the universal within the conjuncts is more clearly blocked. Postal (1998) observed that *wh*-extraction out of a relative clause in a definite DP is notably worse than in other DPs (cf. Szabolcsi, 2006). Building on this, we modify (23) by changing the DP to a definite:

(25)   John knows the professor who wrote every exam.

The judgment in (25) is clearer than in (24): (25) allows a reading with *the > every*, but an inverse reading with *every > the* is marginal. The RNR test case is:

(26)   John knew the professor who wrote and bribed the student who graded every exam.

Most of our informants report that the judgment tracks the baseline: just as *every > and* is marginal in (25), it is marginal in (26). (26) can convey (implausibly) that John knew the professor who wrote every exam and bribed the student who graded every exam (*and > every*), but a plausible reading where professors and students co-vary with exams is at least difficult.

A more direct test of the predictions is to substitute *and* with *or*, which directly interacts with the universal quantifier.

(27)   A: How did John get through the semester?
        B: He blackmailed the professor who wrote or bribed the student who graded every exam.

Again, a wide scope of *every* was at least difficult for most of our consultants.

4   Distributive and cumulative readings

Another source of evidence for our generalizations are interactions with distributive operators such as *different* or *similar*. Consider (28) (see Abels (2004) for a parallel example):

(28)   Carl and Sue married two quite different people.

This sentence conveys that Carl and Sue each married one person, and the person Carl married was quite different from the person Sue married ('distributive reading'). The generalization about distributive readings is that they are possible if the constituent modified by *different* can take scope over a plural or a coordination (cf. Beck, 2000, for cases not involving RNR).

Turning to RNR, a distributive reading is available in non-island configurations (cf. Abbott, 1976, 642, for an example with *similar*):

(29)   Bob and Sally brought their partners over for dinner. Their partners did not get along. **Bob dates and Sally married two quite different people.**

The distributive reading in (29) is expected under an ex situ account, where *two quite different people* attaches above the coordination and can hence scope over it, but not under an in situ account.[8] The in situ baseline is (30), where the distributive reading is clearly impossible:

(30)   Bob dates two quite different people and Sally married two quite different people.

To make explicit how the distributive reading derives if *different* scopes above the coordination, we sketch a possible analysis of (29):

(31)   [[[Bob dates t_1] and [Sally married t_1]] [two quite different people]$_1$]

After Beck (2000), *different* may be analyzed as a reciprocal relational adjective. Simplifying somewhat for exposition, *two quite different people* denotes:

---

[8]This is acknowledged in Abels (2004). However, Abels (2004) argues that this apparent advantage of ex situ analyses is illusory. Consider:

(i)   John says that Friederike must, and that Konrad may, record two quite different songs.

This example is meant to illustrate that raising the VP also licenses a distributive reading, but it is unexpected that the DP could take scope over the coordination from inside the VP. Abels concludes that there must be some other mechanism to allow distributive scoping than movement. We have doubts about this judgment, however. The following at least seems to lack a distributive reading:

(ii.)   Trent Reznor did sing and Johnny Cash will sing two quite different songs.

A second argument that Abels makes in order to show that distributive scope are not sensitive to movement constraints is the following:

(iii.)   Konrad and Frederike know men who have written quite different songs.

But here, it is the constituent *men* that *quite different songs* out-scopes and distributes over, and then *Konrad and Friederike* can distribute over the two men—no scope-taking out of the island is necessary to derive the intended reading.

(32)　　$[\![$two quite different people$]\!]$
　　　　$= \lambda f_{<e,t>}$. $\exists X$ [people(X) & #(X)=2 & $\forall y[y \leq X \rightarrow \forall z[z \leq X$ & $y \neq z \rightarrow$ different(y,z)]] & f(X)]

The conjunction must provide a predicate of pluralities to be taken as the argument of $[\![$two quite different people$]\!]$. To achieve this, we make two assumptions: (i) predicate abstraction can occur separately in each conjunct (λy . *Bob dates y, λz . Sally married z*), and (ii) *and* is optionally analyzed as something other than logical *and*. Link (1984) provides the entry for *and* in (33-a), which composes with the aforementioned predicates to yield the predicate of pluralities in (33-b).

(33)　　a.　$[\![$and$_2]\!] = \lambda P_{<e,t>}$. $\lambda Q_{<e,t>}$. λX . $\exists y \exists z[X=y \oplus z$ & P(y) & Q(z)]
　　　　b.　$[\![\&P]\!] = \lambda X$. $\exists y \exists z[X=y \oplus z$ & Bob dates y & Sally married z]

Applying $[\![\&P]\!]$ to $[\![$two quite different people$]\!]$ delivers the distributive reading: the sentence is true iff there is some two-membered plurality of people $y \oplus z$ where y and z are different from one another (from (32)), and Bob dates y, and Sally married z (from (33-b)).

4.1 Island configurations

Let us now turn to island contexts. In non-RNR structures, the distribution of distributive readings appears to be sensitive to islands such that the distributive reading is unavailable when *different* is trapped in an island:

(34)　　Bob and Sally know a politician that married two quite different people.

Consistent with our second generalization, a distributive reading is similarly absent in RNR if the pivot originates in an island. This effect of islands has already been observed for adjunct islands in Abels (2004), and we illustrate with a relative clause island:

(35)　　Our co-workers, Bill and Sally, do not get along. At Thanksgiving, I always go over to Bill's house and you always go over to Sally's house. So, **#I know the man who married and you know the woman who married these two quite different people.**

4.2 Cumulative readings

The operator *total* seems to show a convergent paradigm to the one seen above with *different*:

(36)　　a.　There seems to have been an odd sort of concert going on in the street last night. **A man hummed and a woman whistled four songs total.**

　　　　b.　There seems to have been an odd sort of concert going on in my street last night and in my sister's street. **I heard a man who hummed and my sister heard a man who whistled four songs total.**

The bolded sentence in (36-a), but not (36-b) allows a "cumulative reading" in which each of the two singers individually sang less than four songs, but four songs in total were sung between them. In sum, the generalizations about distributive and cumulative readings are as expected if there is an ex situ analysis of RNR, but only when respecting island constraints, convergent with our conclusions in the preceding sections.

5　Summary

Data are consistent with the empirical generalizations in (37), restated from the introduction. The most conservative interpretation of these generalizations leads to the analytical conclusions in (38).

(37)   **Empirical generalizations**
  a.   In island-respecting RNR, the pivot can scope above *and* (cf. Sabbagh 2007).
  b.   Island violating RNR is grammatical, but the pivot must scope below *and*.

(38)   **Analytical conclusions**
  a.   RNR has an island-sensitive ex situ analysis.
  b.   RNR has an island-insensitive analysis that is only compatible with *and < pivot*.

Throughout the paper, we have argued that the island-sensitive ex situ analysis in (38-a) involves rightward ATB-movement of the pivot from conjunct-internal base positions to a conjunct-external position. Our data are most straightforwardly analyzed with a "hybrid" analysis of RNR where the ex situ and in situ approaches – rather than being mutually exclusive – co-exist. Island-respecting RNR is ambiguous between this ex situ analysis and an in situ analysis, and island-configurations disambiguate in favor of an in situ analysis. A hybrid approach has been previously proposed in Valmala (2013).

6   A remaining puzzle: conjunctive pivots

To conclude the paper, we point out a class of examples which constitute a residual puzzle:[9]

(39)   Madonna sang and McCartney wrote *American Pie* and *Let it Be*, respectively.

This example has two features: the pivot is a conjunction and *respectively* is present. The interpretation involves pairwise distribution of the conjuncts in the pivot between the conjoined clauses: Madonna sang *American Pie* and McCartney wrote *Let it Be*.[10] This reading can only obtain with the pivot ex situ, as is clear from:

(40)   Madonna sang *A. Pie* and *Let it Be* and McCartney wrote *A. Pie* and *Let it Be* (respectively).

The sentence in (40) is an in situ baseline, and is truth-conditionally distinct from (39): whereas Madonna sang just *American Pie* in (39) and McCartney wrote just *Let it Be*, (40) says that Madonna sang both songs and that McCartney wrote both songs.

    The same profile of example with a distributed reading seems to be available in island-configurations, although this seems to require an ex-situ pivot:

(41)   I know the artist who sang and John has met the artist who wrote *American Pie* and *Let it Be*, respectively.

Hence, the example in (41) is an apparent counter-example to our second generalization: the pivot can take scope above the conjunction despite the island configuration.

    But, the puzzle runs deeper still. Although the pivot must be ex situ, the individual conjuncts in the pivot behave syntactically as though they were within their respective conjuncts. Consider:

(42)   John likes and Mary hates **himself** and **herself**, respectively.

By Condition A, *himself* must be c-commanded by John, and *herself* by Mary. Note that this result cannot be obtained by reconstructing the entire conjunction *himself and herself* into the two clausal conjuncts: if the entire conjunction reconstructed, *himself* and *herself* would both be present in both conjuncts, which would result in an incorrect interpretation and violations of Condition A.[11]

---

[9]Data in this section build on examples in Moltmann (1992) and references therein. The binding data also relate to observations in Hirsch and Marty (2015).

[10]See Gawron & Kehler (2004) for an analysis of *respectively*. As in the examples above with *different*, *and* is analyzed as something other than logical conjunction in their analysis of *respectively*.

[11]*Himself* and *herself* could be 'exempt anaphors' in this example, not subject to Condition A. Reinhart and Reuland (1993) observe that the anaphor in the coordination in (i) is exempt, as diagnosed by the non-complementarity between it and the pronoun:

NPI licensing shows a similar pattern:

(43)   Madonna won't sing and McCartney won't record **any song** and **any album**, respectively.

The NPIs are licensed if *any song* and *any record* are c-commanded by negation separately in their respective conjuncts. If the conjunction itself reconstructed, in addition to a wrong meaning, the NPIs would not be licensed, since *and* is an intervener for NPI licensing (Linebarger, 1987; Guerzoni, 2006).

Further binding data involving Conditions B and C show that the conjuncts in the pivot not only *can* behave as though they were within the conjoined clauses, but in fact *must* (just like in the case of regular RNR, cf. Phillips,i.a.):

(44)   a.   *John$_1$ likes and Mary$_2$ hates **him**$_1$ and **her**$_2$, respectively.   *(Condition B)*
          b.   *He$_1$ likes and she$_2$ hates **John**$_1$ and **Mary**$_2$, respectively.   *(Condition C)*

In (44-a), co-indexation of *John* and *him* is impossible, as is co-indexation of *Mary* and *her*, indicating that *John* necessarily c-commands *him* and *Mary* c-commands *her*. The pattern in (44-b) is parallel.

In sum, it appears that the pivot must be ex situ, but the construction is not island-sensitive and the individual conjuncts in the pivot behave as they though are necessarily in situ in separate conjuncts. While a full account of this pattern is beyond the scope of the present paper, one possibility is that the entire pivot, in (39) *American Pie and Let it Be*, ATB-moves above the conjunction and then the individual conjuncts *American Pie* and *Let it Be* are "metalinguistically" reconstructed as syntactic objects into separate positions in the two clausal conjuncts. If island effects arise through an interaction of syntax and interpretation, this metalinguistic strategy may not respect islands. This type of metalinguistic RNR might also be understood with an approach that views (at least some cases of) RNR as a consequence of incremental structure building during processing (Phillips, 1996) or as a consequence of delayed spell-out (Bachrach and Katzir, 2009). We leave further exploration of this idea for future research.

References

Abbott, B. (1976). Right node raising as a test for constituenthood. *Linguistic Inquiry*, pages 639–642.

Abels, K. (2004). Right node raising: Ellipsis or movement? In Wolf, M. and Moulton, K., editors, *Proceedings of the Northeast Linguistic Society Proceedings Meeting in 2003 at Stony Brook*, pages 45–60.

Bachrach, A. and Katzir, R. (2007). Spelling out QR. In *Proceedings of Sinn und Bedeutung*, volume 11, pages 63–75, Barcelona.

Bachrach, A. and Katzir, R. (2009). Right Node Raising and Delayed Spellout. In *InterPhases: Phase-Theoretic Investigations of Linguistic Interfaces*, pages 283–316.

Barros, M. and Vicente, L. (2011). Right node raising requires both ellipsis and multidomination. *University of Pennsylvania Working Papers in Linguistics*, 3(1):15.

Beck, S. (2000). The semantics of different: Comparison operator and relational adjective. *Linguistics and Philosophy*, 23(2):101–139.

Drubig, H. B. (1994). Island constraints and the syntactic nature of focus and association with focus. In *Arbeitspapiere des Sonderforschungsbereichs 340: Sprachtheoretische Grundlagen der Computerlinguistik*, volume 51. Sonderforschungsbereich 430, Tübingen/Stuttgart.

Fox, D. (2000). *Economy and semantic interpretation*. Number 35. MIT Press.

Guerzoni, E. (2006). Intervention effects on NPIs and feature movement: towards a unified account of intervention. *Natural Language Semantics*, 14(4):359–398.

(i)   Max boasted that the queen invited Lucie and himself/him for a drink.

The datum in (44-a), however, shows that complementarity does hold in the conjunctive pivot examples, calling into question an analysis of (42) in terms of exempt anaphora.

Ha, S. (2008). *Ellipsis, Right Node Raising, and Across-The-Board Movement*. PhD thesis, Boston University.

Hankamer, J. (1979). *Deletion in Coordinate Structures*. Outstanding Dissertations in Linguistics. Garland.

Hartmann, K. (2001). *Right node raising and gapping: Interface conditions on prosodic deletion*. John Benjamins, Amsterdam.

Hirsch, A. and Marty, P. (2015). Binding evidence for conjunction reduction. MS. MIT.

Kayne, R. S. (1994). *The antisymmetry of syntax*. MIT Press, Cambridge, Ma.

Krifka, M. (2006). Association with focus phrases. In Molnár, V. and Winkler, S., editors, *The architecture of focus*, volume 82 of *Studies in Generative Grammar*, chapter 4, pages 105–136. Mouton de Gruyter, Berlin.

Larson, B. (2012). A dilemma with accounts of right node raising. *Linguistic Inquiry*, 43(1):143–150.

Linebarger, M. (1987). Negative polarity and grammatical representation. *Linguistics and Philosophy*, 10:325–387.

McCawley, J. D. (1982). Parentheticals and discontinuous constituent structure. *Linguistic Inquiry*, 13(1):91–106.

Moltmann, F. (1992). *Coordination and comparatives*. PhD thesis, Massachusetts Institute of Technology.

Phillips, C. (1996). *Order and Structure*. PhD thesis, MIT.

Postal, P. M. (1974). *On raising: One rule of English grammar and its theoretical implications*. The MIT Press.

Postal, P. M. (1998). *Three investigations of extraction*, volume 29. MIT Press.

Reinhart, T. and Reuland, E. (1993). Reflexivity. *Linguistic inquiry*, pages 657–720.

Rooth, M. (1985). *Association with Focus*. PhD thesis, University of Massachussetts, Amherst.

Ross, J. R. (1967). *Constraints on variables in syntax*. PhD thesis, MIT.

Sabbagh, J. (2007). Ordering and linearizing rightward movement. *Natural Language and Linguistic Theory*, 25(2):349–401.

Steedman, M. (1985). Dependency and coordination in the grammar of Dutch and English. *Language*, 61:523–568.

Swingle, K. (1995). On the prosody and syntax of right node raising. *Ms., UC Santa Cruz.*

Szabolcsi, A. (2006). Strong vs. weak islands. *The Blackwell companion to syntax*, 4:479–531.

Taglicht, J. (1984). *Message and Emphasis. On focus and scope in English*, volume 15 of *English Language Series*. Longman, London and New York.

Valmala, V. (2013). On right node raising in Catalan and Spanish. *Catalan Journal of Linguistics*, 12:219–251.

Wagner, M. (2006). Association by movement. Evidence from NPI-licensing. *Natural Language Semantics*, 14(4):297–324.

Wagner, M. (2009). Focus, topic, and word order: A compositional view. In van Craenenbroeck, J., editor, *Alternatives to Cartography*, pages 53–86. Mouton De Gruyter, Berlin/New York.

Wexler, K. and Culicover, P. (1980). *Formal principles of language acquisition*. Mit Press.

Wilder, C. (1995). Rightward movement as leftward deletion. *On extraction and extraposition in German*, pages 273–309.

Wilder, C. (1999). Right node raising and the LCA. In *Proceedings of WCCFL*, volume 18, pages 586–598.

Portion Readings are Count Readings, not Measure Readings

Keren Khrizman[1], Fred Landman[2,3], Suzi Lima[4],
Susan Rothstein[1,3], Brigitta R. Schvarcz[1]

[1] Bar-Ilan University [2] Tel Aviv University [3] Universität Tübingen [2015-16]
[4] Universidade Federal do Rio de Janeiro

Abstract

We assume, following Rothstein 2010, 2011, 2016 and Landman 2011, 2013, 2015, a semantic theory of the mass-count distinction which defines the notion of *count* in terms of disjointness, non-overlap, and in which the mass-count distinction applies to noun phrases of any complexity (i.e. not just lexical nouns). We derive, following Rothstein 2011, two interpretations for pseudo-partitives like *three glasses of wine*, a container classifier interpretation which we show to be count and a measure interpretation which we argue to be mass. We then address portion readings. Partee and Borschev 2012 discussed portion readings as a subcase of measure readings. We argue, against this, that portion readings do not pattern with measure readings, because portion readings are count. We discuss three ways of deriving portion readings. This adds, for *three glasses of wine*, two new portion interpretations: a contents-classifier interpretation and a free portion interpretation. We show that, in the semantic framework given, all portion interpretations come out as count, setting them apart from measure interpretations. We show that the distinctions between measure interpretations and portion interpretations derived here hold cross-linguistically in a number of typologically distinct languages.

1 The disjointness principle

Rothstein 2010, 2011 and Landman 2011 develop theories of the mass-count distinction in which *disjointness*, non-overlap, is a characterizing property of count noun denotations, while mass nouns generally denote sets that allow overlap. Based on this, Landman 2013, 2015 develops a formal theory in which the mass-count distinction applies not just to lexical nouns, but to noun phrases in general.

In order to provide a setting for some of the later discussion in this paper, we briefly sketch here some aspects of this theory. Landman 2013, 2015 formulates a compositional mechanism which associates with the standard denotation of a noun phrase a *base* set, a set that generates this denotation under the sum operation $\sqcup$. For count nouns, the base set is the set in terms of which the members of the denotation are counted and to which distribution takes place. Since counting requires disjointness, for count nouns the base set is required to be (contextually) disjoint; *vice versa*, noun phrases whose base set is disjoint at every (relevant) world-time index are thereby classified as count noun phrases.

For example, for singular noun *cat* with denotation $CAT_{w,t}$, the base set is the set $CAT_{w,t}$ itself, and $CAT_{w,t}$ is required to be disjoint. The same disjoint set $CAT_{w,t}$ is the base set for the (non-disjoint) denotation $*CAT_{w,t}$ (the closure of $CAT_{w,t}$ under $\sqcup$) of plural noun *cats* (and it forms a partition within the set of parts of the denotation of *the cats*). For complex noun phrases like *pet cats* and *three pet cats* the base set derived is $CAT_{w,t} \cap PET_{w,t}$. Since $CAT_{w,t}$ is required to be a disjoint set, $CAT_{w,t} \cap PET_{w,t}$ comes out as disjoint as well. This means that the complex noun phrases *pet cats* and *three pet*

cats also come out as count noun phrases. And this means that the pluralities in the denotation of *three pet cats* are correctly counted as *three* in relation to their parts in the set $CAT_{w,t} \cap PET_{w,t}$. It also means that *three pet cats* can combine with distributive predicates that require a plural *count* subject, as in (1), and the distribution is, for a plurality in the denotation of the subject noun phrase, to its parts in the set $CAT_{w,t} \cap PET_{w,t}$:

(1) *Three pet cats* should *each* have their own basket.

The above paragraphs function as an example of the kind of theory that we assume as a background for this paper: a semantic theory in which the mass-count distinction is based on disjointness and applies to all noun phrases. We need something like this here, because we are concerned with the interpretation and mass-count nature of classifier phrases and measure phrases, (subkinds of *pseudopartitives*), which are complex noun phrases. However, for the purposes of the present paper we need little of the background theory: the following principle regulating singular count interpretations of noun phrases will suffice for most of our purposes:

(2) **Disjointness Principle**:
If the interpretation set of a noun phrase is disjoint at to every (relevant) index, the noun phrase is singular count.

2   Classifier readings and measure readings

Rothstein 2011 gives a semantics for expressions like *three glasses of wine*, showing that they are ambiguous between a container classifier interpretation and a measure interpretation.

(3) a. […] and next to them, she would put *three glasses of wine*, three glasses of water, three sugar cubes and one golden coin. [Radomir Ristic, 'The great spirits of fate,' in: The Crooked Path Journal, issue 2, p. 45, 2008]
b. […] and just as you are going to roast, open the paste, pour in *three glasses of Madeira wine*, close the paste well, tie it up securely, roast it two hours […] [Alexis Soyer, *Gastronomic Regenerator, A Simplified and Entirely New System of Cookery*, 6th ed., 1849, Simkin, Marshall and. co., London]

(3a) shows the container classifier interpretation: *three glasses of wine* in (3a) means *three glasses containing wine*. (3b) shows the measure interpretation: *three glasses of wine* in (3b) means *wine to the amount of three glassfuls*.
With Rothstein 2011, we assume that the container-interpretation of *three glasses of wine* is derived by applying the interpretation of *glass* to that of *wine*, deriving NP *glass of wine*. This is pluralized and then modified by *three*. For this to be possible, the count noun *glass,* interpreted as disjoint set $GLASS_{w,t}$, shifts to a classifier *glass* with as interpretation a function from sets to sets.
With Landman 2015, we assume that the classifier interpretation is based on the interpretation of the noun *glass* and a *contents* function (Rothstein 2011 uses a containment relation):

(4) **contents**$_{[P,Q,c]w,t}$ is a function that maps - relative to container property P contents property Q and context c - a P-container onto its relevant contents at index w,t.

We take this to mean that if **contents**$_{[GLASS,WINE,c]w,t}(x) = y$, then x is presupposed to be a glass and y, the relevant contents of glass x, is the wine in glass x, on the condition that what is in the glass is standard enough to count. This means that the amount of wine in the glass should be within the

normality range, relative to GLASS, WINE and c, and the contents should have a constitution which is normal for wine in context c; so, it can be mixed (say with water) if that is normal in the context, but it cannot be polluted, etc. We will suppress the property parameters here (they require an intensional formulation of the theory which is beyond the scope of this paper, see Landman 2015 for more discussion). We derive the following container classifier interpretations:

(5) a. $[_{\text{container classifier}}\ glass\ ] \to \lambda P\lambda x.\text{GLASS}_{w,t}(x) \wedge P(\textbf{contents}_{w,t}(x))$
    b. $[_{\text{NP}}\ glass\ of\ wine\ ] \to \lambda x.\text{GLASS}_{w,t}(x) \wedge \text{WINE}_{w,t}(\textbf{contents}_{w,t}(x))$
       The set glasses whose contents is wine.

Now we observe that, because *glass* is a count noun, its denotation $\text{GLASS}_{w,t}$ is required to be disjoint at every index w,t. Since the interpretation of the noun phrase *glass of wine* is defined at w,t through intersection with $\text{GLASS}_{w,t}$, it follows that the interpretation of *glass of wine* is disjoint at every index. Hence, with the Disjointness Principle:

(6) **Containers are count:** *Glass of wine*, with container classifier *glass*, is a singular count noun phrase.

We assume, with Rothstein 2011, that the measure interpretation of *three glasses of wine* is based on measure function $\mu_{[\text{GLASS,WINE,c}]w,t}$ which maps relative to parameters GLASS, WINE and context c, entities at index w,t onto real values on a measure scale. Here [GLASS,WINE,c] determines what measure amount is one glass in context c, when what is measured is wine. As before, for the **contents** function, we simplify and write $\textbf{glass}_{w,t}$ for $\mu_{[\text{GLASS,WINE,c}]w,t}$.
    Following Landman 2004 and Rothstein 2011, we assume that in the derivation of the measure interpretation of *three glasses of wine*, first the measure *glass* composes with the higher number predicate ($\lambda n.\ n=3$), and then the result (*three glasses*) intersects with the complement *wine*. Thus the interpretation of the measure phrase *three glasses of wine*, based on measure *glass*, is:

(7) a. $[_{\text{measure}}\ glass] \to \lambda N\lambda P\lambda x.(\textbf{glass}_{w,t} \circ N)(x) \wedge P(x)$
    b. *three + glass(es)* $\to \lambda P\lambda x.\textbf{glass}_{w,t}(x)=3 \wedge P(x)$
    c. $[_{\text{NP}}\ three\ glasses\ of\ wine\ ] \to \lambda x.\textbf{glass}_{w,t}(x)=3 \wedge \text{WINE}_{w,t}(x)$
       Wine to the measure of three glassfuls.

Now, first we observe that at a normal index w,t the set $\lambda x.\textbf{glass}_{w,t}(x)=3 \wedge \text{WINE}_{w,t}(x)$ is not disjoint (*normal* here means that we ignore borderline interpretations, like Ø). This, by itself, doesn't tell us much, because, after all, at a normal index w,t the interpretation of plural noun *cats*, $*\text{CAT}_{w,t}$, is not disjoint either. But the latter has a disjoint base $\text{CAT}_{w,t}$.
    Landman 2015 argues that at normal indices w,t, the set $\lambda x.\textbf{glass}_{w,t}(x)=3 \wedge \text{WINE}_{w,t}(x)$ does not have a disjoint base: any base that is strong enough to generate $\lambda x.\textbf{glass}_{w,t}(x)=3 \wedge \text{WINE}_{w,t}(x)$ under $\sqcup$ will by necessity itself overlap (see Landman 2015 for details).
    Without going into technicalities, we observe here that, due to the interpretation of the measure function, the overlap in $\lambda x.\textbf{glass}_{w,t}(x)=3 \wedge \text{WINE}_{w,t}(x)$ is *pervasive*. The measure function is real valued on the set of all parts of the wine; this means that the measure function assigns measure values r to very small, possibly unmeasurably small parts of the wine. Pick any of those, say z. Now take any two quantities of wine x_1 and x_2 that measure $3 - r$ and don't overlap z and don't overlap each other: $x_1 \sqcup z$ and $x_2 \sqcup z$ are going to be overlapping quantities of wine in $\lambda x.\textbf{glass}_{w,t}(x)=3 \wedge \text{WINE}_{w,t}(x)$. Note furthermore that the set $\lambda z.\textbf{glass}_{w,t}(z)=r \wedge \exists x[\text{WINE}_{w,t}(x) \wedge z \sqsubseteq x]$ itself is not disjoint either.
    The above informal discussion points at the reasonableness of the result:

(8) **Measures are mass:** *three glasses of wine* with measure *glass(es)* is a mass noun phrase.

And this is good, because, as Rothstein 2011 argues, measure phrases pattern with mass nouns. In (9a), the partitive predicate *of the two hundred croquette balls* (the famous Dutch *bitterballen*) is count, as shown by the infelicity of the combination with the mass determiner *much*. In (9b) the partitive predicate *of the six kilos of croquette balls* is felicitous, and hence mass, and it has a measure interpretation: croquette balls were eaten, and at the end of the party there's not much left of the huge pile that was on the table originally:

(9) a. #ʔ***Much** of the two hundred croquette balls* was eaten at the party.
    b. ✓***Much** of the six kilos of croquette balls* was eaten at the party.

So far we have derived a two-way ambiguity of *three glasses of wine*: if the complex noun phrase denotes glasses, it has a container classifier reading, if it denotes wine, a measure reading.
    We get the same two interpretations for measure phrases like *three liters of wine*:

(10) a. We poured *three liters of wine* in the brew.
    b. He arrived home and knocked on the door with *one liter of milk*. His mother said to him: "I asked you for two liters. Where is the second one?" Her son said to her: "It broke, mother." [Matilda Koén-Sarano (ed.), Folktales of Joha, Jewish Trickster, p. 22, The Jewish Publication Society, Philadelphia, 2003]

The measure interpretation in (10a) is derived in the same way as the measure interpretation of *three glasses of wine* above:

(11) $[_{NP}$ *three liters of wine*$] \rightarrow \lambda x.\textbf{liter}_{w,t}(x)=3 \wedge \text{WINE}_{w,t}(x)$ mass measure interpretation
                    Wine to the measure of three liters.

The container classifier interpretation is derived by shifting *liter* to a container interpretation.

(12) a. CONTAINER_c is a property that maps index w,t onto disjoint set $\text{CONTAINER}_{w,t}$,
       a set of containers at w,t whose nature is determined by context c.
    b. $\textbf{container}_{\textbf{c}} = \lambda P\lambda x.\text{CONTAINER}_{c,w,t}(x) \wedge P(\textbf{contents}_{w,t}(x))$

We assume that implicit operation $\textbf{container}_{\textbf{c}}$ can compose with the measure *liter* (applied to implicit number predicate **one** $(\lambda n.n=1)$):

(13) a. $[_{\text{container classifier}}$ *liter* $] \rightarrow \textbf{container}_{\textbf{c}} \circ (liter_{\text{measure}}(\textbf{one}))$
              $= \lambda P\lambda x.\text{CONTAINER}_{c,w,t}(x) \wedge P(\textbf{contents}_{w,t}(x)) \wedge \textbf{liter}_{w,t}(\textbf{contents}_{w,t}(x))=1$
    b. $[_{NP}$ *liter of wine*$] \rightarrow$
              $\lambda x. \text{CONTAINER}_{c,w,t}(x) \wedge \text{WINE}_{w,t}(\textbf{contents}_{w,t}(x)) \wedge \textbf{liter}_{w,t}(\textbf{contents}_{w,t}(x))=1$
              The set of containers whose contents is one liter worth of wine.

Clearly, the disjointness requirement on $\text{CONTAINER}_{w,t}$ makes this interpretation of *liter of wine* disjoint at every index w,t, and hence, by the Disjointness Principle:

(14) **Container-shifted measures are count:** *liter of wine*, with *liter* shifted to a container
        classifier, is a singular count noun.

3 Portion readings

So far we have two readings for *three α of wine*: a count container interpretation where it denotes sums of three α-containers, and a mass measure interpretation where it denotes wine to the measure of three α-*fuls*. And this suggests the diagnostics: when the complex noun phrase denotes (sums of) containers it's count, and when it denotes wine it's mass. While the first part of this equation is unproblematic, the second part is made more complex by the existence of portion readings. Portion readings were discussed by Partee and Borschev 2012. They discuss two kinds of portion readings: contents readings, which they relate (as we will) to container readings, and concrete portion readings - which we will rebaptize *free portion readings* -, which they regard as a subcase of measure readings. The status of and relations between contents readings and free portion readings stay a bit up in the air in Partee and Borschev's discussion. Our aim here is to show that developing the semantics for portion readings in a theory that formulates the mass-count distinction for complex noun phrases in terms of disjointness helps to clarify the status of portion readings:

(15) **Portion readings** of *three α of wine* are readings that are count even though the complex noun phrase denotes wine.

Case 1: Shape classifiers

We introduce portion readings by looking at shape classifiers like *hunk, slice, strand, drop, gulp*:

(16) a. A *hunk* of meat / two *slices* of sausage / many *strands* of hair / three *drops* of wine /each *gulp* of wine.
   b. A hunk of meat is meat in the shape of a hunk – a drop of wine is a drop-shaped portion of wine.

As (16a) shows, shape classifier noun phrases, like *slice of sausage* and *drop of wine,* are count, like container classifier noun phrases. But shape classifier noun phrases have interpretations along the lines of (16b): like measure interpretations, they are intersective on the complement.
   The basics of the semantics of shape classifiers is straightforward: they have the same semantics as intersective modifiers:

(17) a. $[_{\text{shape classifier}} \ hunk] \rightarrow \lambda P \lambda x. \text{HUNK}_{w,t}(x) \wedge P(x)$ based on count noun *hunk*
   b. $[_{\text{NP}} \ hunk \ of \ meat] \rightarrow \lambda x. \text{HUNK}_{w,t}(x) \wedge \text{MEAT}_{w,t}(x)$
          Meat in the shape of a hunk, meat that forms a hunk.

Since *hunk* is a count noun, $\text{HUNK}_{w,t}$ is a disjoint set. Hence also $\lambda x. \text{HUNK}_{w,t}(x) \wedge \text{MEAT}_{w,t}(x)$ is a disjoint set, and by the Disjointness Principle *hunk of meat* is a singular count noun phrase.
   We see then that noun phrases like *hunk of meat* denote 'mass stuff', meat. But the meaning of the noun *hunk* tells you that meat in this denotation comes in the form of disjoint countable, separate portions (meaning, among others, that they are more than non-overlapping parts of a bigger body of meat, i.e. that you can pick them up and move them separately, etc.).

Case 2: Contents classifiers

We come back to *three glasses of wine.* With the interpretation of shape classifier noun phrases as our model for portion interpretations, we can show that Partee and Borschev's contents interpretation is a portion interpretation, and on that interpretation, *glass of wine* is a singular count noun phrase.

For this we only need to assume a very reasonable condition on the contents of containers:

(18) **Disjointness of Contents** (for normal situations w,t):
If $GLASS_{w,t}(x) \land GLASS_{w,t}(y) \land x \neq y$ then $\mathbf{contents}_{w,t}(x)$ and $\mathbf{contents}_{w,t}(y)$ are disjoint.

This says that, under normal circumstances, different glasses have disjoint contents. Disjointness of Contents makes the function $\mathbf{contents}_{w,t}$ a one-one function, so that the inverse function $\mathbf{contents}_{w,t}^{-1}$ is also defined. Above we analyzed, following Rothstein 2011, the container classifier reading of *glass of wine* along the lines of: *glass that has wine as contents*. With Partee and Borschev 2012, we assume that a contents-reading is also available, and we assume that it is based on the inverse contents function: *wine that is the contents of a glass*:

(19) a. $[_{\text{contents classifier}}\ glass\ ] \rightarrow \lambda P \lambda x.P(x) \land GLASS_{w,t}(\mathbf{contents}_{w,t}^{-1}(x))$
   b. $[_{NP}\ glass\ of\ wine\ ] \rightarrow \lambda x.WINE_{w,t}(x) \land GLASS_{w,t}(\mathbf{contents}_{w,t}^{-1}(x))$
   The set of wine-portions that are contents of glasses.

Now, if x_1 and x_2 are in the set (19b) and $x_1 \neq x_2$, then x_1 and x_2 are both the $\mathbf{contents}_{w,t}$ of a glass, and, because $\mathbf{contents}_{w,t}$ is a function, these glasses are different. It then follows from Disjointness of Contents that x_1 and x_2 are disjoint. The Disjointness Principle from Section 1 says that a noun phrase is count if its interpretation is disjoint at all *relevant* indices. We make the plausible assumption that relevant indices are restricted to indices that are *normal* with respect to the Contents Principle. With that, it follows that *glass of wine*, with contents classifier *glass*, is a count noun phrase.
   An example showing the contents reading is (20):

(20) I drank *fifteen glasses of beer*, five flutes, five pints, and five steins. I drank five *of the fifteen glasses of beer* before my talk and the rest after it.

In (20), the container classifier reading is irrelevant (I didn't ingest the glasses), what I drank was beer. The relevant reading in (20) is not a measure reading either, because I did not drink 15 glassfuls of beer: one glassful is a contextually fixed amount, but (20) expresses that I drank portions of beer of different size. And the partitive predicate *of the fifteen glasses of beer* is count. So, the natural reading in (20) is the contents reading.

Case 3: Free portion interpretations

Portion interpretations can be implicit, as in the following example:

(21) Eén patat met, één zonder, en één met satésaus, alstublieft.
   One french fries with [mayonnaise], one without, and one with peanut sauce, please.

For (21), we assume that the mass noun *patat* shifts with implicit operation $\mathbf{portion_c}$:

(22) a. $PORTION_c$ is a property whose content is determined by context c such that for all relevant indices w,t: $PORTION_{c,w,t}$ is disjoint.
   b. $\mathbf{portion_c} = \lambda P \lambda x.P(x) \land PORTION_{c,w,t}(x)$
   c. $[_{NP}\ patat] \rightarrow \lambda x.PATAT_{w,t}(x) \land PORTION_{c,w,t}(x)$

In (21) and (22c) mass noun *patat* shifts to singular portion count noun phrase *[portion of] patat*.

Coming back once more to *three glasses/liters of wine*. We assume that the shifting operation **portion_c** adds one more interpretation possibility to the readings derived so far: we assume that implicit operation **portion_c** can compose with the measure interpretation of *glass* (applied to implicit number predicate **one** ($\lambda n.n=1$)):

(23) a. [_portion classifier_ *glass*] $\rightarrow$ **portion_c** $\circ$ (*glass*_measure (**one**))
     b. [_NP_ *glass of wine*] $\rightarrow \lambda x.\text{WINE}_{w,t}(x) \wedge \text{PORTION}_{c,w,t}(x) \wedge \textbf{glass}_{w,t}(x)=1$
                    The set of portions of wine that each amount to one glass-ful.

The interpretation of *glass of wine* derived is a *portion* interpretation, it is *count* (since PORTION_{c,w,t} is disjoint, and hence the interpretation derived is disjoint as well); but it is not a contents interpretation, since the portions in the interpretation are *free*, they are not linked to containers. This interpretation we find in (24):

(24) The instructions are to pour *three cups of soy sauce* in the brew, the *first* after 5 minutes, the *second* after 10 minutes, the *third* after 15 minutes. I have a good eye and a very steady hand, so I pour *them* straight from the bottle.

In (24), when I pour, the soy sauce is never in a cup. But I count what I pour: *them* refers to countable cup-size portions.

We see that formulating the analysis in terms of disjointness allows us to clarify portion readings beyond the discussion of Partee and Borschev. We started out with Rothstein's container classifier reading and measure reading for *three glasses of wine*. We have added to that two portion readings:

-the contents reading relates to the container interpretation of *glass*: take the inverse of the contents function;

-the free portion reading relates to the measure interpretation of *glass*: compose with the portion operation. Both readings derived are portion readings and count.

4 Cross-linguistic evidence

If measure interpretations are mass and portion interpretations count, we expect portion interpretations to occur in contexts where measure interpretations are excluded, In this section we point out that this distinction seems to be supported cross-linguistically.

Case 1. Dutch

As argued in Doetjes 1997 and Rothstein 2011, measures in Dutch are uninflected for number; when a measure occurs with plural inflection, it gets a classifier interpretation. Thus, *vijftien liter water/*fifteen liter[-] of water has a measure interpretation, but *vijftien liters water/*fifteen liter[plur] of water only has a classifier interpretation. (25a,b) show that measure phrases are mass: (25a) has the uninflected form *liter*, it has a measure interpretation, and it is only compatible with the mass determiner *het meeste/*most[sing]: (25a) expresses mass comparison, comparison in terms of volume:

(25) a. ✓*Het meeste* van de vijftien *liter* water was weggelekt.   measure - mass
      Most [sing]   of   the fifteen   liter[-] water had leaked away.
     b. #*De meeste* van de vijftien *liter*   water waren weggelekt.
       Most[plur]   of the fifteen liters[-] water had leaked away.
     c. *De meeste* van de vijftien *liters*   water waren weggelekt.   portion - count
       Most [plur] of   the fifteen   liter[plur] water   had leaked away.

The plural on *liters* in (25c) indicates a classifier reading, which in this case is a portion reading (derived by shifting *liter* to *liter-container* and taking the contents interpretation of the latter).

(25c), with count-determiner *de meeste/*most[plur], expresses count comparison, comparison in terms of *number* of portions.

Case 2. Hungarian

Schvarcz 2014 argues that Hungarian suffix *–nyi* is an operator which takes nouns and shifts them to measures (similar to *–ful* and *–worth* in English, but more general and more productive). The continuation in (26a) triggers a portion interpretation of *Három pohár bort/*three glasses of wine: the expression *refilled twice* points at *three* separate portions:

(26) a. ✓*Három pohár bort* ittam a parti-n,   a pincér kétszer töltötte újra   a poharam.
      three   glass wine drink on the party the waiter twice   fill   again my glass
      I drank three glasses of wine at the party, the waiter refilled my glass twice.
   b. #*Három pohár-nyi bort* ittam a parti-n,   a pincér kétszer töltötte újra a poharam.
      three   glass-ful   wine drink on the party the waiter twice   fill   again my glass
      I drank three glass-fuls of wine at the party, the waiter refilled my glass twice.

When we replace *pohár/*glass by measure *pohár-nyi/*glass-ful, the portion reading disappears: the continuation in (26b) is infelicitous. If portion readings are a special case of measure readings, the infelicity of *-nyi* in the data in (26) is completely unexpected (notice that there is nothing *per se* wrong or 'unportionlike' about drinking three glassfuls of wine). On the analysis given here these data neatly fall into place.

Case 3. Russian

In Russian, the incremental theme of imperfectives is a context in which measure phrases are infelicitous (Košelev 1996, see also Khrizman 2014):

(27) a. #Ivan sidit   na divane i   p'ët   *dva s polovinoj litra*   *vodki.*
      Ivan is sitting[IMP-PRES] on the sofa   and drinking [IMP-PRES] *two and a half*   *liter of vodka.*
   b. ?✓Ivan sidit   na divane i   p'ët   *stakan*   *čaja.*
      Ivan   is sitting[IMP-PRES] on the sofa and drinking [IMP-PRES] *a cup*   *of tea.*

While examples like (27b) are not perfect, there is a strong contrast with cases like (27a), which are completely impossible. Now, the infelicity observed by Košelev's means that we (and everybody else) must assume that in the context of (27a) measures cannot shift to non-measure interpretations. By the same reasoning, the improved felicity of (27b) means that the interpretation of *stakan čaja/*a cup of tea on which the felicity improves is not a measure interpretation. Since the container interpretation is again irrelevant, the interpretation in question will be a portion interpretation. Thus, the contrast in (27) shows that also in Russian, portion interpretations do not pattern with measure interpretations.

Case 4. Hebrew

Rothstein 2009 argues that in Hebrew, measure interpretations are only available if the number predicate can form a *constituent* with the putative measure. She argues that in double construct states the number predicate does *not* form a constituent with the measure. Thus, the double construct state in (28) only allows the first syntactic structure, not the second:

(28) *šlošet bakbukey* hayayin.
    three   bottle   wine[def]   [= the three bottles of wine]
    ✓[*šlošet* [*bakbukey* hayayin]]   #[[*šlošet bakbukey*] hayayin]

The infelicity of (29a) (from Rothstein 2009) shows that double construct states do not allow measure interpretations. But double construct states *do* allow portion interpretations, as shown in (29b):

(29) a. hizmanti esrim orxim ve- hexanti esrim ka'arot marak be sir gadol. rak šiva-asar orxim higiu,
          I invited twenty guests and I prepared twenty bowls (of) soup in a big pot. Only seventeen guests came,
          #ve- *šaloš ka'arot ha- marak ha- axronot* nišaru ba sir.
            and [three bowls   DEF soup   DEF last   ] remained in the pot.
            and [the last three bowls of soup   ] remained in the pot.
    b. ✓šatinu et (kol) *šlošet bakbukey hayayin* še-hu hevi.
        We drank (all) *the three bottles (of) wine*   that he brought.

Similarly, Rothstein 2009 argues that free genitive constructions, as in (30a), do not allow measure interpretations (see Rothstein 2009 for extensive discussion). But, again, free genitive constructions *do* allow portion interpretations, as shown in (30b):

(30) a. *sir šalem* šel marak.
        pot full   of soup   [= a pot full of soup]
    b. zarakti   *sir šalem šel marak*, ve- az šatafti   et ha-sir.
        I threw away *a full pot   of   soup*,   and then I washed out   the pot.

Case 5. Yudja

Lima 2014 argues that Yudja does not grammatically distinguish between mass nouns and count nouns and does not have null classifiers. She argues that nouns denote kinds and that count readings come in through a shift to what she calls 'maximally self-connected instantiations of the kind.'

(31) a. Txabïu apeta pe.
        Three   blood   dripped.
    b. Txabïu y'a a'i.
        Three   water are here.

The numerical phrases in (31) denote *free* portions of blood/water: they need not be the contents of any container, they need not be of any standardized shape, and the context determines what counts as a portion. But: portions of blood are *counted* in (31a).
      We argued above for shifting operation **portion**_c, which produces separable countable portions, but has no further inherent 'lexical' content, meaning that what counts as a portion is determined completely contextually. But: **portion**$_c$(BLOOD$_{w,t}$) is count.
      It seems that the operation **portion**$_c$ is exactly the operation needed to account for the Yudja facts in question, and here again, what is used is the fact that portion interpretations are count.

5 In sum

We started out with a theory of the mass-count distinction for noun phrases which links count nouns to disjointness. We derived, for *three glasses of wine*, a count container classifier interpretation (three glass-containers filled with wine) and a mass measure interpretation (wine measuring three glass-fuls). We then discussed three ways of forming portion interpretations, which added two

portion interpretations for *three glasses of wine*: a contents-classifier interpretation (three wine-portions that are wine-glass contents) and a free portion interpretation (three wine-portions that each measure one glass-ful). We showed that, unlike measure interpretations, which are mass, all portion interpretations are count. Finally we showed that this holds cross-linguistically.

khrizman@walla.com   landman@post.tau.ac.il   suzilima1@gmail.com
susan.rothstein@biu.ac.il   bridget.schvarcz@gmail.com

References

Doetjes, Jenny, 1997, *Quantifiers and Selection*, Ph.D. Dissertation, University of Leiden.

Khrizman, Keren, 2014, 'Genitive case and aspect in Russian,' in: Cassandra Chapman, Olena Kit, and Ivona Kučerová (eds.), *Proceedings of Formal Approaches to Slavic Linguistics* 22, *the McMaster Meeting*, Michigan Slavic Publications.

Košelev, A. D.,1996, 'Referencial'nyj podxod k analizu jazykovyx značenij,' in: *Moskovskij Lingvistǐceskij Al'manax* 1, pp. 82-194.

Landman, Fred, 2004, *Indefinites and the Type of Sets*, Wiley-Blackwell, Oxford.

Landman, Fred, 2011, 'Count nouns – mass nouns – neat nouns – mess nouns,' in: Michael Glanzberg, Barbara H. Partee and Jurǵis Šķilters (eds.), *Formal Semantics and Pragmatics: Discourse, Context and Models. The Baltic International Yearbook of Cognition, Logic and Communication* 6 http://thebalticyearbook.org/journals/baltic/issue/current

Landman, Fred, 2013, 'Iceberg semantics,' talk presented at the *Countability Workshop*, Heinrich Heine Universität, Düsseldorf, September 2013.

Landman, Fred, 2015, 'Iceberg semantics for count nouns and mass nouns: the evidence from portions,' talk presented at the *11th International Symposium of Cognition, Logic and Communication*, University of Latvia, Riga, December 11, 2015.

Lima, Suzi, 2014, *The Grammar of Individuation and Counting*, Ph.D. Dissertation, University of Massachusetts, Amherst.

Partee, Barbara and Vladimir Borschev, 2012, 'Sortal, relational, and functional interpretations of nouns and Russian container constructions,' in: *Journal of Semantics* 29,4, pp. 445-486.

Rothstein, Susan, 2009, 'Individuating and measure readings of classifier constructions: Evidence from Modern Hebrew,' in: *Brill Annual of Afroasiatic Languages and Linguistics* 1, pp. 106-145, Brill, Leiden.

Rothstein, Susan, 2010, 'Counting and the mass-count distinction,' in: *Journal of Semantics* 27.

Rothstein, Susan, 2011, 'Counting, measuring, and the semantics of classifiers,' in: Michael Glanzberg, Barbara H. Partee and Jurǵis Šķilters (eds.), *Formal Semantics and Pragmatics: Discourse, Context and Models. The Baltic International Yearbook of Cognition, Logic and Communication* 6 http://thebalticyearbook.org/journals/baltic/issue/current

Rothstein, Susan, 2016, *Semantics for Counting and Measuring,* Cambridge University Press, Cambridge.

Schvarcz, Brigitta R., 2014, *The Hungarians who Say -nyi: Issues in Counting and Measuring in Hungarian*, MA Thesis, Bar-Ilan University.

Homogeneity, Trivalence, and Embedded Questions*

Manuel Križ

Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS)
Département d'Études Cognitives, École Normale Supérieure, PSL Research University

Abstract

Plural predication is known for the homogeneity property: 'Adam wrote the books' is true if Adam wrote (roughly) all of the books, but its negation is true only if he wrote (roughly) *none* of them. Embedded questions show in many ways analogous behaviour: 'Agatha knows who was at the party' is intuitively true if Agatha is fully informed about who the guests were, whereas its negation 'Agatha doesn't know who was at the party' conveys that she has pretty much no idea who was there. We argue that the properties of questions in this connection can be explained as a direct consequence of the homogeneity of plural predication once the latter is viewed through the lense of trivalent logic.

1 Introduction

Sentences with definite plurals are known for the property of *homogeneity*:[1]

(1)   a.   Mr. Benfleet published the books. ⤳ Mr. Benfleet published **all** of the books.
         b.   Mr. Benfleet didn't publish the books. ⤳ Mr. Benfleet published **none** of the books.

Sentences with embedded questions show an analogous homogeneity effect:[2]

(2)   a.   Agatha knows who was at the party.
             ⤳ For **everybody** who was at the party, Agatha knows that they were there.
         b.   Agatha doesn't know who was at the party.
             ⤳ For **nobody** who was at the party does Agatha know that they were there.

The main goal of this contribution is to demonstrate that once homogeneity with plurals is viewed through the lense of trivalent logic, an explanation for the facts with embedded question arises naturally due to the fact that the answers to the question, which are propositions of the form *x was at the party*, are trivalent propositions when *x* is instantiated by a plurality of individuals.

2 Pluralities and Homogeneity

The facts with plural definites can be conceptualised naturally in terms of trivalent logic. If we assume, as is standard, that negation simply switches truth and falsity, but leaves the third truth value untouched, we arrive at the following diagnosis:

---

[1]Cf. [18], [15], [6], [16], [10], among others.
[2]See [21] for experimental corroboration.

(3)   Mr. Benfleet published the books.

> **true** iff Mr. Benfleet published all of the books.
> **false** iff Mr. Benfleet published none of the books.
> **neither** otherwise (i. e. if he published some, but not all of the books).

If we further assume, as is standard since [14], that definite plurals denote plural individuals, so that these sentencs are simple predicational structures with no quantification involved, we can state for following generalisation to capture the pattern according to which (distributive) predicates, as applied to pluralities, are trivalent.[3]

(4)   **Homogeneity for Distributive Predicates**
A (distributive) predicate P is true of a plurality a iff it is true of all parts of a, and false iff it is false of all parts of a. Otherwise it is undefined.

2.1   Quantifiers and Homogeneity Removers

A natural question to ask, then, is how quantifiers (ranging over atomic individuals) behave in this trivalent world, which was empirically investigated by [12]. What they found was this: A quantifier is true of its scope predicate iff it is true no matter how the undefined cases of the scope predicate are resolved, and false iff it is false no matter how the undefined cases of the scope predicate are resolved.[4]

Consider, for example, the sentence (5) when there are three publishers a, b, and c.

(5)   Every publisher is P.

Assume that P is as below, with # as the third truth value. There are two ways of resolving the undefined case of P, yielding P^1 and P^0.

$$
P = \begin{bmatrix} a \mapsto 1 \\ b \mapsto 1 \\ c \mapsto \# \end{bmatrix}
P^1 = \begin{bmatrix} a \mapsto 1 \\ b \mapsto 1 \\ c \mapsto 1 \end{bmatrix}
P^0 = \begin{bmatrix} a \mapsto 1 \\ b \mapsto 1 \\ c \mapsto 0 \end{bmatrix}
$$

The universal quantifier is true of P^1, but false of P^0. Thus, it matters which way the undefined cases are resolved and the universal quantifier is undefined of P.

Considering the alternative situation below, however, it does not matter how the undefined case is resolved, since in either case there is at least one individual of which the predicate is false, so that the universal quantification is automatically fase.

$$
P = \begin{bmatrix} a \mapsto 1 \\ b \mapsto 0 \\ c \mapsto \# \end{bmatrix}
P^1 = \begin{bmatrix} a \mapsto 1 \\ b \mapsto 0 \\ c \mapsto 1 \end{bmatrix}
P^0 = \begin{bmatrix} a \mapsto 1 \\ b \mapsto 0 \\ c \mapsto 0 \end{bmatrix}
$$

We thus arrive at the following pattern for the universal quantifier:

(6)   Every publisher is P.

> **true** iff P is true of all publishers.
> **false** iff there is at least one publisher of whom P is false.
> **undef.** otherwise.

---

[3]In what follows, we will assume that we live in a walled garden where no collective predicates exist. See [9] for a discussion of homogeneity with collective predicates.

[4]For first-order definable monotonic quantifiers, this amounts to Strong Kleene logic.

If we instantiate P with *accepted the books*, we obtain the following:

(7)   Every publisher accepted the books.

      true iff all publishers accepted all the books.
      false iff at least one publisher accepted none of the books.
      undef. otherwise (i.e. iff every publisher accepted at least some of the books, but at least one publisher didn't accept all of the books).

Ignoring collective predication, we can assume that *all* is just a universal quantifier over atomic individuals: It turns a plural predication into quantification over the parts of the plurality.

(8)   a. $[\![\text{The students came}]\!] = came(\iota x.students(x))$
       b. $[\![\text{All the students came}]\!] = \forall x' \preceq_{AT} \iota x.students(x) : came(x')$

If we look at predicates that are defined for all atomic individuals, such as simple intransitive verbs, then we find that *all* effectively functions as a homogeneity remover ([15]). (8b), for example, has the same truth conditions as (8a), but is false whenever (8a) is either false or undefined, and never undefined. Thus, the effect of *all* is to collapse undefinedness into falsity.

(9)   Mr. Benfleet published all the books.

      true iff Mr. Benfleet published all of the books.
      false iff there is at least one book that Mr. Benfleet didn't publish.
      neither never.

We will further assume that there is no difference between *all* in adnominal position and *all* in adverbial position, so that (10a) and (10b) are semantically equivalent.

(10)   a.   All the students came.
       b.   The students all came.

3   Embedded Questions

Assume that embedded questions semantically have the weakly exhaustive reading. Then viewing questions from the same perspective, we arrive at the following description:

(11)   Agatha knows Q.

      true iff Agatha knows all the true answers to Q.
      false iff Agatha knows none of the true answers to Q.
      neither otherwise (i.e. if she knows some, but not all true answers).

Some languages, like German, Dutch, and Chinese,[5] have elements which, when added *within* the question, remove homogeneity:

(12)   a.   Agatha weiß   nicht, wer auf der Feier war.
          Agatha knows not   who at   the party was
          roughly: 'Agatha has no idea who was at the party.'
       b.   Agatha weiß   nicht, wer aller auf der Feier war.
          Agatha knows not   who all   at   the party was
          'There is somebody who Agatha doesn't know was at the party.'

---

[5]Yimei Xiang (p.c.) on *dou*.

Depending on the *wh*-word, these elements may take different forms (*aller*, *alles*, *überall*; there is also some speaker variation), but they always correspond to the adverbial universal quantifier over the respective type which can also appear in declarative clauses.

3.1 Knowledge of Trivalent Answers

The above effects can be explained as a consequence of the fact that the answers to questions are trivalent propositions.[6] The way that the Hamblin set of a question is usually obtained is, intuitively, by replacing the *wh*-word with various individuals of the relevant sort. These individuals can be atomic individuals or pluralities, and in the latter case, the resulting proposition is trivalent. Thus, the Hamblin set of (13a), given in (13b), contains many trivalent propositions, namely all those where x is instantiated by a plurality of persons.

(13) a. Who came?
      b. $\{\lambda w.came_w(x) \mid x \text{ is a person or plurality of persons}\}$

In order to see the further consequences of this, we first have to look at what happens when *know* is applied to a trivalent declarative complement.

(14) Agatha knows that the girls came to the party.

On the usual analysis, *know* is a universal quantifier over worlds, its scope argument being, in this case, the proposition *that the girls came to the party*. This proposition is a trivalent predicate of worlds: it is true in those world where all the girls came, false in those worlds where none of the girls came, and otherwise undefined. Thus, we can apply the established behaviour of the universal quantifier as applied to a trivalent scope predicate and obtain the following prediction (for the *de re* reading, and ignoring the presupposition), which appears reasonable enough:[7]

(15) $\lambda w.\forall w' \in \text{Dox}_w(Agatha) : came_{w'}(\iota x.girl_w(x)))$

   true iff all of the y. p. came in all of Agatha's belief worlds.
   false iff in at least one of Agatha's belief words, none of the y. p. came.
   undef. otherwise

On the weakly exhaustive reading, *know Q* is always extensionally equivalent to *know p* for a particular p, namely the strongest true member of the Hamblin set. Thus, to know who came is to know *that x came*, where x is the maximal plurality of people who actually came. Since this is a trivalent proposition (if more than one person came), we can apply the pattern from (15) and immediately obtain a prediction for truth and falsity conditions as well as cases of undefinedness.

(16) Agatha knows who came ≡ Agatha knows that those came who did, in fact, come.
      $\lambda w.\forall w' \in \text{Dox}_w(Agatha) : came_{w'}(\iota y.came_w(y))$

   true in w iff in all of Agatha's belief worlds, all the people who came in w came.
   false in w iff in at least one of Agatha's belief worlds, none of the people who came in w came.
   undef. otherwise.

---

[6]Particular thanks are due to Benjamin Spector for encouraging me to explore this idea.
[7]The presupposition of *know* is, of course, being ignored here.

Note that beyond the view on the trivalence of plural predication presented above, no additional assumptions were needed to obtain these effects with embedded questions — in fact, extra assumptions would be needed to prevent them!

Since the homogeneity effects with embedded questions are a consequence of the fact that the answers to the question are trivalent propositions, we now have an avenue for explaining how universal adverbials within the question manage to remove homogeneity. They do just what they do in declarative sentences as well: They replace plural predication by universal quantification over atoms and thereby remove homogeneity, making the members of the Hamblin set bivalent.

(17)   a.   Wer ist aller gekommen?
             who is all   come
        b. $\{\lambda w.\forall y : y \preceq_{AT} x \to came_w(y) \mid x$ is a person or plurality of persons$\}$

With a bivalent proposition — *that all the people came who did, in fact, come* — as the argument of *know*, no undefinedness arises.

(18)   Agatha knows who all came $\equiv$ Agatha knows that all the people came who did, in fact, come.

        true in w iff in all of Agatha's belief worlds, all the people who came in w came.
        false in w iff in at least one of Agatha's belief worlds, at least one of the people who came in w didn't come.
        undef. never.

3.2   More on Homogeneity Removers

Besides being able to explain how a universal adverbial quantifier within the question causes the homogeneity effect to disappear, we can make sense of a number of further facts regarding the behaviour of adverbial quantifiers within questions.

We observe that *mostly* and *partly* do not have question-internal correlates. In particular, we do not know of any language in which the following is possible:[8]

(19)   #Agatha weiß, wer großteils gekommen ist.
        Agatha knows who mostly   came
        'For most of the people who came, Agatha knows they came.'

Our theory would predict the following meaning for a question containing the adverbial *mostly*:

(20)   a. #Wer ist großteils gekommen?
             who is mostly   come
        b. $\{\lambda w.\text{most}(\lambda y.y \preceq_{AT} x)(\lambda y.came_w(y)) \mid x \in D_\mathbf{e}\}$

This Hamblin set has no unique strongest member, which many theories of question embedding take to give rise to a *mention some*-reading, predicting the following meaning:

---

[8][2] point out German *so*, which has a somewhat unclear effect that may be seen as lying in this direction.

(i)   Agatha weiß, wer so auf der Feier war.
        Agatha knows who SO at the party was
        'Agatha has a pretty good idea about who was at the party.'

We are not aware of any similar items, and not convinced that this single example isn't to be analysed in some entirely different fashion. Note also that *so* is entirely unrelated to the German existential adverbial quantifier *teilweise* 'partly'.

(21) #Agatha weiß, wer großteils gekommen ist.
      Agatha knows who mostly   came
      'There is a plurality x such that most members of x came and Agatha knows that most
      of x came.'

This is a decidedly odd meaning, which intuitively doesn't conform to the way in which we
normally package information. Thus, while we have no formal constraint to propose that
would rule it out, it seems quite natural that such a statement would tend to be perceived as
infelicitous.[9] In the case of *partly*, we predict that knowledge of the question collapses into
knowledge of the corresponding existential statement, which could plausibly be infelicitous due
to some blocking or manner effect.

    The lack of alternatives to adverbial universal in normal questions also explains why the
homogeneity removers in questions cannot be contrastively stressed:

(22) #Agatha weiß   nicht, wer ALLER gekommen ist, aber sie weiß, dass Miles da   war.
      Agatha knows not   who all   come   is but   she knows that Miles there was
      'Agatha doesn't know about ALL the people who came, but she knows Miles was there.'

Finally, it follows that *aller* and its ilk do not have their homogeneity-removing effect with
rogative verbs, since their homogeneity, if they show any, cannot be due to the trivalence of the
answers. This is correct: (23) still means that Peter has no influence over who comes. It seems
that *aller* is essentially vacuous here (though not unacceptable).

(23) Es hängt   nicht von Peter ab, wer aller kommt.
      it depends not   from Peter PRT who all   comes
      not: 'There is somebody whose coming doesn't depend on Peter.'

3.3   An Odd Prediction?

The analysis just presented actually predicts a stronger meaning for negated sentences with
embedded questions than we have presented at the outset. According to it, (24a) entails not
only (24b), but the stronger (24c).

(24) a. Agatha doesn't know which girls came.
      b. Agatha doesn't know of any of the girls who in fact came that they came.
      c. Agatha doesn't know that any of those girls came who in fact came.

If this prediction is wrong, then one should be able to find a sentence which under some
conditions is predicted to be undefined, but is actually true. It is not clear that such a case can
be identified.

    If it is assumed that *all* the girls came, (25a) is predicted to entail (25b).

(25) a. Agatha doesn't know which girls came.
      b. Agatha doesn't know that any of the girls came.

---

[9]Questions with *mostly* become marginally possible on a reading with higher-order plurals or contextually
partitioned plurals, which is expected on our theory:

(i)   Context: *There are three groups of students. One group consists mostly of women.*
        ??Agatha knows which students are mostly female.
        'Agatha knows which group of students is the one that is mostly women.'

Consequently, (26a) should be undefined in such a situation. And indeed, the sentence would seem to be somewhat unnatural in this case, (26b) being a preferred way of describing the situation.

(26)   a.   Agatha knows that at least some of the girls came, but she doesn't know which ones.
        b.   Agatha knows that at least some of the girls came, but she doesn't know that all of them did.

There is a plausible independent explanation for this: (26) has a stronger presupposition (that all the girls came) than (25), and it is well-known that sentences with stronger presuppositions are preferred as long as those presuppositions are true ([7] and many thereafter). Thus, we cannot take the oddness of (25) to *confirm* our theory of homogeneity in embedded questions. However, it doesn't disconfirm it, either, since the sentence is, after all, odd in the context in question.

Consider, alternatively, a situation in which both Nina and Mary came. Then (27a) cannot be true — it can only be either false or undefined. And indeed, the sentence does seem to have a slightly contradictory flavour about it, unlike the perfectly acceptable (27b).

(27)   a. ??Agatha doesn't know which girls came, but she knows that at least one of Nina and Mary came.
        b.   Agatha doesn't know exactly which girls came, but she knows that at least one of Nina and Mary came.

This is similar to the analogous case with definite plurals.

(28)   a. ??Mr. Benfleet didn't publish the books, but he published the autobiography.
        b.   Mr. Benfleet didn't publish all the books, but he published the autobiography.

However, (27) seems equally odd when it is not presupposed that both Nina and Mary came, although our theory doesn't predict this. The explanation for this oddness effect is therefore likely a different one, potentially to do with a requirement for contrastive stress, so that, again, the predictions of our theory are masked and prevented from being conclusively put to the test.

In the face of this empirical unclarity, we might at least ask if we *could* avoid the strong meaning (24c) and replace it with the weaker meaning in (24b), should we wish to do so. Since the behaviour of embedded questions is, on our account, tied to the behaviour of declarative complements of *know*, this would require that if a definite plural is embedded in a declarative under *know*, the sentence should have a reading on which the definite plural effectively takes distributive scope over *know*.

(29)   Agatha doesn't know that the young people came.
        a. $\lambda w.\neg\forall w' \in \mathrm{Dox}_w(Agatha) : came_{w'}(typ_w))$
              $\vDash$ Agatha doesn't know that any of the y. p. came.
        b. $\lambda w.\mathrm{Dist}(typ_w)(\lambda x.\neg\forall w' \in \mathrm{Dox}_w(Agatha) : came_{w'}(x))$
              $\nvDash$ Agatha doesn't know that any of the y. p. came.
              $\vDash$ For none of the y. p. does Agatha know that they came.

It seems doubtful that this reading actually exist, and indeed definite plurals are known to not normally take distributive inverse scope ([20]):

(30)   Two students read the books.

> **not:** 'For every book, there are two boys who read it.'

Assuming, however, that the reading in question for (29) does exist, in order to harness it for questions, we would need a semantic, not a syntactic way for the plurality to take distributive wide scope, since in the case of the embedded question, there is no constituent that denotes a plurality and could, for example, undergo quantifier raising. The development of a system that allows this, but also allows low-scope readings of pluralities, and perhaps even disallows distributive inverse scope over regular quantifiers, posits a significant technical challenge for which no obvious solution is currently in sight (see [17] for some of the problems involved).

Pending further technical developments, the weaker reading is therefore not accessible for our approach — and, we repeat, it is not even clear that it needs to be accessible.

4   A Non-Alternative

The way we superficially described the data in (11), repeated here as (31), suggests a quite different approach to the one that we have taken. It seems that one could take the complement of *know* to be a complex algebraic object and have *know* be a homogeneous distributive predicate, as per (4), with respect to this object.

(31)   Agatha knows Q.

> **true** iff Agatha knows all the true answers to Q.
> **false** iff Agatha knows none of the true answers to Q.
> **neither** otherwise (i. e. if she knows some, but not all true answers).

And indeed, the use of algebraic ideas in the treatment of embedded questions is not a new idea: it has been used to approach the phenomenon of quantificational variability effects ([4], [13], [3]).

(32)   Agatha mostly knows who was at the party.
       ⇝ For most people who were at the party, Agatha knows they were there.

According to Lahiri, there is an algebra of answers and quantification is over (atomic) true answers.

(33)   Agatha mostly knows who was at the party.
       'For most true p of the form x *was at the party* (x an atom), Agatha knows p.'

If the object of *know* is the plurality of true answers to the question, then this is just analogous to adverbial quantification with definite plurals:[10]

(34)   The books were mostly published. ≈ Most of the books were published.

We will call this approach to homogeneity with embedded questions — which conceives of the argument of *know* as an algebraic object and assumes that homogeneity holds with respect to that algebra, and identifies that algebra with the one used to analyse QVE — as the QVE approach to homogeneity. This approach has indeed been taken in [5].

A number of problems for the QVE approach is posed by the homogeneity-removing adverbials within the question. The first puzzle that these items raise is why they are inside the

---

[10]On Beck & Sharvit's analysis of QVE, the question would instead be viewed as a plurality of subquestions. This difference is immaterial for the present discussion.

question and clearly within the scope of the *wh*-word, while normal adverbial quantifiers are in the matrix clause. On the QVE approach, some compositional mechanism would have to be devised by which question-internal *all* manages to turn the question (or perhaps worse, if one follows Lahiri, an answer) into a universal quantifier over its parts.

This, however, leads to the second puzzle: If question-internal *all* works this way, why does it not have counterparts with different quantificational strength? Why is there no question-internal counterpart to *mostly* and *partly*, as discussed in section 3.2 above?

The third and most severe problem for the QVE approach is that the question-internal homogeneity remover can, in fact, co-occur with a quantificational adverb in the matrix clause. This is very much unlike the case of definite plurals, where an adverbial quantifier cannot associate with a quantificational noun phrase.

(35)   a. *Alle Buben sind großteils gekommen.
              all boys   are mostly   come
         b. Agatha weiß großteils, wer aller auf der Feier war.
              Agatha knows mostly   who all at the party was
         c. Wer aller zugelassen wird, hängt   großteils (ausschließlich) von diesem
              who all admitted is   depends mostly (exclusively)   on this
              Komitee   ab.
              committee PRT

We take these facts to indicate that, tempting as it may be, quantificational variability effects and homogeneity with embedded questions should not be unified.

5   Conclusion and Outlook

We have argued that once the homogeneity property of plural predication in natural language is viewed as a phenomenon of logical trivalence, homogeneity effects with embedded questions follow immediately due to the fact that the answers to *wh*-questions are trivalent propositions (when pluralities are involved). This has the additional benefit of explaining how adverbial universal quantifiers *within* the question can remove the homogeneity effect: they simply turn the answers into bivalent propositions, in the same way that they do when inserted into declarative clauses. A number of further properties of these adverbial homogeneity-removers within questions can also be made sense of. It is this feature which makes our approach superior to one which attempts to unify homogeneity and quantificational variability effects in embedded questions, as the latter cannot account for these question-internal adverbial quantifiers.

Yet a number of open questions remain. One aspect of adverbial universal is not explained on our theory: Why can they not occur inside questions in all language, but are often infelicitous, such as in most varieties of English?

(36)   a. The young people will all come.
         b. *Who/Which people will all come?

Furthermore, we identified an empirical unclarity concerning the particular truth and falsity conditions that our theory predicts for sentences with embedded questions, which might turn out to be too strong. Should these predictions turn out to be wrong, a serious technical challenge for our approach would arise.

Finally, it is unclear at this point how our approach to homogeneity effects fares beyond the case of weakly exhaustive readings. It would seem that the present trivalent semantics

would be most easily combined with an approach that derives stronger readings from weakly exhaustive ones through some sort of exhaustification ([8]), but no fully satisfactory account of this kind has been given. The exploration of possible trivalent versions of partition semantics also remains as a task for future research.

References

[1] Pascal Amsili and Claire Beyssade. Obligatory presuppositions in discourse. In Anton Benz, Peter Kuehnlein, and Candace Sidner, editors, *Constraints in Discourse 2*, Pragmatics & Beyond. Benjamins Publishers, Amsterdam & Philadelphia, 2009.

[2] Sigrid Beck and Hotze Rullmann. Degree questions, maximal informativeness, and exhaustivity. In *Proceedings of the Tenth Amsterdam Colloquium*, Amsterdam, 1996. ILLC.

[3] Sigrid Beck and Yael Sharvit. Pluralities of questions. *Journal of Semantics*, 19:105–157, 2002.

[4] Stephen Berman. *On the Semantics and Logical Form of wh-Clauses*. PhD thesis, University of Massachusetts, Amherst, MA, 1991.

[5] Alexandre Cremers. Plurality effects and exhaustive readings of embedded questions. Presented at NYU, May 2015.

[6] Jon Gajewski. *Neg-Raising: Polarity and Presupposition*. PhD thesis, MIT, 2005.

[7] Irene Heim. Artikel und definitheit. In Arnim von Stechow and Dieter Wunderlich, editors, *Semantik: Ein Internationales Handbuch der zeitgenössischen Forschung*, pages 487–535. de Gruyter, Berlin, 1991.

[8] Nathan Klinedinst and Daniel Rothschild. Exhaustivity in questions with non-factives. *Semantics & Pragmatics*, 4(2):1–23, 2011.

[9] Manuel Križ. *Aspects of Homogeneity in the Semantics of Natural Language*. PhD thesis, University of Vienna, 2015.

[10] Manuel Križ. Homogeneity, non-maximality, and 'all'. *Journal of Semantics*, 2015.

[11] Manuel Križ. Question embedding, plurals, and homogeneity. Presentation at University College London, November 2015.

[12] Manuel Križ and Emmanuel Chemla. Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics*, 2015.

[13] Utpal Lahiri. *Questions and Answers in Embedded Contexts*. Oxford University Press, 2002.

[14] Godehard Link. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Baeuerle, C. Schwarze, and Arnim von Stechow, editors, *Meaning, Use and Interpretation of Language*. de Gruyter, 1983.

[15] Sebastian Löbner. Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy*, 23:213–308, 2000.

[16] Giorgio Magri. An account for the homogeneity effects triggered by plural definites and conjunction based on double strengthening. In Salvatore Pistoia Reda, editor, *Pragmatics, Semantics and the Case of Scalar Implicatures*, Palgrave Studies in Pragmatics, Language and Cognition, pages 99–145. Palgrave Macmillan, Houndsmills, 2014.

[17] Viola Schmitt. *More Pluralities*. PhD thesis, University of Vienna, 2012.

[18] Roger Schwarzschild. Plurals, presuppositions and the sources of distributivity. *Natural Language Semantics*, 2(3):201–248, 1994.

[19] Benjamin Spector and Yasutada Sudo. Presupposed ignorance and exhaustification: How scalar implicature and presupposition interact. Manuscript, Institut Jean Nicod and UCL, 2014.

[20] Mark Steedman. *Taking Scope*. MIT Press, 2012.

[21] Yimei Xiang. Complete and true: A uniform analysis for mention-some and mention-all. Presented at the LF Reading Group, MIT, 2014.

Performative uses and the temporal interpretation of modals[*]

Sven Lauer

University of Konstanz, Germany
`sven.lauer@uni-konstanz.de`

Abstract

Expressions with a modal semantics vary with respect to whether they are suited to 'performative uses'. Modals like *must* and *have to* can easily be used to give a command and hence create an obligation, but the same does not seem to be true for predicates like *be obligated to* and *be under an obligation to*. This fact poses a challenge for an otherwise attractive class of analyses that take the 'performative effect' (i.e., the creation of an obligation) to arise pragmatically from a claim made with the usual, descriptive modal semantics. Seeing as the different modals are typically assigned the same truth-conditional content, a pragmatic account predicts that there should not be a difference in the availability of performative uses.

In this paper, I will explore possible avenues for meeting this challenge while preserving the attractive features of a pragmatic account. My starting point will be the observation that certain commonly-made assumptions about temporal interpretation in fact block a pragmatic derivation of the performative effect. Then I will consider how this conclusion can be avoided for modals that have performative uses. We either have to assume that these modals have a more liberal temporal interpretation or that the performative effect arises in a different manner than assumed by existing accounts. Quite independently from the issue of anti-performativity, this paper demonstrates that two seemingly independent phenomena—temporal interpretation and performative uses of modals—are in fact intertwined, hence we can shed light on one by studying the other.

1 Introduction

Modal sentences with *must*, *have to*, *may* and *can* can be used not only to *state* that an obligation or permission is in force, but also to *create* an obligation or permission, provided the speaker has the requisite authority. Thus an utterance of (1a) can count as a command, and bring about an obligation for the child to be home by 7pm, and an utterance of (1b) can constitute the granting of a permission to go out and play. In what follows I will focus on the case of necessity modals, as permission-granting uses raise a range of issues that are orthogonal my present concerns (most notably, Lewis's (1979) 'problem about permission').

(1)   [Parent to child.]
   a.   You have to / must be home by 7pm.
   b.   You may / can go out and play.

There are other expressions with modal meanings that appear to resist such 'performative uses'. While the sentences in (2) can be used to inform the addressee about a pre-existing obligation, they are not well-suited as the initial command that brings the obligation into effect.[1]

---

[1]Some of these sentences, especially (2d), have an irrelevant reading as a verbal (eventive) passive of an explicit performative like *I (hereby) require you to be home by 7pm*, on which they can be used performatively. This reading can be forced by inserting *hereby* (*You are hereby required to be home by 7pm*). I assume that, on this reading, they are semantically equivalent to their active counterparts, and function as spelled out in Condoravdi and Lauer (2011).

(2)   a.   You are obligated to be home by 7pm.
          b.   You are supposed to be home by 7pm.
          c.   You are obliged to be home by 7pm.
          d.   You are required to be home by 7pm.
          e.   You are under an obligation to be home by 7pm.

I take it to be obvious that there is a contrast between the modals in (1) and the expressions in (2), which I am going to refer to as 'anti-performative modals'.[2] For most of this paper, I will talk as if anti-performativity is categorical, *i.e.*, I will take for granted that there are at least some modal expressions that are completely incompatible with performative uses. In Section 3.3.3, I will consider the possibility that anti-performativity is a matter of degree.

If there are modals that lack performative uses, then any satisfactory theory of modal meaning should explain not only why some modals have performative uses, but also why others do not. One class of accounts attributes the performative use to a dynamic 'performative meaning', which directly updates the obligations of the addressee (Lewis 1979, van Rooy 2000). Given that *have to* and *must* also have descriptive, non-performative uses, such accounts essentially assume these modals to be ambiguous. As a consequence, they can easily account for the existence of anti-performative modals by assuming that these lack the performative reading. Besides stipulating an ambiguity, however, such accounts are incompatible with an attractive conception of the form-force mapping, according to which all declarative sentences (including modal sentences), have the same type of denotation and a uniform grammatically-determined dynamic effect on the context.

Another kind of approach to performative uses, first suggested by Kamp (1978) and recently defended by Kaufmann (2012), is compatible with this conception. It takes modals to be unambiguous, and to always have their 'descriptive' meanings, which determine ordinary truth-conditional contents, and are always used to make modal claims. Performative uses are explained by assuming that a speaker who has the requisite authority can create an obligation by claiming that it exists. This pragmatic approach is quite attractive, but the existence of anti-performative modals poses a significant challenge for it. Standardly, the sentences in (2) are assigned the same meanings as the sentences in (1), and consequently utterances of both kinds of sentences result in the same modal claim. But then, on the pragmatic approach, there should be no difference in the availability of performative uses.

In this paper, I will explore possible avenues for reconciling a pragmatic account of performative uses with the existence of anti-performative modals. Ultimately, the pragmatic approach will have to assume that there is *some* difference in meaning between modals that have performative uses and those that do not. The question is what this difference in meaning could be, given that the two classes of modals appear to make equivalent claims on their descriptive uses.

My starting point will be the observation that certain commonly-made assumptions about temporal interpretation in fact block a pragmatic derivation of the performative effect for anti-performative modals (Section 2). Then I consider how the same conclusion can be avoided for modals that *do* have performative uses. We either have to assume that these modals have a more liberal temporal interpretation than anti-performative modals (Section 3.1) or that the performative effect arises in a different manner than assumed by existing accounts (Section 3.3).

---

[2]The issue of anti-performativity was brought to my attention by Igor Yanovich (p.c.), who made me realize that Csipak and Bochnak's (2015) observation that *be supposed to* apparently does not have performative uses generalizes to various other modal expressions.

2   Blocking performative uses via temporal interpretation

Pragmatic accounts of performative uses have the following four features (for the purposes of the present section, it does not matter how the crucial (iii) comes about).

(i) On its performative use, a modal sentence 'Must(p)' denotes the same proposition as on its descriptive use, *viz.*, the proposition that *p* is deontically necessary.

(ii) On a performative use, the speaker claims that this proposition is true (*i.e.*, he commits to the the truth of the proposition and/or proposes to add it to the conversational common ground in the sense of Stalnaker (1978)).

(iii) In so doing, the speaker creates an obligation for *p*.

(iv) Thereby, the proposition denoted by the modal claim is made true.

In this section, I will argue that, given some plausible assumptions about their temporal interpretation, anti-performative modals will fail (iv), even if (i)–(iii) are true. This arguably is sufficient to explain why anti-performative modals do not have performative uses.

For the sake of concreteness, I use the following formal set-up, though the argument made here is largely independent from the technical details. I assume that untensed sentences denote properties of intervals, which are convex sets of (linearly ordered) moments in time. The interpretation of (deontic) modals is stated in terms of a modality $\Box^d$, indexed to moments. $\Box^d_m(\phi)$ hence is to be read as 'at moment *m*, ϕ is obligatory'.[3] I also assume an ontology that contains *events* as concrete particulars, along with a function τ from events to intervals that maps any event to its run-time. Furthermore, I will assume that the domain of events consists not only of the eventualities talked about, but also contains *utterance events*. For reasons that will become transparent shortly, it will be convenient to assume that the interpretation function $[\![\cdot]\!]^u$ receives the current utterance *u* as a parameter.

With these preliminaries out of the way, we can state the six assumptions that jointly block the performative use for anti-performative modals. While not universally accepted, assumptions I–V all have some independent motivation, assumption VI captures the essence of the idea that performative uses are truly performative.

Assumption I: Anti-performative modals are stative predicates.   That is, I take *be obligated to* to be of the same aspectual type as *be asleep*. For present purposes, it is immaterial how states are represented. For simplicity, I will describe the interpretation of modal statives directly in terms of quantification over moments.

Assumption II: A stative predicate is true at an interval *i* if the state holds throughout *i*.   While certainly not uncontroversial, this assumption is familiar from the literature (*e.g.* Bennet and Partee 1972, Taylor 1977, Partee 1984, Dowty 1986, Ogihara 2007). For our modal statives, it amounts to the following (here and throughout, I represent the prejacent of the modal as a propositional atom, as its interpretation does not concern us):

(3) $[\![$ You be obligated to be home by seven $]\!]^u = \lambda i \, . \, \forall m \in i : \Box^d_m(\text{you-home-by-7})$

---

[3]The *d* in $\Box^d$ serves as reminder that the necessity is construed deontically. How obligatoriness is cashed out is irrelevant for present concerns. For concreteness, the reader may assume that $\Box^d_m$ universally quantifies over the set of worlds that are deontically ideal at moment *m*.

Assumption III: A matrix present tense stative sentence requires that the stative predicate is true at the speech time s^*.   Consequently:

(4) $[\![$You are obligated to be home by seven$]\!]^u = 1$ iff $\forall m \in s^* : \Box^d_m(\text{you-home-by-7})$

where s^* is the speech time of u

Assumption IV: For any given utterance $u : s^* = \tau(u)$.   For present purposes, we need to be specific as to which interval, exactly, the 'speech time' is supposed to be. An obvious idea is that s^* is just the run-time of the current utterance event. This is indeed what Ogihara (2007) assumes, and we adopt this assumption here. Hence:

(5) $[\![$You are obligated to be home by seven$]\!]^u = 1$ iff $\forall m \in \tau(u) : \Box^d_m(\text{you-home-by-7})$

Assumption V: If a state of affairs s is a result of an event e, then s will not obtain before the final moment of $\tau(e)$.   That is, if some state of affairs is the result of an event, the state will not obtain before the event is completed. Consequently, a resultant state will either temporally abut the run-time of the event, or overlap it in the final moment.

Assumption VI: The obligation is created *as a result of* the utterance of the modal sentence.

Assumptions V and VI directly imply that, even if an utterance of *you are obligated to be home by seven* were to create the obligation in question, the proposition in (5) would still be false, since for any non-final moment m' in $\tau(u)$ it is not the case that $\Box^d_{m'}(\text{you-home-by-7})$—the obligation comes into existence 'too late' to make the denoted proposition true.

Given assumptions I–VI, then, anti-performative modals are predicted to lack performative uses. It is noteworthy that while the assumptions are arguably all plausible, they are not without alternatives.[4] The foregoing should be seen as a proof-of-concept—it shows that there are plausible assumptions which would block the performative use for anti-performative modals on a pragmatic account.

3   Performative modals revisited

Assumptions II–VI jointly imply that no use of a present tense stative sentence can be self-verifying. This turns the table on the pragmatic approach: It is no longer a puzzle why anti-performative modals do not have a performative use, but rather why other modals do. To account for their performative uses, we must either assume that modals like *must* and *have to* are not stative, or we must give up the assumption that performative uses of modals make the proposition they assert true.

3.1   Performative modals as non-stative

The pragmatic account is home free if we assume that modals that can be used performatively have a more liberal temporal interpretation than anti-performative modals. Concretely, suppose *must* and *have to* only require that the state of obligation *overlap* with the time of evaluation, as in (6). Then the considerations in Section 2 do not apply, and claims made with these modals can be self-verifying, provided the necessity comes about in the final moment of $\tau(u)$.

---

[4]In particular, assumption II has a salient alternative: Instead of requiring that a stative predicate hold *throughout* the interval of evaluation, we may require that it the state only temporally *overlap* with the interval of evaluation (Kamp and Reyle 1993, Stechow 1995, Condoravdi 2002). The argument in this section would obviously not go through in this case—unless assumption VI is strengthened to enforce non-overlap of events and their result states.

(6) $[\![\text{You have to be home by 7pm}]\!]^u = 1$ iff $\exists m \in \tau(u) : \square_m^d(\text{you-home-by-7})$

This idea is lent some plausibility by the observation that there is a syntactic difference between the anti-performative modals we have seen, and modals like *must* and *have to*: The former involve copular constructions, the latter do not. It is hence tempting to assume this syntactic difference implies a difference in semantic types—say, true statives denote properties of Davidsonian eventualities, while proper modals denote properties of intervals directly, as in Condoravdi (2002). This difference in semantic types could then be responsible for the difference in temporal interpretation.

On this view, adjectival modals like *be obligated to* lack performative uses because they are true statives, whose temporal interpretation precludes self-verification. This would solve the problem posed by the anti-performativity of such predicates in a rather elegant manner. However, unless we find independent evidence from descriptive uses for the hypothesized difference in temporal interpretation, it remains somewhat stipulative. Unfortunately, I have not been able to find any such evidence. This lack of independent motivation does not show that there is no difference in temporal interpretation, but it should give the defender of the pragmatic account pause.[5]

Potentially even more problematic is the fact that it is not clear that the class of anti-performative modals coincides with the class of adjectival modal expressions. Csipak and Bochnak (2015) claim that the German modal *sollen* does not have performative uses.[6] Similarly, some adjectival modal predicates, like *be allowed to* and *be permitted to*, seem to have performative uses, *cf.* (7).

(7)   You are now allowed to enter the building.

3.2   Performative uses as non-performative

Given these doubts, let us assume that the temporal interpretation of all modals is the same (even if they have different semantic types), and that self-verification is blocked for all of them via the reasoning in Section 2. Is there any way to account for the performative uses of *must* and *have to*, without assuming that they are ambiguous between a descriptive and a performative meaning?

Before giving my own answer, I will briefly discuss a possibility which I do not think is viable.[7] It will be useful to compare and contrast it with the proposal I spell out in the following section. The idea is the following: Maybe it is wrong to assume that it is the utterance itself that creates the obligation. Instead, it is created by a mental act of the agent who has deontic authority; what brings the obligation into effect is the speaker 'making up his mind' as to what the addressee should do. Then, the putatively-performative modal utterance simply reports the existence of the newly-created obligation. In fact, this idea seems to be implicit in the version of the pragmatic account that Kamp (1978) considered (and criticized). Here is his sketch:

> Suppose that *A* has authority over *B* and that this fact is common knowledge shared between *A* and *B*. Then *B* may be expected to react to *A*'s utterance: 'You may take an apple' with the reflection: 'It is up to *A* whether I may take an apple or not. Therefore he knows whether what he says is true or false. It may be assumed moreover that he is not saying

---

[5]It certainly is not surprising that no such evidence can be found in the present tense—even if it is semantically possible to report an obligation with a modal sentence at the very same time at which it comes into effect, there would be very few occasions where speakers would plausibly do so. But given a number of non-trivial assumptions, we would expect to find evidence of the hypothesized difference in past tense sentences with *had to* and *was obligated to* (anchored to a particular evaluation time by means of a *when*-clause, say). I leave this issue for future research.

[6]They attribute the absence of performative uses to a lexical idiosyncrasy of *sollen*, *viz.* an evidential meaning component that requires that the modal claim be based on a prior utterance event. I do not have the space to discuss this proposal in detail here, but if it turns out to be correct, *sollen*'s incompatibility with performative uses would not be an argument against the line pursued in this section.

[7]Igor Yanovich suggested this possibility to me in conversation (without endorsing it).

what he knows to be false, as this would go against established principles of conversational propriety. So I may conclude that I have the permission to take an apple.'

<div align="right">(Kamp 1978, p. 275)</div>

On this view, performative uses are not truly performative—they are regular, descriptive uses. This avoids the 'timing problem' from Section 2, since the speaker will have made up his mind prior to making his utterance, and hence the obligation will be in effect throughout $\tau(u)$. However, it should be clear that, were we to adopt this view, the problem posed by anti-performative modals is back (with a vengeance). That is, it is once again a mystery why expressions like *be obligated to* cannot be used to report mentally-created obligations, while modals like *have to* can.

Moreover, the suggested account is otherwise quite problematic. In particular, it falls prey to a challenge raised by Hans Kamp:

> The problem with this explanation is that it doesn't go quite far enough. [. . .] Suppose *A* says to *B* 'You may take an apple', *B* then takes an apple, whereupon *A* berates him for doing so, claiming he had no permission to take an apple. In such a situation it is not just that *B* can excuse himself by pointing out that he was *misled* by *A*'s utterance. No, *B* can justly claim that he *had* the permission, in virtue of what *A* said to him. There are situations where *A* just cannot mislead *B* simply because his utterance constitutes the granting of the permission.   (Kamp 1978, p. 275)

Kamp's objection is that an account like the one considered here fails to predict that a performative modal utterance is *sufficient* for creating an obligation or permission. We may add that it also fails to account for the fact that, intuitively, the utterance is *necessary* for creating the obligation or permission: Suppose *A* and *B* are talking on the phone, *A* has authority over *B*. *A* makes up her mind that *B* should do something. She draws her breath, ready to utter the corresponding modal sentence. In that moment, the connection is cut off, which *A* notices, hence she does not say anything. On the account considered in this section, we are forced to say that *B* has an obligation in this case (he just does not know it). But it seems intuitively much more appropriate to say that *A* did not get a chance to impose the obligation, and hence that no obligation is in effect.

In summary, assuming that allegedly-performative uses of modal sentences are really descriptive uses that report a mentally-created obligation not only reinstates the problematic prediction of the pragmatic account we started out with, it also fails to accord with the intuition that, under appropriate circumstances, the utterance itself is both necessary and sufficient for creating the obligation or permission in question. I hence conclude that this view is not viable.

3.3   Performative uses as claims about preferences

We can have our cake and eat it, too. We can accept that modal sentences can never be self-verifying, but hold on to the idea that performative uses are truly performative, in the sense that it is the utterance itself that creates the obligation or permission. And we can do so in a way that predicts that anti-performative modals do not have performative uses.

What we have to do is to give up the assumption that performatively-used modals denote propositions about what is deontically required. I want to suggest that, on their performative uses, these modals instead express propositions about what is necessary, given the preferences of the speaker (*i.e.*, they have a broadly-speaking bouletic interpretation). We already know that assertions about speaker preferences can be 'performative', in the sense that they can count as commands and create obligations. Consider (8), after Condoravdi and Lauer (2009).

(8)   [Parent to child] I want you to be home by 7pm.

By virtue of the use of the predicate *want*, (8) is a claim about the speaker's preferences, and yet, in the right context, an utterance of the sentence can create the same obligation that a performative use of a modal would create. Taking this observation as a starting point, I propose the following conception of performative uses: If a speaker is a deontic authority, then whenever a speaker has a public preference for *p*, *p* is obligatory. With a performatively-used modal, the speaker publicizes a preference, thereby creating an obligation. The rest of this paper spells this idea out using the apparatus of the *dynamic pragmatics* of Lauer (2013), in a way that meets Kamp's challenge.

3.3.1   Preliminaries

I can only sketch the required formal set-up here, and refer the reader to Condoravdi and Lauer (2011), Lauer (2013) for the details. First, besides their deontic interpretation $\Box^d$, I assume that modals have a preference-related interpretation $\Box^{Sp}$. '$\Box^{Sp}_m(\phi)$' represents that ϕ is necessarily true if the speaker's preferences are optimally satisfied. This does not require us to assume that modals are lexically ambiguous, instead we can take modals to be underspecified in the style of Kratzer (1981).[8]

Second, I assume that our models represent the *commitments* of interlocutors, which come in two kinds. *Doxastic commitments* are commitments to treat a proposition as true and are represented via an operator PB_m (for 'public belief'). *Preferential commitments* are commitments to treat a proposition as desirable and are represented via an operator PEP_m (for 'public effective preference').

Third, I assume that these commitments are subject to a number of consistency constraints, which can be implemented as constraints on admissible models (*cf.* Lauer (2013, Ch. 5.3.2)). Only two of these will be relevant here:

(9)   If ϕ entails ψ, then for all a, m :

$$\text{PB}_m(a, \phi) \supset \text{PB}_m(a, \psi)$$

(10)   For all moments m, agents a and propositions φ:

$$\text{PB}_m(a, \Box^a_m \varphi) \supset \text{PEP}_m(a, \varphi)$$

(9) requires that doxastic commitment is closed under entailment. (10) requires that if an agent is committed to believe that φ is necessary for optimally realizing her preferences, she is also committed to treat φ as desirable.[9]

Finally, I assume the following Declarative Convention (*cf.* Condoravdi and Lauer 2011), intended to model the conventionally-determined dynamic effect of declarative sentences. As in Lauer (2013, Ch. 5.4), such conventions can likewise be implemented as constraints on admissible models. Here I state it informally.

(11)   Declarative Convention
         If a speaker Sp makes an utterance u of a declarative sentence with content ϕ he incurs the following commitment (where m is the final moment of $\tau(u)$):

$$\text{PB}_m(Sp, \phi)$$

---

[8] See Condoravdi and Lauer (2015) for a Kratzer-style implementation of the $\Box^{Sp}$-reading. As we do there, I furthermore assume that the not *all* the speaker's preferences are taken into account, but only her *effective* preferences, in the sense of Condoravdi and Lauer (2011, 2012), Lauer (2013), Condoravdi and Lauer (2015). What matters for present purposes is that these are preferences that 'win out' against any conflicting desires the agent may have.

[9] In Lauer (2013, Ch. 6.3, p. 158) I consider a similar, but weaker consistency constraint, which requires only that a speaker who is committed to believe that φ is one of his (basic) effective preferences be also committed to prefer φ. (11) is stronger in that it also requires that a speaker is committed to such a preference if he is committed to believe that φ merely is true whenever his basic effective preferences are optimally realized. This would be true, *e.g.*, if φ is merely a precondition for something the agent prefers.

3.3.2 Deriving the performative effect

Suppose now that Sp makes an utterance u of (12), on its preference reading, and let n be the final moment of $\tau(u)$. By (11), this will result in the commitment in (13).

(12)   You have to be home by 7pm.

(13) $\mathrm{PB}_n(Sp, \forall m \in \tau(u) : \Box_m^{Sp}(\mathsf{you\text{-}home\text{-}by\text{-}7}))$

By (9), this means that Sp also has the commitment in (14).

(14) $\mathrm{PB}_n(Sp, \Box_n^{Sp}(\mathsf{you\text{-}home\text{-}by\text{-}7}))$

From this, it follows by (10) that Sp also has the commitment in (15).

(15) $\mathrm{PEP}_n(\mathsf{you\text{-}home\text{-}by\text{-}7})$

That is, if a speaker utters (12), on its preference reading, he thereby becomes publicly committed to a preference for you-home-by-7.

The final ingredient of the account is a suitable conception of what it is to have deontic authority. In Lauer (2013, p. 147), I proposed the following conception (building on Condoravdi and Lauer (2009)). An agent a is a deontic authority with respect to a proposition p if p is obligatory whenever a is publicly committed to prefer p.[10]

(16)   **Deontic authority with respect to p**
         An agent a is a deontic authority with respect to p iff:
         $$\forall m : \mathrm{PEP}_m(a, p) \supset \Box_m^d(p)$$

Now suppose that, in the context of u, the speaker Sp is a deontic authority with respect to you-home-by-7. Then from (15), it follows that (17) is true.

(17) $\Box_n^d(\mathsf{you\text{-}home\text{-}by\text{-}7})$

That is, by uttering *You have to be home by 7pm*, the speaker has created an obligation for the addressee to be home by 7pm. Thus we account for the performative effect. And we do so in a way that meets Kamp's challenge, because according to (16) the existence of the obligation is dependent on the *public commitments* of speaker. Hence, the obligation will come into existence regardless of whether the agent has the preference he professes to have—all that matters is whether he is committed to having it. Furthermore, it is the utterance itself that creates the commitment (and, therewith, the obligation), and it cannot fail to do so, in virtue of the Declarative Convention and the consistency constraints on commitments. The present account hence correctly accounts for the intuition that the utterance of a performatively-used sentence is both sufficient and necessary for the obligation to come about.[11]

---

[10] This conception also can be used to account for command-uses of desideratives like (8), as well as command-uses of imperatives, on either the account defended in Condoravdi and Lauer (2012, on which *Be home by 7pm!* directly induces the commitment in (15)) or a variant of the account of Kaufmann (2012, on which *Be home by 7pm!* denotes the proposition $\Box_n^{Sp}(\mathsf{you\text{-}home\text{-}by\text{-}7})$).

[11] Even though Kamp's challenge is met, we may wonder (as an astute reviewer did) whether the present account can explain the fact that it is odd to challenge a performatively used modal sentence as false.

(i) A : You have to be home by 7pm.
        B : ??That's not true, you don't want me to be home by 7pm.

Nothing that has been said so far excludes such challenges. However, on the present account, B can certainly be said to be 'missing the point'. Even if A's claim were in fact false, he still would have created an obligation, and his utterance would still constitute a command, which presumably this was A's intention. But this would be irrational if A did not in fact have the effective preference he claims to have. Perhaps this is enough to explain why B's challenge seems infelicitous. I leave this issue for future work.

3.3.3 Anti-performative modals

On the proposed account, modal expressions like *be under an obligation to* are predicted to lack performative uses because they are lexically constrained to express deontic necessities. That is, unlike (12), (18) only has the reading in (18a), but not the one in (18b). Since performative uses proceed via an assertion with content (18b), no performative use is predicted for (18).[12] Anti-performativity is due, then, to lexical constraints on the kind of necessity a modal is compatible with. We hence do not expect that there is a clean separation between, say, adjectival modal expressions and others.[13]

(18)   You are under an obligation to be home by 7pm.

    a.   can mean: $\forall m \in \tau(u) : \Box_m^d(\text{you-home-by-7})$

    b.   cannot mean: $\forall m \in \tau(u) : \Box_m^{Sp}(\text{you-home-by-7})$

The proposal is also compatible with the possibility that anti-performativity is not always categorical. For *be under an obligation to*, performative uses seem to be categorically ruled out, but it has been suggested to me (by Tom Wasow, Chris Potts and Dan Lassiter) that things may be different for other predicates, especially *be required to*.

(19)   You are required to be home by 7pm.

(19) is at least somewhat resistant to performative uses, but some speakers seem to be able to imagine such a use. We can make sense on this by assuming that, in addition to (or instead of) categorical lexical constraints on modal backgrounds, speakers have gradable preferences for particular construals. Perhaps *be required to* is not strictly incompatible with a preference construal, but just strongly biased against it. Given enough contextual pressure, speakers may be able to overcome this bias, resulting in a performative use.

It is noteworthy that even if *all* anti-performative modals are only biased against, rather than incompatible with, performative uses, what has been said in this paper remains relevant. For on the unamended Kamp/Kaufmann approach, graded anti-performativity is just as unexpected as categorical anti-performativity—the anti-performative modals all are either categorically restricted to, or heavily biased in favor of, deontic construals. If performative uses proceeded via a deontic modal claim, modals like *be obligated to* should not resist performative uses, but be biased in favor of them.

4   Conclusion

If a pragmatic account of performative uses is to be reconciled with the existence of anti-performative modals, such modals must differ semantically from the ones that have performative uses. This paper has investigated two possibilities for such differences in meaning: Either the two classes of modals differ in their temporal interpretation, or they differ with respect to the kinds of modal backgrounds they are compatible with. The latter necessitates significant revisions to the pragmatic account. Quite independently from the issue of anti-performativity, this paper has demonstrated that two seemingly independent phenomena—temporal interpretation and performative uses of modals—are in fact intertwined, hence we can shed light on one by studying the other.

---

[12]In principle, this explanation for the absence of a performative effect for (17) would be available even if we don't make the assumptions in Section 2. Except we would have to worry, in that case, whether there aren't *two* routes to the performative effect, one via an assertion with content (18b) and one (à la Kamp/Kaufmann) via an assertion with content (18a), which would again predict a performative use for (18). Adopting the assumptions in Section 2 precludes this possibility.

[13]But we may expect tendencies. As Hacquard (2013) points out, 'lexical' modals (which include the adjectival ones) tend to be more constrained with respect to the kinds of modal backgrounds they combine with then 'grammatical' ones like *must*.

References

Bennet, M. and Partee, B.: 1972, Toward the logic of tense and aspect in english, *Indiana University Linguistics Club*, Bloomington, Indiana.

Condoravdi, C.: 2002, Temporal interpretation of modals: Modals for the present and for the past, *in* D. Beaver, L. C. Martínez, B. Clark and S. Kaufmann (eds), *The Construction of Meaning*, CSLI Publications, Stanford, CA, pp. 59–88.

Condoravdi, C. and Lauer, S.: 2009, Performing a wish: Desiderative assertions and performativity. Talk at California Universities Semantics and Pragmatics (CUSP) 2, UC Santa Cruz, November 2009.

Condoravdi, C. and Lauer, S.: 2011, Performative verbs and performative acts, *in* I. Reich, E. Horch and D. Pauly (eds), *Sinn and Bedeutung 15*, Universaar – Saarland University Press, Saarbrücken, pp. 149–164.

Condoravdi, C. and Lauer, S.: 2012, Imperatives: Meaning and illocutionary force, *in* C. Piñón (ed.), *Empirical Issues in Syntax and Semantics 9*, pp. 37–58.

Condoravdi, C. and Lauer, S.: 2015, Anankastic conditionals are just conditionals. To appear in *Semantics & Pragmatics*.

Csipak, E. and Bochnak, R.: 2015, The semantics of 'supposed to' as a reportative evidential. Talk at the LSA Annual meeting, Portland, OR.

Dowty, D.: 1986, The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics?, *Linguistics and Philosophy* **9**, 37–61.

Hacquard, V.: 2013, The grammatical category of modality, *in* M. Aloni, M. Franke and F. Roelofsen (eds), *Proceedings of the Nineteenth Amsterdam Colloquium*, pp. 19–26.

Kamp, H.: 1978, Semantics versus pragmatics, *in* F. Guenthner and S. J. Schmidt (eds), *Formal Semantics and Pragmatics for Natural Languages*, Reidel, Dordrecht, pp. 255–287.

Kamp, H. and Reyle, U.: 1993, *From Discourse to Logic*, Reidel, Dordrecht.

Kaufmann, M.: 2012, *Interpreting Imperatives*, Springer, Dordrecht/New York.

Kratzer, A.: 1981, The notional category of modality, *in* H.-J. Eikmeyer and H. Rieser (eds), *Words, Worlds, and Contexts. New Approaches in Word Semantics*, de Gruyter, Berlin, pp. 38–74.

Lauer, S.: 2013, *Towards a dynamic pragmatics*, PhD thesis, Stanford University.

Lewis, D.: 1979, A problem about permission, *in* E. Saarinen, R. Hilpinen, I. Niiniluoto and M. Provence (eds), *Essays in Honor of Jaakko Hintikka*, Reidel, Dordrecht, pp. 163–175.

Ogihara, T.: 2007, Tense and aspect in truth-conditional semantics, *Lingua* **117**, 392–418.

Partee, B.: 1984, Nominal and temporal anaphora, *Linguistics and Philosophy* **7**(3), 243–286.

Stalnaker, R.: 1978, Assertion, *in* P. Cole (ed.), *Syntax and Semantics 9: Pragmatics*, New York Academic Press, New York, pp. 315—332.

Stechow, A. V.: 1995, On the proper treatment of tense, *Semantics and Linguistic Theory (SALT) 5*.

Taylor, B.: 1977, Tense and continuity, *Linguistics and Philosophy* **1**, 199–220.

van Rooy, R.: 2000, Permission to change, *Journal of Semantics* **17**, 119–145.

Definiteness: from L1 Mandarin to Mandarin L2 English*

Bert Le Bruyn and Xiaoli Dong

Utrecht Institute of Linguistics, Utrecht University, The Netherlands
`b.s.w.lebruyn@uu.nl, x.dong@uu.nl`

1 Introduction

In this paper we present a preview of our study on the L2 acquisition of English definites by L1 Mandarin speakers, highlighting those aspects that are relevant to the current theoretical debate on definiteness. Section 2 zooms in on the learners' L1 and argues that – next to standard definite uses of bare nouns – it also has a construction that marks non-maximal familiar definiteness. Section 3 zooms in on the learners' L2, argues that they don't confuse definiteness/specificity and locates the main acquisition problem in the constraints that govern accommodation.

Our pretheoretical working definitions are as follows: we assume maximality is to be understood as exhaustivity at the level of the model and that familiarity implies previous mention in the discourse. Both the model and the discourse can be contextually restricted (see e.g. Brisson, 1998; Walker et al., 1998).

2 Mandarin

In the recent literature on cross-linguistic variation in the domain of definiteness, two papers stand out, *viz.* Schwarz (2009) and Arkoh and Matthewson (2013). Both argue that there are languages with a definite article that marks non-maximal familiar definiteness. Continuing the work of Yang (2005) and Partee (2006), we argue that Mandarin is relevant in this discussion as well. Next to the definiteness conveyed by bare nouns (see e.g. Cheng and Sybesma, 1999, 2012), it also has a construction that marks non-maximal familiar definiteness. The study of non-maximal familiar expressions is important for familiarity- as well as for uniqueness-based theories of definiteness.

The construction we are referring to consists of the sequence *modifier + numeral + classifier + noun*, henceforth *MNCN*. The modifier can be – among other things – a possessive or a relative clause:

(1) Zhangsan de san   ben shu
     John   DE three CL   book
     'John's three books'

(2) wan tiaosan   de liang ge xuesheng
     play parachute DE two   CL student
     'the two skydiving students'

Previous analyses have qualified the MNCN as specific, referential, not non-definite (Huang, 1982), familiar, non-maximal definite (Yang, 2005; Partee, 2006), specific indefinite (Sio, 2006; Zhang, 2006), and familiar, maximal definite (Hall, 2015). We ran two experiments to assess the exact nature of the construction.

2.1 Experiment 1: familiarity

The aim of the first experiment was to establish under which circumstances the use of the MNCN would be licit. We distinguished between three conditions: one in which the speaker could not identify the referent, one in which the speaker could identify the referent and one in which the speaker referred to a previously introduced referent.

The different items were all integrated into one overarching story about a guard in an elite sports school in China. The school had 6 students and both the girls and the boys could easily be distinguished during the day but happened to have the same silhouettes, thus making identification during the night impossible. The day vs. night setting was the crucial difference between the first two conditions.

(3)   The students

Participants (n=15) were asked to play the guard who reported on what he had seen to the principal. Their input was a visualization of what the guard had seen and they had to select a description with an MNCN or one in which the modifier was put between the classifier and the noun. These descriptions were put into text balloons in pictures representing the guard reporting to the principal.

The third condition involved objects that had been stolen from the principal's office and that the principal had asked the guard to locate (e.g. sports trophies). Given the importance of these objects, the principal had asked the guard to call him as soon as he found them. The MNCN and non-MNCN reports of the discovery of the stolen items were put into text balloons in pictures representing the guard calling the principal.[1]

(4) to (6) give an impression of each of the respective conditions. In (4) we also provide a written out version of the text balloons, a gloss and a translation.

---

[1]We also designed a version of this experiment in which the third condition involved students instead of objects. The results were comparable but the overarching story was slightly more elaborate. For reasons of space we only report on the experiment with objects.

(4) Non-identifiable referents
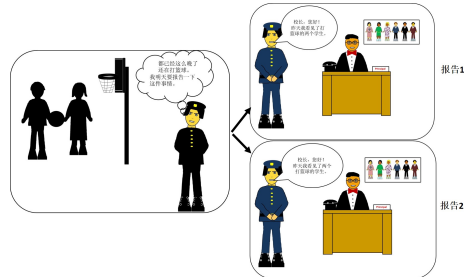
    a. Dou yijing   zheme wan le,   hai zai da   lanqiu,   wo mingtian yao baogao
       PRT already so   late PRT, still PRT play basketball, I   tomorrow will report
       yixia   zhe jian shiqing.
       quickly this CL   affair
       'It's already so late and they're still playing basketball, I'll briefly report about this
       affair tomorrow.'

    b. Xiaozhang, ninhao! Zuotian   wo kanjian le da   lanqiu   de liang ge xuesheng.
       Principal, hello! Yesterday I   see   LE play basketball DE two CL student
       'Principal, hello! Yesterday I saw two students playing basketball.'

    c. Xiaozhang, ninhao! Zuotian   wo kanjian le liang ge da   lanqiu   de xuesheng.
       Principal, hello! Yesterday I   see   LE two CL play basketball DE student
       'Principal, hello! Yesterday I saw two students playing basketball.'

(5) Identifiable referents
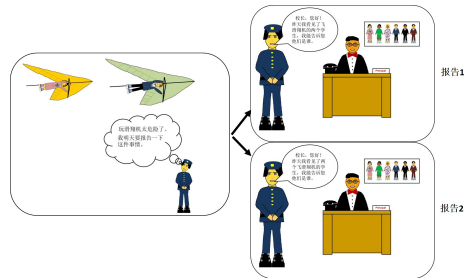
(6) Previously introduced referents
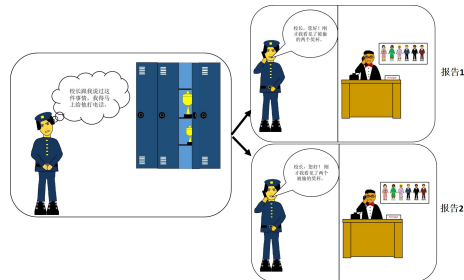
We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the familiar condition was the only one that was significantly different from the other two in allowing the use of the MNCN (familiar/simple indefinite: t(177)=5.278; p<.000 | familiar/ specific indefinite t(177)=4.987; p<.000). The interpretation we give of these results is that familiarity is a necessary condition for using the MNCN in Mandarin.

2.2   Experiment 2: maximality

The second experiment assessed maximality. The rationale behind it was that definites that allow for maximal readings will favor those in out of the blue contexts, thus making sentences like (7) sound as inconsistent as sentences like (8), involving *all*:

(7) ${}^?$I stole the two pens of Mary's, so now she only has one left.

(8) ${}^?$I stole all of Jacky's books, so now she only has one left.

An experiment with English native speakers (n=15) recruited via Crowdflower confirmed that (7) and (8) are not significantly different from each other and that both differ from sentences with indefinites as in (9):

(9)   Harry smoked 5 cigarettes, so now he only has 3 left.

Examples of the Mandarin items corresponding to (7) and (8) are given in (10) and (11), (10) involving the MNCN as the counterpart of the English definite, and (11) containing *suoyou* as the counterpart of English *all*.

(10)   Wo chi le Xiaowang de san   ge jidan,   suoyi ta xianzai zhiyou san ge le.
        I   eat LE Xiaowang DE three CL chicken egg, so   he   now   only CL LE

(11)   Xiaoding chi le suoyou de dangao, suoyi xianzai zhiyou san   ge dangao le.
        Xiaoding eat LE all   DE cake,   so   now   only   three CL cake   LE

Next to items like (10) that have a non-relational noun (*egg*) and a possessor as modifier (*Xiaowang de*), we also included MNCNs with relational nouns and relative clauses as modifiers, giving rise to three subtypes of MNCN conditions (possessor modifier + relational noun; possessor modifier + non-relational noun; relative clause modifier + non-relational noun).

Participants (n=17) were asked to assess the consistency of the sentences we presented them with, focusing on the link between the first and the second part. They received three answer options (compatible, possibly compatible and incompatible). In our analysis we reduced these to two options (compatible/possibly compatible vs. incompatible).

We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that all MNCN conditions were judged significantly more consistent than the all condition (t(183)=6.615 | 6.801 | 6.897; p<.000) but did not differ from each other. The interpretation we give of these results is that maximality is not part of the interpretation of the MNCN in Mandarin.

2.3   The MNCN: conclusion

Based on the two experiments, we conclude that familiarity is a necessary condition for the use of the MNCN and that maximality is not part of its interpretation. The only analysis in the literature that is compatible with these data is that of Yang (2005) and Partee (2006), who assume the MNCN expresses non-maximal familiar definiteness.

3   Mandarin L2 English

In the past decade the literature on the L2 acquisition of definiteness by L1 speakers of articleless languages (like Korean, Russian, Mandarin, Polish, etc.) has almost exclusively focused on the complicating role of specificity in this process. The consensus seems to be that L1 speakers of articleless languages – independently of proficiency – often have a hard time deciding what the exact semantics of the definite article is, fluctuating between definiteness and specificity. We argue that specificity is not at play (3.1) and that the main acquisition problem lies in the acquisition of maximality (3.2).

3.1   Assessing the influence of specificity

In order to assess the influence of specificity on the L2 acquisition of definites, we designed a new experimental paradigm. The standard paradigm used in the literature is that of Ionin et al. (2004) but has evolved over the years (compare e.g. Ionin et al. (2004) and Ko et al. (2010)) and the construct *specificity* in the experiment has never been fully validated (see Ionin (2006) for discussion).[2]

Ionin (2006) takes the type of specificity that comes into play in the L2 acquisition of definiteness to be identical to the one expressed by indefinite *this* in English:

(12)   If you did get in any trouble, there was **this man**, Joe, and he was a really big man, he was a black-belt in karate and everything, and . . .   (example taken from the BNC)

Ionin describes the use of *this* in examples like (12) as signaling that the speaker intends to refer to exactly one individual who is moreover noteworthy in some way or other. We operationalized these intuitions in various ways and ran a series of experiments with native speakers of English we recruited online via Crowdflower. The crucial criterion for success was a significant difference between non-specific and specific contexts in welcoming the use of *this* rather than *a*. We succeeded in meeting this criterion by playing with a foreground/background distinction as in (13) and (14) where indefinite *this* was judged more appropriate for the foregrounded referent of *this boy* in (13) than for the backgrounded referent of *this brick* in (14).

(13)   Have I already told you my favorite soccer story? Well, one day I saw **this boy** playing soccer in our street... All of a sudden he jumps up, makes an amazing salto and makes the ball hit a can. It turned out he was the son of a famous soccer player.

(14)   Have I already told you about the scariest moment of my life? Well, one day I saw a girl on top of a building... All of a sudden, she starts to dance, slips on **this brick** and falls off the building! Fortunately she landed on some cardboard boxes and didnt get hurt...

The specific question we asked the participants (n=15) was whether the use of *this* was felicitous in the given contexts or whether they would prefer to replace it by *a*.

We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the foregrounded uses of this were significantly more felicitous than the backgrounded uses (t(118)=2.326; p=.022).

With a successful operationalization of specificity in place, we set up an L2 experiment in which we included non-specific and specific indefinite conditions as well as a familiar definite condition. In each item we presented two versions of a small story, one with an indefinite article and one with a definite article. We asked Mandarin L2 learners of English (n=19)

---

[2]See the end of section 3.2 for some comparison between the two paradigms.

recruited through Crowdflower to choose the story that sounded most natural to them.[3] To help participants, we visualized the discourse setting of the story (the story was told by one guy to another guy in a pub) as well as the story itself. (15) to (17) illustrate the different conditions:

(15) Non-specific indefinite

(16) Specific indefinite

(17)   (Familiar) definite

If specificity were to play the role it has been claimed to play in the literature, we would expect the definite condition and the specific indefinite condition to be eliciting significantly more definites than the non-specific indefinite condition. To check this prediction, we ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the non-specific indefinite condition elicited significantly more definites than the specific indefinite condition (t(117)=3.096; p=.002) and that the definite condition still elicited significantly more definites than the non-specific indefinite condition (t(117)=6.152; p<.000). This effect is not the one we would expect if Mandarin L2 English learners were to fluctuate between a specific and a definite semantics for the definite article. It however receives a straightforward – be it *post hoc* – explanation if we assume Mandarin L2 English learners assume the definite article marks familiarity. The higher rate of definites in the non-specific condition then follows from the well-known connection between backgrounded and given information.

3.2   The acquisition of maximality

In 3.1 we established that specificity is not interfering with the acquisition of definiteness. We however know that Mandarin L2 English learners do overgenerate definites (see e.g. Tryzna, 2009; Yang and Ionin, 2009b, etc.). We propose an explanation of these data couched in theories of definiteness.[4]

We assume Mandarin L2 English learners start with an interlanguage grammar in which definites are familiar but not maximal. Support for this comes from three experiments. The first is the one we presented in 3.1, the other two are briefly summarized here:

- An exploratory production experiment (a culturally modified version of the paradigm used by Schaeffer and Matthewson 2005):

   We tested 2nd grade high school students from the Zhejiang Ruian High School (n=16) and found that they performed at ceiling for the production of definites in the familiar

---

[4]A comparison with Yang and Ionin (2009a) will appear in Le Bruyn and Dong (prep).

condition. This indicates that Mandarin L2 learners of English associate definites with familiarity.

- A felicity judgement experiment (identical to the one we discussed for native English speakers in section 2.2):

We tested students from the Beijing International Studies University (n=20) and found that they differed from native English speakers in finding sentences like *I stole the two pens of Mary's, so now she only has one left* significantly more consistent than sentences like *I stole all of Jacky's books, so now she only has one left*. At the same time they however also judged sentences with *the* less consistent than sentences like *Harry smoked 5 cigarettes, so now he only has 3 left*.
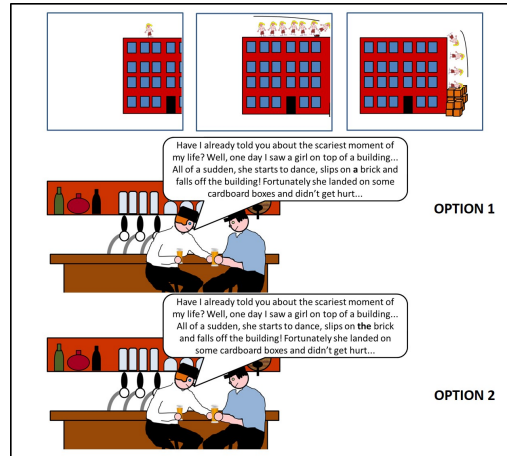
We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the definite condition was significantly different from the *all* condition (t(237)=12.820; p<.000) and from the indefinite condition (t(237)=3.184; p=.002). These data indicate that Mandarin L2 learners of English differ from native speakers in not always associating definites with maximality.

In the process of acquisition, Mandarin L2 English learners are confronted with maximality as another ingredient of definiteness and pick up on this. Support for this comes from one experimental study and one corpus study:

- An exploratory forced choice elicitation task (an experiment including part of the items from Ko et al. 2010):

We tested 2nd grade high school students from the Zhejiang Ruian High School (n=35) and found that they performed almost at ceiling for the production of lexically unique referents (*the capital of North Dakota, the president of our university*) as well as classical bridging definites (*the cover* (of a previously mentioned book), *the screen* (of a previously mentioned laptop)). These data indicate that Mandarin L2 English learners – even though they don't establish a one-to-one correspondence between definites and maximality – are sensitive to maximality for the definite article.

- A study of the Portsmouth learner corpus:

We had two native speakers error tag 500 random occurrences of the definite article in texts produced by Mandarin L2 English learners (kappa: 0.82). The native speakers agreed on 58 problematic uses of *the*, several of which involved overproduction of the definite for kind reference. Given that these cases don't occur in the L2 and cannot be due to L1 transfer either, they are most likely cases in which the maximality semantics of the definite article was (erroneously) generalized to kind reference.

The acquisition problem that presents itself in moving from a familiarity to a familiarity + maximality semantics of the definite article lies in merging a familiarity and a maximality view on definiteness. The contexts that pose most problems are those in which nouns are used that are not lexically unique but come with sufficient modification to warrant a contextually unique interpretation. The felicity of a definite in these contexts depends on the availability of accommodation, a process whose conditions seem to be well understood by native speakers but pose a challenge for semanticists working on definiteness as well as for L2 learners with an articleless L1. An example is given in (18):

(18) Phone conversation

Jennifer:	Hello, Helen? This is Jennifer!
Helen:	Hi, Jennifer! It's wonderful to hear from you. I suppose you want to talk to my sister?
Jennifer:	Yes, I haven't spoken to her in years! I'd like to talk to her now if possible.
Helen:	I'm very sorry, but she doesn't have time to talk right now. She is meeting with **a very important client from Seattle. He is quite rich, and she really wants to get his business for our company**! She'll call you back later.

Chances are real that the company Jennifer is working for doesn't have more than one important client in Seattle. This means that – in principle – a contextually unique interpretation should be viable. The particular way of phrasing chosen in (18) however proscribes accommodation for native speakers and forces the use of *a*. Some slight variations would however felicitously host a definite:

(19) She's meeting with the very important client who arrived from Seattle.
She's meeting with the very important client she met in Seattle.
She's meeting with the very important client from across the street.

The challenge for definiteness theories as well as for L2 learners with an articleless L1 is to master the subtle differences between (18) and (19).

The choice of (18) is not accidental. It's a specific indefinite item from one of the more recent test batteries of Ionin and colleagues (Ko et al., 2010). One of the most striking differences between their specific indefinite items and ours is that we have operationalized specificity by making the corresponding referents into the protagonists of a story that – at least according to the speaker – was worth telling. For Ionin and colleagues, specificity is operationalized as the enumeration of noteworthy properties. Given that these easily give rise to the impression that a contextually unique reading is targeted, it doesn't come as a surprise that they elicit overgeneration of definites by L2 learners with an incomplete mastery of the constraints on accommodation.

4   Conclusion

In this paper, we have given a preview of our work on the L2 acquisition of English definites by L1 Mandarin speakers. We have highlighted those aspects that are of relevance for current discussions about the nature of definiteness, in particular the relation between maximality and familiarity.

References

Arkoh, R. and Matthewson, L. (2013). A familiar definite article in Akan. *Lingua*, 123:1–30.

Brisson, C. (1998). *Distributivity, maximality, and floating quantifiers*. PhD thesis, Rutgers, The State University of New Jersey.

Cheng, L. and Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, 30(4):509–542.

Cheng, L. and Sybesma, R. (2012). Classifiers and DP. *Linguistic Inquiry*, 43(4):634–650.

Hall, D. (2015). Licensing definiteness. Ms. Queen Mary University London.

Huang, C. T. (1982). *Logical relations in Chinese and the theory of grammar*. PhD thesis, MIT.

Ionin, T. (2006). *This* is definitely specific: specificity and definiteness in article systems. *Natural Language Semantics*, 14(2):175–234.

Ionin, T., Ko, H., and Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1):3–69.

Ko, H., Ionin, T., and Wexler, K. (2010). The role of presuppositionality in the second language acquisition of English articles. *Linguistic Inquiry*, 41(2):213–254.

Le Bruyn, B. and Dong, X. (prep). The L2 acquisition of English definites by L1 Mandarin speakers. In preparation. Utrecht institute of Linguistics.

Partee, B. H. (2006). A note on Mandarin possessives, demonstratives, and definiteness. In Birner, B. J. and Ward, G., editors, *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, pages 263–280. Amsterdam: John Benjamins.

Schaeffer, J. and Matthewson, L. (2005). Grammar and pragmatics in the acquisition of article systems. *Natural Language & Linguistic Theory*, 23:53–101.

Schwarz, F. (2009). *Two types of definites in natural language*. PhD thesis, UMass, Amherst.

Sio, J. (2006). *Modification and reference in the Chinese nominal*. PhD thesis, Leiden University.

Tryzna, M. (2009). Questioning the validity of the article choice parameter and the fluctuation hypothesis. In del Pilar García Mayo, M. and Hawkins, R., editors, *Second language acquisition of articles*, pages 67–86. Amsterdam: John Benjamins.

Walker, M., Joshi, A., and Prince, E. (1998). *Centering theory in discourse*. Oxford University Press.

Yang, H. (2005). *Plurality and modification in Mandarin nominal phrases*. PhD thesis, University of Texas at Austin.

Yang, M. and Ionin, T. (2009a). L2 English articles and the computation of uniqueness. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, pages 325–335. Connecticut, US.

Yang, M. and Ionin, T. (2009b). Specificity and speaker knowledge in the second language acquisition of English articles. Ms. Guangdong University of Foreign Studies and University of Illinois at Urbana-Champaign.

Zhang, N. (2006). Representing specificity by the internal order of indefinites. *Linguistics*, 44(1):1–21.

An Inquisitive Semantics Analysis for the Chinese Polar Question Particle *Ma*

Shih-Yueh Lin*

New York University
syjefflin@nyu.edu

Abstract

Inquisitive Semantics provides two operators that serve to modify the inquisitive content of a given proposition. The goal of this paper is to show that the two operators together provide an account for the semantics of the Chinese polar question particle *ma*, arguing that the *ma*-particle is the lexical realization of the combination of non-inquisitive and inquisitive operators. It will also be shown that the proposed analysis yields a novel account for the (in-)compatibility of the *ma*-particle and different types of Chinese questions.

1 Introduction

The syntax and semantics of question formation in *wh*-in-situ languages are often related to sentence-final particles. For example, the Clausal Typing Hypothesis [2] states that while questions are typed by movement of *wh*-phrases in *wh*-movement languages, in a *wh*-in-situ language, questions are clause-typed by question particles. It is also argued that the *ma*-particle in Chinese serves to clause-type polar questions, for instance:

(1) a. John líkaī-le   ma?   b. John kàn-guò shéme rén   ma?
     John leave-ASP MA   John see-ASP   what   person MA
     'Did John leave?'   'Did John see someone?'

The goal of this paper is to show that Inquisitive Semantics [7] offers the right set of operators that accounts for the process of clause-typing by the *ma*-particle. Inquisitive Semantics provides two closure operators that serve to modify the inquisitive content of a given proposition $\mathcal{P}$. The inquisitive operator *?* forms an inquisitive proposition by adding a pseudo-complement as a novel possibility for $\mathcal{P}$. The non-inquisitive closure *!* "flattens" a proposition by uniting all possibilities already contained in $\mathcal{P}$ into a single possibility.

(2)   a. $? =_{def} \lambda \mathcal{P}_{\langle\langle s,t\rangle,t\rangle}.\lambda p_{\langle s,t\rangle}.[\mathcal{P}(p) \vee {}^{*}\mathcal{P}(p)]$

     b. $! =_{def} \lambda \mathcal{P}_{\langle\langle s,t\rangle,t\rangle}.\lambda p_{\langle s,t\rangle}.\forall w.[w \in p \rightarrow \exists q.[q \in \mathcal{P} \wedge q(w)]]$

In what follows, I will argue that the *ma*-particle denotes the lexical realization of the combination of *?* and *!*. In other words, when *ma* is attached to a proposition, the proposition is flattened by *!*, followed by a novel possibility being added through disjoining by *?*. The analysis predicts that any question formed with *ma* is always a polar question, because by adding the *ma*-particle, the non-inquisitive operator *!* flattens the possibilities, and the outcome is always a singleton set. Once the inquisitive operator *?* brings in the pseudo-complement, the final result will always denote a proposition that contains only two possibilities – a polar question.

Furthermore, it will be shown that the analysis not only explains the possible patterns, but also yields a novel account of the incompatibility of the *ma*-particle and different types of questions. For instance, as noted by [10], the *ma*-particle is incompatible with A-not-A questions (hereafter, ANAQ). The novel observation made in this paper is that *ma* also cannot co-occur with *wh*-adjunct questions (ADJQ) and quantity questions (QNTQ). I will argue that this is because the pseudo-complement formed by *?* is the empty set, because (i) all the possible worlds have been exhausted by the two mutually exclusive possibilities in ANAQ, and (ii) the ADJQ and QNTQ do not have negative answers. Thus, the outcome of applying *?* cannot be interpreted as a question of any sort because it is NON-INQUISITIVE and INSIGNIFICANT, based on the Inquisitive Principle [1]. I will specify the semantics for the *ma*-particle in section 2, and turn to discuss the patterns of incompatibility in section 3. Section 4 concludes this paper.

2 Polar Q-particle and Inquisitive Semantics

Section 2.1 provides an argument from the response to the *ma*-questions to support the claim that *ma* is a polar question particle. Section 2.2 lays out the theoretical background for the semantics of *ma*, and shows how the proposed semantics for *ma* "types" polar questions.

2.1 *Yes/no*-particle as a diagnostic for polar question

In Chinese, *ma* is characterized as the polar question particle because a clause must be interpreted as a polar question when it is attached by the *ma*-particle,[1] as shown in example (1a) where *ma* is attached to an atomic declarative and the sentence has become a polar question. In example (1b), even when it contains a *wh*-phrase, the meaning for the clause is still a polar question because of the appearance of the *ma*-particle.

A solid diagnostic for the interpretation of polar question comes from the *yes/no*-particle used in the response. As pointed out by [5], the *yes/no*-particle in English is sensitive to question type. Particularly important for this paper is that the *yes/no*-particle can only be contained in a response to polar question, but not to *wh*-question.

(3) a. Q: What did you eat yesterday?   A: (*Yes) I ate the hamburger.

    b. Q: Did you eat something yesterday?   A: Yes, I ate the hamburger.

---

[1] It is worth to point out here that sometimes Chinese questions need not be formed with the *ma*-particle [10]. For instance, the question in (i) is formed with a rising intonation in the end of the sentence.

(i) I know that there will be a conference tomorrow in the Department of Philosophy, and John was making a note about the program and schedule. I ask:

nǐ   míngtiān huì qù zhéxué-suǒ   de huìyì?
you tomorrow will go philosophy-department DE conference
'You will go to the conference in the Department of Philosophy tomorrow?'

However, it should be noted that this kind of question actually corresponds to the rising declarative questions discussed in [8]. One typical property of Gunlogsonian declarative questions is that it has to be accompanied with a biased context. Thus, the question is not felicitous in context where the speaker does not have biased attitude, as shown in example (ii). In this paper, I will assume Gunlogson's analysis for declarative questions.

(ii) Context: I know that there will be a conference tomorrow in the Department of Philosophy, but I have no idea whether John will be there or not. At the lobby, I see John and ask:

#nǐ míngtiān huì qù zhéxué-suǒ   de huìyì?
you tomorrow will go philosophy-department DE conference
'You will go to the conference in the Department of Philosophy tomorrow?'

Similarly, the answer to the question in Chinese example (1b) can contain the *yes/no*-particle, *duì/bùduì* 'correct/incorrect,' as illustrated in the question-answer pair in (4). In contrast, without the *ma*-particle in that question as in (5), the response cannot contain the *yes/no*-particle.

(4) a. John kàn-guò shéme rén   ma?
       John see-ASP what   person MA
       'Did John see someone?'

    b. duì,   tā kàn-guò Bill.
       correct he see-ASP Bill
       'Yes, he saw Bill.'

(5) a. John kàn-guò shéme rén?
       John see-ASP what   person
       'Who did John see?'

    b. (*duì,) tā kàn-guò Bill.
       correct he see-ASP Bill
       'He saw Bill.'

In short, the minimal pair in example (4-5) demonstrate the effect of *ma* as the polar question particle. In the next subsection I will outline the theoretical background and show the explanation.

2.2 Semantics of the *ma*-particle

Section 2.2.1 outlines the assumptions and definitions of Inquisitive Semantics. Section 2.2.2 shows the application of the proposed semantics to the *ma*-particle.

2.2.1 Inquisitive Semantics

In Inquisitive Semantics, the semantic value of a sentence is conceived as one or several ways to update the common ground. In this way, a sentence expresses a proposition (of type $\langle st, t \rangle$) that denotes the set of one or multiple maximal possibilities $p_{\langle s,t \rangle}$. The informative content of a sentence is defined through info function. Specifically, a proposition that eliminates some possible world(s) from the common ground is informative, and a proposition that does not eliminate any world is non-informative. Throughout this paper I will assume a fixed set of worlds ω.

Definition 1. A *possibility* is the set $p \subseteq \omega$ of possible worlds.

Definition 2. A *proposition* is the set $\mathcal{P} \subseteq \wp(\omega)$ of possibilities.

Definition 3. The *informative content* of a proposition $\mathsf{info}(\mathcal{P}) = \bigcup \mathcal{P}$.

Definition 4. A proposition is *informative* iff $\mathsf{info}(\mathcal{P}) \subset \omega$.

Definition 5. A proposition is *non-informative* iff $\mathsf{info}(\mathcal{P}) = \omega$.

Crucial here is the notion of inquisitive content. When a proposition raises an issue to settle, it contains more than one maximal possibility. Specifically, a proposition that contains more than one maximal possibility is inquisitive, and a proposition that contains only one maximal possibility is non-inquisitive.

Definition 6. A proposition is *inquisitive* iff $\mathsf{info}(\mathcal{P}) \notin \mathcal{P}$.

Definition 7. A proposition is *non-inquisitive* iff $\mathsf{info}(\mathcal{P}) \in \mathcal{P}$.

Accordingly, the Inquisitive Principle that characterizes four different types of propositions is defined based on informativity and inquisitivity, as shown in Table 1. Two types of propositions are particularly important in the context of this paper: question and insignificance. A proposition is characterized as a question iff it is inquisitive and non-informative, and a proposition is insignificant iff it is non-inquisitive and non-informative.

Table 1: Inquisitive Principle

	Informative	Non-informative
Inquisitive	Hybrid	Question
Non-inquisitive	Assertion	Insignificant

(6) A proposition is a question iff $\text{info}(\mathcal{P}) = \omega$ and $\text{info}(\mathcal{P}) \notin \mathcal{P}$.

(7) A proposition is insignificant iff $\text{info}(\mathcal{P}) = \omega$ and $\text{info}(\mathcal{P}) \in \mathcal{P}$.

Finally, I adopt three operators defined by [12]. The inquisitive operator $?$ in (8) forms an inquisitive proposition by disjoining the proposition and its pseudo-complement as a novel possibility. Since negation is not treated as the complement set of the original denotation [12, p.22], a novel definition is required for the * (negation) operator, as in (9). The non-inquisitive closure $!$ serves to flatten a proposition by uniting all possibilities already contained in $\mathcal{P}$ into a single possibility. The proposed semantics for the *ma*-particle is that *ma* denotes the lexical complex of the combination of the two closure operators in (8) and (10) that serve to modify the inquisitive content.

(8) Inquisitive operator: $? =_{def} \lambda \mathcal{P}_{\langle\langle s,t\rangle,t\rangle}.\lambda p_{\langle s,t\rangle}.[\mathcal{P}(p) \vee {}^*\mathcal{P}(p)]$

(9) Pseudo-complement: $* =_{def} \lambda \mathcal{P}_{\langle\langle s,t\rangle,t\rangle}.\lambda p_{\langle s,t\rangle}.\forall w.[w \in p \to \neg\exists q.[q \in \mathcal{P} \wedge q(w)]]$

(10) Non-inquisitive operator: $! =_{def} \lambda \mathcal{P}_{\langle\langle s,t\rangle,t\rangle}.\lambda p_{\langle s,t\rangle}.\forall w.[w \in p \to \exists q.[q \in \mathcal{P} \wedge q(w)]]$

(11) Polar question particle: $[\![ma]\!] = ?!$

2.2.2 Application

The application of *ma* as the realization of the $?!$ complex to atomic declaratives is straightforward. Since an atomic declarative denotes a singleton set that contains only one possibility, the application of $!$ is vacuous in the sense that it does not change the inquisitive content. It is the novel possibility that is added by $?$ that turns the proposition into a polar question by making it inquisitive and non-informative.

(12) a. $[\![\text{John likai-le}]\!] = [\![!(\text{John likai-le})]\!] = \{\{w \mid \text{leave}_w(j)\}\}$

    b. $[\![?(!(\text{John likai-le}))]\!] = \{\{w \mid \text{leave}_w(j)\}\} \cup \{\{w \mid \neg\text{leave}_w(j)\}\}$
          $= \{\{w \mid \text{leave}_w(j)\}, \{w \mid \neg\text{leave}_w(j)\}\}$ POLAR Q

To explain how the *ma*-particle types the clause that contains a *wh*-phrase as in (1b), I assume, following [4], that Chinese in-situ *wh*-phrase denotes the set of entity that is composed through pointwise functional application [11]. The outcome of the semantic composition is the

set of multiple possibilities that corresponds to the classic semantic value of *wh*-question. By adding the *ma*-particle, *!* flattens the possibilities and the result is a non-inquisitive singleton set. Once the inquisitive operator *?* brings in the pseudo-complement, the outcome denotes a proposition that contains only two possibilities, i.e., the meaning for polar questions. As we have seen earlier, the *yes/no*-particle is sensitive to the meaning of question type. Since the semantic value of the *wh*-question has been changed, the answer to the question can naturally contain the *yes/no*-particle as a part of the response. This is illustrated in (13), assuming the universe consists of John, Bill, Cathy, and Mary.

(13) a. $[\![$John kàn-guò shéme rén$]\!] = \{\{w \mid \mathsf{see}_w(j,b)\}, \{w \mid \mathsf{see}_w(j,c)\}, \{w \mid \mathsf{see}_w(j,m)\}\}$

    b. $[\![!(\text{John kàn-guò shéme rén})]\!] = \{\{w \mid \mathsf{see}_w(j,b) \vee \mathsf{see}_w(j,c) \vee \mathsf{see}_w(j,m)\}\}$

    c. $[\![?(!(\text{John kàn-guò shéme rén}))]\!]$
$= \{\{w \mid \mathsf{see}_w(j,b) \vee \mathsf{see}_w(j,c) \vee \mathsf{see}_w(j,m)\}\} \cup \{\{w \mid \neg\exists x.\mathsf{see}_w(j,x)\}\}$
$= \{\{w \mid \mathsf{see}_w(j,b) \vee \mathsf{see}_w(j,c) \vee \mathsf{see}_w(j,m)\}, \{w \mid \neg\exists x.\mathsf{see}_w(j,x)\}\}$ POLAR Q

In Summary, the application of *!* operator produces a non-inquisitive singleton set. The *?* operator turns this set into an inquisitive proposition by disjoining the pseudo-complement formed by *. As a result, the outcome is always a proposition that contains two possibilities – a polar question. This corresponds to the analysis for *ma* as a clause-typer for polar questions in [2].

3   Insignificance

The analysis proposed in this paper yields the following prediction in (14)

(14) Throughout the composition, if the pseudo-complement formed by * is the empty set, the *ma*-question will be unacceptable because the output of disjoining is insignificant and non-inquisitive; i.e., $!(\mathcal{P}) \cup \varnothing = !(\mathcal{P})$.

If the prediction is borne out, we will be able to find some Chinese examples that are incompatible with the *ma*-particle. In this section I discuss the observation made by [10] for ANAQ, and provide some novel observation for ADJQ and QNTQ as empirical evidences that support the Inquisitive Semantics analysis, as the *ma*-particle really cannot co-occur with any of these questions.

3.1   A-not-A question and the *ma*-particle

As noted by [10], the *ma*-particle cannot co-occur with ANAQ. At first sight, this seems to be an argument against *ma* as a polar question particle, because ANAQ is a subtype of polar questions.

(15) ANAQ and *ma*:
John xǐ(huān)-bù-xǐhuān Bill *ma?
John li(ke)-NEG-like   Bill MA

One might suggest that ANAQ is a special kind of polar questions that cannot be typed by the *ma*-particle. For example, one observation made in the literature for ANAQ is that its response cannot contain the *yes/no*-particle, and as we have seen above, whether the response can contain the *yes/no*-particle is a crucial factor to characterize polar questions.

(16) Q: John xǐ(huān)-bù-xǐhuān Bill?
          John li(ke)-NEG-like   Bill
          'Does John like Bill or not?'

      A: (*duì,) tā xǐhuān Bill.
          correct he like   Bill
          'He likes Bill'

However, this observation is not accurate. For example, when ANAQ is formed with a copular verb *shì*, it immediately becomes plausible for the response to contain the *yes/no*-particle. Thus, an account is still called for the incompatibility of ANAQ and *ma*.

(17) Q: John shì-bú-shì   xǐhuān Bill?
          John copular-NEG-copular like   Bill
          'Does John like Bill or not?'

      A: duì,   tā xǐhuān Bill.
          correct he like   Bill
          'Yes, he likes Bill'

(18) John shì-bú-shì   xǐhuān Bill *ma?
       John copular-NEG-copular like   Bill MA

I argue that the example in (15) is not a counterexample against *ma* as a polar question particle. On the contrary, the incompatibility between *ma* and ANAQ is actually a strong argument in favor of the Inquisitive Semantics analysis for *ma* as a polar question particle. Following [13], I assume that the semantic value of an ANAQ is the set of two mutually exclusive possibilities.

(19) ⟦John xǐ-bù-xǐhuān Bill⟧ = $\{\{w \mid \mathsf{like}_w(j,b)\}, \{w \mid \neg\mathsf{like}_w(j,b)\}\}$

By adding the *ma*-particle to the ANAQ, three operations are performed. First, the non-inquisitive operator *!* flattens the proposition. Unlike the cases discussed earlier, at this point the proposition is insignificant because the possibility is exhaustive, i.e., $\mathcal{P} = \omega$.

(20) ⟦!(John xǐ-bù-xǐhuān Bill)⟧ = $\{\{w \mid \mathsf{like}_w(j,b) \vee \neg\mathsf{like}_w(j,b)\}\} = \omega$

INSIGNIFICANT

The pseudo-complement formed by the * operator now is the empty set. As a consequence, the application of *?* is vacuous because the inquisitive content is not changed, and the analysis correctly predicts ANAQ to be incompatible with the *ma*-particle, since the outcome of disjoining the proposition and the empty set is still insignificant.

(21) ⟦?(!(John xǐ-bù-xǐhuān Bill))⟧ = $\{\{w \mid \mathsf{like}_w(j,b) \vee \neg\mathsf{like}_w(j,b)\}\} \cup \varnothing$
                                      $= \{\{w \mid \mathsf{like}_w(j,b) \vee \neg\mathsf{like}_w(j,b)\}\}$

INSIGNIFICANT

In short, the analysis yields a novel account for the incompatibility of the *ma*-particle and ANAQ. This solves the conundrum raised by [10] that the polar question particle *ma* is incompatible with a polar question (ANAQ).

3.2   The absence of semantic negative answers

The novel observation in this paper is that the *ma*-particle also cannot co-occur with ADJQ and QNTQ, for instance:

(22) a. ADJQ
         John weìshéme zaì   kū?
         John why   PROG cry
         'Why is John crying?'

      b. ADJQ + *ma*
         *John weìshéme zaì   kū ma?
         John why   PROG cry MA

(23) a. QNTQ:
         John maĭ-le   jĭ-běn   shū?
         John buy-ASP how.many-CL book
         'How many books did John buy?'

      b. QNTQ + *ma*:
         *John maĭ-le   jĭ-běn   shū ma?
         John   buy-ASP how.many-CL book MA

   I argue that this is because of the existential presupposition involved in ADJQ and QNTQ. Specifically, in (22a), the ADJQ presupposes the existence of a reason for why John is crying. In (23a), the QNTQ presupposes that a certain number of books have been purchased by John. The evidence comes from the negative response to the questions. While a negative response to ADJQ is usually felicitous, the reason itself cannot be negated alone. In other words, *John is not crying for any reason* is not considered a felicitous answer.

(24) Q: Why is John crying?

      a. John meí zaì   kū.
         John NEG PROG cry
         'John is not crying.'

      b. #John meí yīnwèi   rènhé yuányīn zaì   kū.
         John   NEG because any   reason   PROG cry
         'John is not crying for any reason.'

Similarly, a negative response is usually allowed for QNTQ. However, again the number expression cannot be negated alone. The response to the QNTQ can be either *John bought some number of books*, or *John didn't buy any book*, but *líng-běn shū* '0 book' is not considered a felicitous answer.

(25) Q: How many books did John buy?

      a. John meí maĭ shū.
         John NEG buy book
         'John did not buy any book.'

b. #John maǐ-le   líng-běn shū.
   John   buy-ASP 0-CL   book
   'John bought 0 book.'

Thus, I assume that ADJQ and QNTQ do not have any negative answer, so the responses in (24b) and (25b) are infelicitous. The negative responses in (24a) and (25a) are to be conceived as a presupposition denial against the existential presupposition in ADJQ and QNTQ. The incompatibility of the *ma*-particle and the two types of *wh*-questions is now explained because the pseudo-complement formed by the inquisitive operator ? will become the empty set since no semantic negative answer is involved in the semantic value of ADJQ and QNTQ. As we have seen in the case of ANAQ, these questions are predicted to be unacceptable when the *ma*-particle is attached, because of insignificance.

This analysis yields the following theoretical implication for the debate of existential presupposition in questions. The central issue of the debate is whether questions have existential presuppositions, and generally there are two mainstream theories for the debate. [6] argues that the negative responses are actually semantic answers to questions. Thus, *wh*-questions do not have any existential presupposition. In contrast, it is argued that the negative responses are denials against existential presupposition in questions [9, 3]. Given our discussion, the *ma*-particle becomes an effective diagnostic for existential presupposition in questions, because whether a *wh*-question has a semantic negative answer is crucial for the (in-)compatibility of the *ma*-particle. Our answer to the question of the existential presupposition debate is that some questions like (1b) do have a negative answer. In this way, they are compatible with the *ma*-particle and do not have any existential presupposition in nature. This is in line with the proposal made in [6]. However, some question types like ADJQ and QNTQ are different. The negative responses in (24a) and (25a) are not semantic answers involved in the questions. Instead, they are presupposition denials. Thus, since the pseudo-complement formed by * is the empty set, the proposition is deemed to be insignificant.

In short, the Inquisitive Semantics analysis yields a novel account for the incompatibility puzzle raised by [10]. Crucial here is that the pseudo-complement set formed by * operator is the empty set. It is also shown that the compatibility of the *ma*-particle and other different types of *wh*-questions depends on whether the question has a negative answer. A conclusion drawn from the discussion is that existential presupposition is not an inherent and general properties for all kinds of questions, but rather it depends on the meaning of question types.

4   Conclusion

In this paper I present a novel analysis for the Chinese polar question particle *ma*. It is argued that Inquisitive Semantics provides the right set of operators that can well account for the semantics of the *ma*-particle. The two operators ? and ! together produce the set of two possibilities by flattening the proposition, followed by adding the pseudo-complement as a novel possibility. As a result, the outcome is always the set that contains only two possibilities – a polar question.

It is further shown that the analysis can also explain the impossible patterns. The incompatibility between the *ma*-particle and other types of questions results from (i) the possibilities are exhaustive (ANAQ), or (ii) the absence of negative answers (ADJQ and QNTQ). The theoretical implication of the analysis is questions do not generically have existential presupposition, but rather whether existential presupposition is involved in a question depends on the meaning of the question type.

In this way, the analysis corresponds to the Clausal Typing Hypothesis in [2], and further offers a novel semantics for the process of clause-typing. In addition, the analysis also complements the hypothesis by predicting the incompatibility of the *ma*-particle and different types of questions.

References

[1] Scott AnderBois. Focus and uninformativity in Yucatec Maya questions. *Natural Language Semantics*, 20:349–390, 2012.

[2] Lisa Lai-Shen Cheng. *On the Typology of Wh-Questions*. PhD thesis, Massachusetts Institute of Technology, 1991.

[3] Ileana Comorovski. *Interrogative phrases and the Syntax-semantics Interface*. Kluwer Academic Publisher, 1996.

[4] Hongyuan Dong. *Issues in the Semantics of Mandarin Questions*. PhD thesis, Cornell University, 2009.

[5] Donka Farkas and Floris Roelofsen. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language*, 91(2):329–414, 2015.

[6] J. Groenendijk and M. Stokhof. *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam, Amsterdam, 1984.

[7] Jeroen Groenendijk and Floris Roelofsen. Inquisitive semantics and pragmatics. In Zubeldia-L. Larrzabal, J., editor, *Proceedings of the 7th international Tbilisi symposium on language, logic, and computation*, pages 41–72, Leioa, 2009. University of the Basque Country Publication Service.

[8] Christine Gunlogson. Declarative questions. In Brendan Jackson, editor, *Proceedings of SALT XII*, pages 124–143, Cornell University, Ithaca, NY, 2002. CLC Publications.

[9] Lawrence Horn. *On the semantics of properties of logical operators in English*. PhD thesis, UCLA, Los Angeles, CA, 1972.

[10] Boya Li. *Chinese final particles and the Syntax of the periphery*. PhD thesis, Leiden University, 2006.

[11] Junko Shimoyama. Indeterminate phrase quantification in Japanese. *Natural Language Semantics*, 139–173(14):2, 2006.

[12] Nadine Theiler. A multitude of answers: embedded questions in Typed Inquisitive Semantics. Master's thesis, University of Amsterdam, Amsterdam, 2014.

[13] Mengxi Yuan and Yurie Hara. Questioning and Asserting at the same time: the L tone in A-not-A questions. In Maria Aloni; Michael Franke; Floris Roelofsen, editor, *Proceedings of 19th Amsterdam Colloquium*, pages 265–272, University of Amsterdam, 2013.

Three puzzles about negation in non-canonical speech reports

Emar Maier[1] and Corien Bary[2*]

[1] University of Groningen, Groningen, The Netherlands
emar.maier@gmail.com
[2] Radboud University, Nijmegen, The Netherlands
c.bary@ftr.ru.nl

1 Non-canonical speech reports

The canonical forms of reported speech are direct (1) and indirect (2) speech.

(1)   Mary said "I'm running late, so don't wait up."

(2)   Mary said that she was running late so we shouldn't wait up.

Both of these are ways to report what Mary said. The difference is that in direct speech we report her words (by quoting them verbatim, or by giving a demonstration of the original utterance, depending on your quotation theory of choice), while in indirect speech we report the content of the original words.

Beyond the two well-established report types in (1) and (2) there are various forms of reported speech in which the aspect of reporting is somehow backgrounded and the content of the report itself, i.e., the complement or quoted phrase, serves as the main point. This backgrounding may be achieved syntactically, as in parenthetically framed direct (3) and indirect speech (4).

(3)   "I'll be late again," she said, "so don't wait up."

(4)   She would be late again, she said, so we shouldn't wait up.

Alternatively, the backgrounding effect may be achieved morphologically, as in reportative evidential marking in Cuzco Quechua (5), the reportative subjunctive in German (6) and Dutch (7), or reportative modals (7b).[1]

(5)   Para-sha-n-si.
       rain-prog-3-REP
       'It is raining, I am told'   (Cuzco Quechua, Faller 2002)

(6)   (Er sagte, sie sei   schön.) Sie habe   grüne Augen.
       (He said   she be-REP pretty.) She have-REP green eyes
       (He said she's pretty.) She has green eyes, he said.'   (German, Jäger 1971)

(7)   a.   Anne zou thuis zijn.   b.   Anne schijnt thuis te zijn
            Anne would.REP at-home be   Anne seems.REP at-home to be
            'Anne is at home, reportedly'   (Dutch)

---

[1]Cf. also Schenner (2009) on the reportative modals *sollen* and *wollen* in German.

Finally, the backgrounding of the reporting may also be entirely unmarked and left to pragmatics. This happens in so-called semantically parenthetical indirect reports (Simons 2007; Hunter et al. 2006; Hunter 2015). In (8), for instance, B's answer takes the form of a canonical indirect report of something Mary said. But the fact that Mary said something does not address the question under discussion, so the main contribution of the report seems to be just the complement, that John is ill.

(8)   A: Why was John not at the meeting?
        B: Well, Mary said that he's ill.

We identify in this paper three puzzles concerning the peculiar interaction between these non-canonical types of reporting and negation. We then propose an analysis to solve all three.

2   Three Puzzles

2.1   Puzzle 1: De Cornulier's observation.

De Cornulier (1978) observes that parenthetically framed direct discourse does not allow negation in the saying-frame (9a), in contrast to canonical direct or indirect discourse, (9b-c) (cf. Recanati 2001:646–647):

(9)   a.   *"I'll be late", she didn't say.
        b.   She didn't say, "I'll be late."   c. She didn't say that she'd be late.

Recanati explains this syntactic negation-blocking in terms of the mimetic character of quotation in parenthetical direct discourse. Following Clark and Gerrig's (1990) demonstration theory of quotation, Recanati argues that in a parenthetical direct speech construction the quoted phrase is itself a demonstration/depiction of an original speech act, so that it doesn't make sense to explicitly deny that such a speech act took place. To explain the felicity of (9b), he stipulates that this canonical form can be re-analyzed as a form of pure mention rather than genuine direct speech.

However, we find similar negation-blocking effects in parenthetical indirect discourse:

(10)   *She would be late again, she didn't say, so we shouldn't wait up.

Moreover, for unembedded reportative subjunctives and evidentials, wherever we add a negation, it never manages to take scope over just the saying. For reportative evidentials this is illustrated in (11):

(11)   Mana=s   phalay-ta ati-n=chu.
        not=REP fly-acc   can-3=neg
        'It cannot fly, I was told.'   (Cuzco Quechua, Faller 2014)

The negation in (11) does not affect the information that the speaker reports what someone told her – we will refer to this as the *reporting component* of the semantic content of a non-canonical report – but only the content of that report, the *reported component*. (12) shows the same point for the German unembedded subjunctive in the second sentence. If we add a negation to the unembedded subjunctive clause and/or to the preceding overtly embedded report construction, the inference that the content of the clause was said by someone survives:

(12)   Rau hatte ihm (nicht) geraten, im Amt zu bleiben. Doch müsse er (nicht) selbst die Entschei-
        dung treffen.
        'Rau had (not) advised him to stay in office. But he himself would-REP (not) have to make the
        decision (⤳, he said).'   (adapted from Fabricius-Hansen and Sæbø 2004:214)

Finally, for semantically parenthetical reports like (8) above it is syntactically possible to add a negation and even have it take scope over the saying. However, if we do this the parenthetical reading is no longer available and B's answer becomes infelicitous.

(13)   A: Why was John not at the meeting?
          B: #Well, Mary didn't say that he's ill.

Since most of these non-canonical report types fall on the indirect discourse, non-quotational side of the reporting spectrum, Recanati's explanation in terms of the mimetic character of quotation cannot capture the more general pattern.

2.2   Puzzle 2: Non-deniability of the reporting.

In canonical direct and indirect discourse the reporting itself is easily denied by another discourse participant.

(14)   A: She said {"I'm innocent"/ that she was innocent}
          B: No, she didn't (= she didn't say that)

More precisely, depending on the context, we can in principle deny either the reporting or the reported component of a canonical report, as illustrated in (15).

(15)   B': Nonsense, she may be innocent, but she would never say that.
          B": Nonsense, she's guilty, regardless of what she told you.

By contrast, for our non-canonical varieties a denial targeting the reporting content is impossible, or at least much more difficult.

Let's start with evidentials. Faller (2014:6) observes that in Cuzco Quechua one cannot challenge (11) ("It cannot fly, reportedly") with "That's not true, nobody told you this" (cf. Murray (2011) for a similar observation about Cheyenne).

The same seems to hold for the Dutch and German reportative moods and modals. For instance, with a Dutch reportative *zou* a denial may target the reported content, as in B's answer in (16), but not the just the reporting content, as in B':

(16)   A: Anne zou thuis zijn. ('Anne would.REP be at home')
          B: Onzin. Ik weet niet wat je gehoord hebt, maar ze is niet thuis
              'Nonsense. I don't know what you heard, but she's not at home'
          B': #Onzin. Niemand heeft dat beweerd, ook al is ze misschien wel thuis
              'Nonsense, nobody said that, though she may be at home'

For semantically parenthetical reports, the situation is more subtle. A third speaker may well interrupt our dialogue to object that Mary didn't say that. Note however that explicitly targeting only the reporting, as in (17), sounds slightly off, as if C is deliberately and uncooperatively misinterpreting B's answer to hijack the conversation and start talking about Mary instead of John.

(17)   A: Why was John not at the meeting?   B: Well, Mary said he's ill.   C: No she didn't.

Parenthetical direct and indirect speech reports introduce another complication. These forms typically occur only in narrative text and seem awkward in a spontaneous dialogue setting where speakers are arguing about what's true and/or who said what. For this reason it's hard to see which component is more readily deniable. We'll return to this genre dependence of parenthetical direct/indirect speech below.

Let's assume that an assertion may express various layers of contents (Bach, 1999; Potts, 2005; Geurts and Maier, 2013), of which one (or more) can serve as the main point and other are backgrounded. Then the fact that denials typically target one such content layer is evidence that that is the main point. In our case we've been distinguishing a reporting meaning component (x said p) and a reported component (p) associated with non-canonical reports. The denial facts collected above then indicate that the main point of a parenthetical report is the reported meaning component, while the reporting component is backgrounded. This general diagnosis is in line with the literature on these subjects, including the cases for which our denial tests were somewhat inconclusive, cf. e.g. De Vries (2006) on backgrounding in parenthetical direct discourse, and Simons (2007) on the main point of a semantically parenthetical indirect report. In sum then, our second puzzle can be paraphrased without denial: How is it that in non-canonical reports the reported component constitutes the main point of the utterance, with the reporting component pushed into the background?

2.3   Puzzle 3: Reportative commitment.

In puzzle 2 we observed that only the reported component is readily available for denial in a dialogue. We concluded that the reporting component is backgrounded while the reported component contributes the at-issue information. However, it has been noted, primarily in the evidentiality literature, that despite this apparent foregrounding of the reported component, the speaker herself still need not be fully committed to its truth. We see this in examples like (18) where the reporter herself explicitly denies the reported content:

(18)   Pay-kuna=s qulqi-ta   saqiy-wa-n. Mana=ma, ni un sol-ta   saqi-sha-wa-n=chu
       they=REP   money-acc leave-1o-3 no=impr   not one Sol-acc leave-prog-1O-3=neg
       'They left me money, I was told. But no, they didn't leave me one Sol.' (Cuzco Quechua, Faller 2002)

This apparent lack of commitment in reportative evidential constructions is dubbed 'reportative exceptionality' in the evidentiality literature (cf. AnderBois 2014 for a detailed overview), referring to the fact that this pattern is not found in any other types of evidentiality marking.

Let's investigate the speaker's commitment to the reported component in some of our other non-canonical report types. We'll start with parenthetical direct and indirect speech. In a typical narrative setting, where these constructions are most at home, the narrator doesn't usually disagree with what a protagonist says, but sometimes she does:

(19)   {"I'm innocent," she said, "so let me go". / She was innocent, she said, so we should let her go.}
       But I didn't believe her. And it turned out she was guilty and letting her go was a big mistake.

The lack of commitment to the reported component is also clear with Germanic reportative modals and mood:[2]

(20)   Anneloes schijnt   thuis   te zijn, maar ik geloof er   niets   van.
       Anneloes seems.REP at-home to be,   but   I believe there nothing of
       'Anneloes is at home, I am told, but I don't believe it.'   (Dutch, Koring 2013)

(21)   Er sagte, sie sei schön. Sie habe grüne Augen. In Wirklichkeit hat sie aber blaue Augen.
       'He said she is beautiful. She has green eyes, he said. But in reality she has blue eyes.' (German)

For semantically parenthetical indirect speech, Hunter (2015) likewise observes a "hedged commitment", which we could bring out as follows:

---

[2]Cf. also Schenner 2009:184 on "shifted responsibility" with reportative *sollen* in German.

(22)   A: Where is Mary?   B: Bill said she's still at work, but I don't believe it.

All in all, with respect to commitment, non-canonical reports pattern with canonical direct and indirect speech, i.e. the speaker is not committed to the truth of the content of the report:

(23)   John said {"I didn't do it"/that he didn't do it}, but he clearly did.

The third puzzle then consists in reconciling this behavior with the apparent backgrounding brought out with denials in puzzle 2 above: If the main point of a report is the reported content, how come the speaker is not committed to that?

3 Two dimensions of meaning

As pointed out above, our non-canonical report types seem to emphasize the content of the reported clause over that of the saying-frame. In formal semantics, evidential and parenthetical content in general are analyzed as somehow secondary or supplementary to the main point. For instance, Potts (2005) analyzes appositives and expressives generally as conventional implicatures, which contribute their content to a separate meaning dimension. Likewise, the "Baseline Conception" (AnderBois, 2014) in the evidentiality literature is that an evidential contributes it's evidential content as a special, not-at-issue meaning component, with an ongoing controversy over what type of meaning that is, i.e. a presupposition, conventional implicature, or speech act modifier. In discussing puzzle 2 above we've confirmed such a special, not-at-issue status of the reporting meaning component in all our non-canonical report forms. Below we propose to apply the semantic tools previously developed to deal with not-at-issue meaning components to give a uniform analysis of non-canonical reports that explains our three negation-related puzzles.

The first step towards this unified analysis is to represent the two meaning components in a simple two-dimensional logical framework. Roughly, the first dimension represents the main point, while the second dimension represents the not-at-issue content. Applied to our non-canonical report constructions, the first dimension contains the reported component, the second the reporting component. Formally, our logical forms consist of pairs of expressions of dynamic predicate logic with intensional variables (p is a variable over propositions, type st). The 2D logical forms for some of our key examples are below:

(24)   a. She would be late again, she said [= (4)]   b. Anne zou thuis zijn. [= (7)]

$$\left\langle \begin{matrix} p \\ \exists p[say(x,p) \land p =^\wedge late(x)] \end{matrix} \right\rangle \qquad \left\langle \begin{matrix} p \\ \exists p[say(x,p) \land p =^\wedge home(a)] \end{matrix} \right\rangle$$

Before we go into the compositional generation, and the dynamic/pragmatic interpretation of these 2D logical forms, let's discuss the logical forms themselves.

Note first of all the free variable x which represents the reported speaker as an anaphoric element to be resolved in context. We'll assume an existential closure operation to interpret such a free variable when the context provides no salient speaker to serve as the source of the reporting.

Second, note that regular indirect reports are now ambiguous between a one-dimensional canonical logical form, (25a), and a two-dimensional semantically parenthetical logical form, (25b).

(25)   B: Well, Mary said he's ill. [=(8)]

    a. $say(m,^\wedge ill(x))$ b. $\left\langle \begin{matrix} p \\ \exists p[say(m,p) \land p =^\wedge ill(x)] \end{matrix} \right\rangle$

Third, note that in our non-canonical logical forms we use the intensional variable p as main point, instead of just copying the scope material. In this way we avoid 'binding problems' with anaphora and

existential quantifiers, as in:

(26)   Iemand zou Jan verleid hebben.
        'Someone seduced John, reportedly'   (Dutch)

On our two-dimensional approach to non-canonical reports both dimensions contain a representation of the reported component (somebody seduced John), but a representation like (27) that involves two independent existential quantifiers is too weak, as it allows the cheater in the at-issue component to be chosen completely independently from the alleged cheater in the backgrounded component (cf. Geurts and Maier 2013 for discussion of similar multi-dimensional binding problems).

(27) $\left\langle \begin{array}{c} \exists x[seduce(x,j)] \\ say(z,^\wedge \exists x[seduce(x,j)]) \end{array} \right\rangle$

One reason that we choose a propositional variable p of type st rather than one of the standard sentence type t (i.e. the type of both dimensions in (27)) is that that this allows us to account for the vacuous use of the reportative subjunctive when embedded in a canonical indirect discourse (without any additional morphosyntactic assumptions about "concord"): *Er sagte sie sei schön* 'he said that she was-REP beautiful' means the same as *Sie sei schön* 'she was-REP beautiful' (Bary and Maier, 2014). When discussing puzzle 2 below we'll incorporate an extensionalizer $^\vee$ in the definition of the dialogue move of acceptance (and rejection).

We'll give the dynamic semantic/pragmatic interpretation of our 2D logical forms in the course of solving puzzles 2 and 3 below. In the remainder of this section, let's focus on the compositional derivation in the syntax/semantics interface. Following Bary and Maier (2014) (who in turn build on Fabricius-Hansen and Sæbø 2004) we analyze the reportative evidential morpheme, modal, or mood as introducing a sentential operator REP in the syntax at LF:

(28)   Iemand zou Jan verleid hebben $\rightsquigarrow$ [REP [someone [seduced John]]

The REP operator is interpreted with the following rule:

(29) $[\![\text{REP } \varphi]\!] = \left\langle \begin{array}{c} p \\ \exists p[say(x,p) \wedge p =^\wedge [\![\varphi]\!]] \end{array} \right\rangle$

For overtly parenthetical direct and indirect reports we introduce PAR, which abstracts away the saying from the REP operator, and an additional movement of the complement of a report, leading to the following LF:

(30)   She would be late, she said. $\rightsquigarrow$ [PAR [she be late]] [λp. she said p]

(31) $[\![\text{PAR } \varphi]\!] = \lambda Q \left\langle \begin{array}{c} p \\ \exists p[Q(p) \wedge p =^\wedge [\![\varphi]\!]] \end{array} \right\rangle$

For semantically parenthetical reports one option would be to say that the ambiguity observed above is indeed essentially syntactic, i.e. a canonical indirect discourse surface form may hide a syntactic LF like (30). Ultimately, perhaps a semantic/pragmatic derivation of the ambiguity may be preferable here, but we leave this for future research.

4   Re: puzzle 1. Evidentiality

In many of the examples above the intuitive interpretation of the second component is merely to indicate the source from which the speaker learned the main point. For evidentials this is part of the

aforementioned Baseline Conception. A direct evidential indicates that the speaker's evidence for the scope proposition (the at-issue component) is of a direct perceptual nature, while a reportative evidential indicates hearsay evidence. So the second dimension in our 2D logical forms of reportative evidentials intuitively indicates the evidence that supports the information in the first component. This evidential interpretation naturally extends to the Germanic reportative modals and moods, e.g. in *Anne zou thuis zijn*, (7), the proposition that someone said she's at home is presented as evidence for the proposition that she is.

To capture the puzzling negation-blocking effects with these evidential reports, we just need to assume that there are natural restrictions on what counts as evidence: seeing that p, or someone telling you that p are acceptable sources of evidence for p, while not seeing that p or someone not saying that p are not. It follows that 'not saying that p' is blocked from occurring on the second dimension of our logical forms. This evidential interpretation of our 2D logical forms thus correctly rules out certain readings in evidential reports:

(32)   Anne zou niet thuis zijn. (Anne would.REP not be at home)

$$* \left\langle \begin{matrix} p \\ \neg \exists p [say(x,p) \wedge p =^\wedge home(a)] \end{matrix} \right\rangle \quad \text{ok:} \left\langle \begin{matrix} p \\ \exists p [say(x,p) \wedge p =^\wedge \neg home(a)] \end{matrix} \right\rangle$$

In this way, general restrictions on admissable evidence effectively explain why negation never takes scope over the hearsay component in a reportative evidential construction. But how far can we extend the evidential interpretation of the reporting dimension to our other non-canonical reports, i.e. semantically and syntactically parenthetical reports?

Let's start with the semantically parenthetical report type. In line with observations from Simons (2007) and Hunter et al. (2006), who explicitly compare semantically parenthetical readings with hearsay evidentials, an evidential interpretation of the secondary dimension fits these examples rather well. Intuitively, in (8) B answers A's question by saying why John isn't at the meeting, using the report construction to indicate that she has it on hearsay evidence.

However, for syntactically parenthetical direct and indirect speech an evidential interpretation often seems rather far-fetched. In a typical narrative setting, where these types of reporting are especially common, as we saw earlier, we have a more or less omniscient narrator, who surely doesn't need to base her assertions on hearsay evidence from a (fictional) character. Lacking a clear understanding of the meaning dimensions at play in these narrative reports we suggest that Recanati's (2001) mimetic explanation applies to this restricted subset set of parenthetical reports. That is, first of all, parenthetical direct reports involve an act of quotational demonstration, which itself already presupposes the existence of the speech act so demonstrated which rules out explicit negation. Second, we'll assume that parenthetical indirect reports come in two varieties. The ones we find in narratives are actually cases of free indirect discourse and as such are arguably just as quotational and hence demonstrational as direct reports Maier (2015), so Recanati's explanation applies. The second variety are genuine indirect reports, with a parenthetical reporting clause. We submit that these truly indirect forms of overtly parenthetical reporting have the same distribution and evidential interpretation as the semantically parenthetical reports we already discussed above, as illustrated in (33).

(33)   Why is John not at the meeting?
        Well, he's ill, {Mary said/I hear/I'm told/he says}, so he's probably in bed.

5   Re: puzzle 2. Backgrounding

To capture the non-deniability/backgrounding observation we follow a recently popular idea about the interpretation of appositives and (some) other conventional implicatures, viz. that they impose a non-

negotiable, forced "pre-update" of the common ground (Murray, 2011; Koev, 2013; AnderBois et al., 2015; Griffiths, 2015). By contrast, following ideas from Farkas and Bruce (2010), and Inquisitive Semantics (Groenendijk and Roelofsen, 2009), at-issue content is then analyzed as an "update proposal", that can be accepted or rejected by other discourse participants.

The interpretation of a sentence with an appositive is thus analyzed as a three-step process: first the appositive content directly updates the common ground. Then an additional update is proposed. Only this proposed update is at-issue and on the table for discussion by the interlocutors. If the proposal is not challenged at the next turn in the dialogue, it is effectively accepted ("grounded") and we proceed to perform the second update.

As Murray (2011) points out, this special dynamic interpretation of appositives is in line with Faller's (2002) speech act modifier view of evidentials as comprising a "presentation" of the reported component, along with an assertion of the reporting component. In this paper we extend it to all our non-canonical reports. Concretely, in this section we propose a three-step dynamic, semantic/pragmatic interpretation of the compositionally generated 2D logical forms from section 3. The resulting system predicts the behavior in puzzle 2, i.e., the non-deniability/backgrounding of the reporting content in non-canonical reports as opposed to canonical direct and indirect speech.

Rather than adopting one of the cited proposals for dealing with appositives, we'll provide here our own stripped down, toy model, containing just the key features relevant to our puzzles. As a first approximation we propose the following dynamic interpretation rule for 2D logical forms:

(34) $c + \left\langle \begin{matrix} \varphi \\ \psi \end{matrix} \right\rangle = (c +_1 \psi) +_2 \varphi$

In this formula, c is the context set (i.e., a set of world–assignment pairs), representing the common ground; $+_1$ is the standard dynamic semantic update; and $+_2$ denotes a proposed update. Let's illustrate what this means with a concrete example, our German reportative subjunctive discourse.

(35)   Er sagte sie sei schön. Sie habe grüne Augen. [=(6)]

Consider the point after we just interpreted the first sentence, i.e. c contains the information that there is a man, associated with discourse referent x, and a woman, y, and x said that y is pretty. The second sentence contains a reportative subjunctive, syntactically analyzed with the sentential operator REP at LF:

(36)   [REP [she have green eyes]]

By (29) REP compositionally triggers the generation of a 2D logical form. Let's further assume that the overt pronoun *sie* picks out y, and the covert subject of the speech report is identified as x, the speaker of the overt speech report from the previous sentence:

(37) $\left\langle \begin{matrix} p \\ \exists p[say(x,p) \wedge p =^{\wedge} green.eyes(y)] \end{matrix} \right\rangle$

By (34), the dynamic interpretation of (37) in context proceeds by updating c with the evidential information in the second dimension. The existential quantifier extends the domain of the assignments in c by mapping p to the proposition that y has green eyes. Note that this allows us to subsequently interpret the apparently free variable p in the top dimension of (37). The condition $say(x,p)$ throws out possibilities from c where x didn't say that y has green eyes. In sum, the effect of this first step, $+_1$, is that we've now updated the context with the information that x said that y has green eyes and made the proposition that y has green eyes available for future anaphora.

As for $+_2$ the most important point for now is that it should make the reported content, that she

has green eyes, at issue, i.e., open for discussion. To model this type of issue-proposing we have to say something about acceptance and denial as moves in a dialogue. The idea is that once proposed, a proposition will either be accepted as true (by default), or rejected by the interlocutors. Formalizing this properly requires formulating a dialogue model that includes operations like proposing ($+_2$), accepting, and denying. (38) gives some key definitions: proposing a proposition means just pairing it with the current context set, and accepting (denying) such a pair means updating with the information that the proposal is true (not true).

(38)   a. $c +_2 \varphi = \langle c, \varphi \rangle$
        b. $accept(\langle c, \varphi \rangle) = c +_1 {}^\vee \varphi$
        c. $deny(\langle c, \varphi \rangle) = c +_1 \neg^\vee \varphi$

In addition, the dialogue system should include a rule to the effect that each proposal is automatically accepted at the start of the next turn, unless that next turn is a denial. Finally, for regular, one-dimensional logical forms, of type t, we simply update the context with $+_2$ (but note that we have to adjust (38b-c) so that $^\vee$ is only applied if φ is of type st).

6   Re: puzzle 3. Commitment

In section 4 we have seen that the first dimension of our logical forms contains a proposition, representing the reported content. Crucially this information is presented as an issue rather then claimed to be true. In section 5 we have modeled this idea in a little dialogue system. Essentially, the presented issue is only added to the common ground if none of the interlocutors objects. In section 2.3 however we established that not only is the truth of φ negotiable in dialogue, even the speaker herself is not necessarily committed to it.

In the evidentiality literature, accounts differ on whether speaker commitment to the reported content is a default, built into the semantics, or a pragmatic affair. On the one side, AnderBois (2014) argues that this commitment is part of the semantics of evidentials, analyzing exceptions in terms of a pragmatic perspective shift of the sort proposed by Harris and Potts (2010) to explain non-speaker orientation of expressives and appositives. On the opposite side, Murray (2011) claims that the reported content is not proposed to be added to the common ground at all.

On our account the speaker herself initially only proposes (or *presents*, as Faller puts it) the reported content as an issue, represented as a propositional variable on the first dimension of the logical form. At the end of her turn, if unchallenged, the issue effectively turns into an assertion ($^\vee \varphi$), by implicit acceptance as defined in (38). But before the end of her turn the speaker herself may still modify her proposal and weaken her eventual commitments. This is, we suggest, precisely what happens in the typical examples of reportative exceptionality discussed in section 2.3. By saying *reportedly φ, but I don't believe it*, the speaker effectively cancels the issue, whether φ, the truth of which she would otherwise become pragmatically committed to by the end of her turn. A precise logical description of this idea of issue-cancelation (for instance, in terms of inquisitive semantics) is left for future research.

7   Conclusion.

Drawing on recent insights from research on evidentiality and appositives, we have proposed a unified solution to three puzzles involving negation, denial and commitment in 'non-canonical reports'. At the very least, we have shown that this is a natural class of reported speech constructions that shares some interesting properties and deserves further attention.

Particularly interesting for future research is the point where the behaviors of the different subtypes diverge. We encountered one instance of this when discussing puzzle 1, viz. what we might call the split between narrative and evidential reports. Although overtly parenthetical direct and indirect speech reports seem to share the basic backgrounding (puzzle 2) and commitment (puzzle 3) properties with the other non-canonical varieties, they do not share the evidentiality aspect (puzzle 1). That is, while a German subjunctive or semantically parenthetical construction is used to indicate that the at issue (but non-committed) proposition is based on hearsay evidence, the reporting content in a direct or free indirect report in a narrative setting serves a very different discourse function. Further research is required to figure out exactly what this discourse function is and how it can be derived from features of their narrative environment. This will shed new light on the status of free indirect discourse and its relation to direct and indirect discourse.

References

AnderBois, S. (2014). On the exceptional status of reportative evidentials. *Proceedings of Semantics and Linguistic Theory (SALT) 24*, 234–254.

AnderBois, S., A. Brasoveanu, and R. Henderson (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics 32*, 93–138.

Bach, K. (1999). The Semantics-Pragmatics Distinction: What It is and Why It Matters. In K. Turner (Ed.), *The Semantics-Pragmatics Interface from Different Points of View*, pp. 65–84. Oxford: Elsevier.

Bary, C. and E. Maier (2014). Unembedded Indirect Discourse. *Sinn und Bedeutung 18*, 77–94.

Clark, H. and R. Gerrig (1990). Quotations as Demonstrations. *Language 66*(4), 764–805.

de Cornulier, B. (1978). L'incise, la classe des verbes parenthétiques et le signe mimique. *Cahier de linguistique* 8.

De Vries, M. (2006). Reported direct speech in Dutch. *Linguistics in the Netherlands 23*(1), 212–223.

Fabricius-Hansen, C. and K. J. Sæbø (2004). In a mediative mood: The semantics of the German reportive subjunctive. *Natural Language Semantics 12*, 213–257.

Faller, M. (2014). Reportativity, (Non-)at-issueness, and Assertion. *Proceedings of BLS 40*.

Faller, M. (2002). *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. Ph.D. thesis, Stanford.

Farkas, D. and K. Bruce (2010). On reacting to assertions and polar questions. *Journal of Semantics 27*(1), 81–118.

Geurts, B. and E. Maier (2013). Layered Discourse Representation Theory. In A. Capone, F. Lo Piparo, and M. Carapezza (Eds.), *Perspectives on Linguistic Pragmatics*, pp. 311–327. Berlin: Springer.

Griffiths, J. (2015). *On appositives*. Ph.D. thesis, Groningen.

Groenendijk, J. and F. Roelofsen (2009). Inquisitive semantics and pragmatics. Presented at the Stanford workshop on Language, Communication and Rational Agency.

Harris, J. and C. Potts (2010). Perspective-shifting with appositives and expressives. *Linguistics and Philosophy 32*(6), 523–552.

Hunter, J. (2015). *Reports in Discourse*. Ms. Toulouse.

Hunter, J., N. Asher, B. Reese, and P. Denis (2006). Evidentiality and intensionality: Two uses of reportative constructions in discourse. *Proceedings of Constraints in Discourse 2006*, 99–106.

Jäger, S. (1971). *Der Konjunktiv in der deutschen Sprache der Gegenwart*. München: Hüber.

Koev, T. (2013). *Apposition and the Structure of Discourse*. Ph. D. thesis, Rutgers.

Koring, L. (2013). *Seemingly similar: Subjects and displacement in grammar, processing, and acquisition*. Ph.D. thesis, Utrecht University.

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind & Language 30*, 235–73.

Murray, S. (2011). A Hamblin semantics for evidentials. *Semantics and Linguistic Theory 19*.

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Recanati, F. (2001). Open Quotation. *Mind 110*(439), 637–687.

Schenner, M. (2009). Semantics of evidentials: German reportative modals. *ConSOLE XVI*.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua 117*, 1034–56.

I Believe I Can φ*

Matthew Aaron Mandelkern, Ginger Schultheis and David Boylan

Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.
mandelk@mit.edu, vks@mit.edu, dboylan@mit.edu

Abstract

We propose a new analysis of ability modals. After briefly criticizing extant approaches, we turn our attention to the venerable but vexed conditional analysis of ability ascriptions. We give an account that builds on the conditional analysis, but avoids its weaknesses by incorporating a layer of quantification over a contextually supplied set of actions.

1 Introduction

Our topic is ability modals, modals of the kind found in the following sentences:

(1)   John can go swimming this evening.

(2)   Mary cannot eat another bite of this rotten meal.

(3)   Louise is able to pick Roger up from work today.

As a simple heuristic, we can identify ability modals as modals that appear in sentences that can be paraphrased ⌜S is able to φ⌝ (on its most prominent reading) or ⌜S has the ability/power to φ⌝ (or with their negations).[1]

Our topic in particular is ability modals which have a specific action—an action tagged with a specific time—as the modal's prejacent, as in (1), (2), and (3). Other ability modals, as in

(4)   Susie can swim.

(5)   Jim is able to touch his nose with the tip of his tongue.

have as their prejacent a *generic* action, one not tied to a specific time. We assume that ability ascriptions with generic actions are just specific ability ascriptions embedded under a generic operator, and thus that a semantics for generic ability ascriptions will fall out of our proposal together with a suitable semantics for the generic operator.

We will survey three extant accounts of ability modals—the orthodox account, the universal account, and the conditional account—and argue that none is satisfactory. We then propose a new account that builds on the insights of the conditional account but avoids its well-known problems.

---

*The authors would like to thank audiences at MIT and at the Semantics and Philosophy in Europe Eighth Colloquium at the University of Cambridge, three anonymous reviewers for the Amsterdam Colloquium, Alex Byrne, Kai von Fintel, Martin Hackl, Samia Hesni, Sabine Iatridou, Matthias Jenny, Justin Khoo, Milo Phillips-Brown, Kieran Setiya, Bradford Skow, Robert Stalnaker, Stephen Yablo, and each other for valuable discussion and feedback. We thank Matthias Jenny for the inspired title; unfortunately we have not been able to discover anything noteworthy about the interaction between attitude predicates and ability modals.

[1]Unless otherwise noted, all uses of 'can' and 'is able to' in examples and definitions that follow are to be read so that they can be paraphrased this way, and are to be treated as interchangeable.

Proceedings of the 20th Amsterdam Colloquium   256

2   The Orthodox Account

Kratzer (1977, 1981) gives a unified account of natural language modals according to which 'can' is an existential quantifier whose domain is the set of worlds that are 'best' according to a contextually supplied *modal base h* and *ordering source g*, both functions from worlds to sets of propositions. Given a world w, these together determine a preorder $\preceq_{g,w}$ on $\bigcap h(w)$ as follows: $u \preceq_{g,w} v$ iff $\{\psi \in g(w) : v \in \psi\} \subseteq \{\psi \in g(w) : u \in \psi\}$. Letting $\text{BEST}_{h,g,w}$ be the set of worlds in $\bigcap h(w)$ that are minimal according to $\preceq_{g,w}$, we have:

(6)   **Orthodox Account**: $[\![\text{S can } \varphi]\!]^{h,g,w} = 1$ iff $\exists w' \in \text{BEST}_{h,g,w} : [\![\varphi]\!]^{h,g,w'} = 1$.

This account of modals is widely enough accepted that it has fair claim to being orthodoxy. But it's hard to see how to implement it for ability modals. The problem is that there is no natural value for the ordering source to take that makes the right predictions about ability modals.

On the standard implementation, the modal base is circumstantial, and the ordering source takes each world to a set of propositions that 'holds fixed certain intrinsic features of the agent in question' at that world.[2] But this approach makes predictions that are much too weak. Suppose Jim and Jo are at a crucial stage in a game of darts. Jo's young child Susie exclaims

(7)   Let me play! I can hit the bullseye on this throw.

But Susie can hardly ever hit the dartboard, and she has never even gotten a dart to stick in the dartboard; she can't, isn't able to, doesn't have the ability to, hit a bullseye on this throw. (To fix intuitions, imagine that Susie does go for the shot, and that the dart falls far short of the dartboard.) In light of these facts about Susie, (7) is clearly false on its abilitative reading. But on the approach just sketched, (7) is predicted to be true, since it is *compatible* with Susie's intrinsic properties, and local circumstances, that she hit a bullseye. (Note that (7) *does* have a true reading which can be paraphrased as

(8)   It can happen that I hit the bullseye on this throw.

This is a circumstantial/metaphysical reading of 'can'. The standard proposal thus adequately accounts for this reading, but not for the prominent abilitative reading of (7), paraphrasable not as (8) but as

(9)   I'm able to hit the bullseye on this throw.)

A natural first reaction to this issue is to include in the value of the ordering source at w propositions which describe what is *normal* at w. It's not obvious this helps even in this case: hitting the bullseye is unlikely but not obviously *abnormal*. But even granting that it is abnormal in a relevant sense for Susie to hit the bullseye, this proposal makes the wrong predictions in other cases: someone can be able to do something even if doing it is highly abnormal. For example, Susie is a competent speaker of English, and thus is able to utter the sentence

(10)   Every art, inquiry, action, and pursuit is thought to aim at some good.

But, being a small child and non-philosopher, she only utters this sentence in circumstances that are, intuitively, at least as abnormal as ones in which she hits the bullseye. The present

---

[2]Vetter (2013, p. 7); see Portner (2009). The following criticism also applies if *abilities* are in the ordering source.

proposal would thus wrongly predict that (11) is false:[3]

(11)   Susie can utter (10) now.

Incorporating normality into the ordering source thus will help only if we can spell out a notion of normality that treats Susie's hitting a bullseye as abnormal, but Susie's uttering (10) as normal. We don't see a natural way to do this, and thus we don't see a natural way for the orthodox semantics to account for ability modals (which isn't to say there is *no* way for the orthodox approach to account for ability modals; see Section 6).[4]

3   The Universal Account and the Dual

In response to worries like these, some have argued that abilitative 'can' has universal, rather than existential, force.[5] We do not think such an account could work.

   To show this, we need to take a brief excursus to discuss the dual of 'can'. Some have claimed that abilitative 'can' has no dual.[6] We find this puzzling. As with any other modal operator, we can form the dual of 'can' simply by putting a negation above and below it. Intuitions suggest that both 'cannot but' and 'cannot *not*' (italics indicating stress), as in

(12)   Ginger cannot but eat another cookie right now.

are natural language realizations of this semantic pattern, and thus are both duals of 'can'.[7] 'Must' and 'have to' can also have the meaning of the dual of 'can', as in

(13)   I have to sneeze. (Kratzer, 1977)

   Why have these data generally been ignored in accounts of ability modals? We believe an infelicity of nomenclature has promoted confusion here. 'Cannot but' is not an *ability* modal; rather, it expresses compulsion of some kind, whereas ability has to do not with compulsion but with potentiality. But this does not show that 'cannot but' is not the dual of 'can': it simply illustrates the humdrum point that duals do not have the same meaning as each other, even though they may have the same subject matter. Compare the situation with deontic 'may', which is a *permission* modal. Its dual does not express permission but rather (deontic) compulsion. But it does not follow that deontic 'may' lacks a dual; rather, 'may' and 'must' are simply part of a larger unified class, namely that of deontic modals.

   Likewise, 'can' and its dual do not both express something about ability, but they nonetheless belong to a larger unified class, which we propose to call the class of *practical modals*.[8] Carefully delimiting this class is important for the study both of natural language and of traditional philosophical problems. For instance, we suspect that the modals that appear in anankastic constructions are best analyzed as practical modals, as are the modals often adverted to in

---

[3]Portner (2009)'s suggestion to make 'can' a *good possibility* modal runs into the same problem.

[4]Kenny (1976) gives a different critique of the orthodoxy, pointing out that on a modal analysis, ⌜S can (φ or ψ)⌝ ⊨ ⌜S can φ or S can ψ⌝. The account we'll give inherits this objection. However, we are not persuaded by it; we believe that a supervaluationist approach along the lines of Stalnaker (1981) can answer this criticism.

[5]E.g. Kenny (1976, p. 139), Brown (1988), Giannakidou (2001), Giannakidou and Staraki (2012).

[6]See Hackl (1998). Hackl correctly notes that it is harder than we might expect to hear an equivalence between 'can' and, e.g. 'not must not'. We are not sure what to make of this fact, but we don't take it to show that ability modals 'lack a universal dual' (Hackl, 1998, p. 10).

[7]Thanks to Stephen Yablo for pointing out the first of these.

[8]Following a suggestion of Kieran Setiya (p.c.).

philosophical discussions of freedom of will and practical necessity.[9]

Once we have the dual of 'can' clearly in sight, it is hard to see how a universal analysis of 'can' could suffice. On such an analysis 'cannot but' would have existential force;[10] this cannot be squared with the intuition that 'cannot but' expresses compulsion of some sort.

4   The Conditional Analysis

A more promising approach than the orthodox or universal accounts treats ⌜S can φ⌝ as meaning the same as ⌜S would φ if S tried to φ⌝.[11] Let $f_c(w, \psi)$ be the selection function from Stalnaker (1968)'s theory of conditionals: a contextually supplied function from a world w and proposition ψ to the 'closest' world where ψ is true, and to w iff $w \in \psi$.[12] Let '$try(S, A)$' abbreviate ⟦S tries to do A⟧^{c}. Then:

(14)   **Conditional Analysis**: ⟦S can φ⟧$^{c,w} = 1$ iff ⟦$\varphi(S)$⟧$^{c,f_c(w,try(S,\varphi))} = 1$.

φ ranges over actions, which we will model as properties; ⟦$\varphi(S)$⟧^{c} is the proposition that φ holds of S, i.e., that S does φ.

The conditional analysis (henceforth 'CA') gives much more intuitive truth-conditions than the orthodox account, and avoids the problems sketched for the orthodox account: e.g., if Susie tried to hit the bullseye, she wouldn't; but if she tried to utter (10), she would. But the CA has a number of problems of its own.

4.1   Monotonicity

First, the CA makes the wrong predictions about the monotonicity of 'can'. Intuitions about entailment suggest that 'can' is upward monotone:[13]

(15)   John can swim the butterfly tonight.

seems to entail

(16)   John can swim tonight.

Facts about NPI-licensing support this impression. 'Cannot' licenses NPIs, as in

(17)   John can't find any dance partners at this party.

According to a leading theory of NPI-licensing, the Fauconnier-Ladusaw analysis,[14] NPIs are licensed only in downward monotone environments; and 'not can' is downward monotone iff 'can' is upward monotone.

The CA, however, predicts that 'can' is non-monotone, not upward monotone. Some will take this to be a virtue; indeed, it's not clear whether a sentence like

(18)   I can ride a bike with training wheels right now.

---

[9]See e.g. von Fintel and Iatridou (2005), Aristotle (2002), Frankfurt (1969), Williams (1982), Setiya (2007).

[10]Surprisingly, Brown (1988) takes this prediction on board, claiming that the dual of 'can' is 'might'.

[11]An old philosophical idea, traceable to Hume (1748), taken up by Moore (1912) a.o.; for formulations in a model-theoretic framework, see especially Lehrer (1976), Cross (1986), and Thomason (2005). We use 'try' in an inclusive sense which treats things like cars and elevators as the kinds of things that can try.

[12]Stalnaker's uniqueness assumption is crucial for accounting for the negation and dual of 'can'.

[13]M is upward monotone iff if $\rho \vDash \psi$, then $M(\rho) \vDash M(\psi)$; downward monotone iff if $\rho \vDash \psi$, then $M(\psi) \vDash M(\rho)$.

[14]See e.g. Ladusaw (1979), von Fintel (1999).

entails

(19)   I can ride a bike right now.

We suspect that data like this can be explained pragmatically, however. For note that someone who can only ride a bike with training wheels could well insist

(20)   I can ride a bike with training wheels now, so *technically*, I can ride a bike now.

This discussion is inadequate to the complexity of the issues at hand, but we think the data suggest, *pace* the CA, that 'can' is upward monotone.

4.2   Dual

Second, the CA makes the wrong predictions about the dual of 'can':

(21)   **Dual (Conditional Analysis)**: $[\![S \text{ cannot but } \varphi]\!]^{c,w} = 1$ iff $[\![\varphi(S)]\!]^{c,f_c(w,try(S,\bar{\varphi}))} = 1$.

i.e. iff the closest world where S tries to not φ is one where S still φ's. This is too weak. Intuitively ⌜S cannot but φ⌝ means not only that S φ's if she tries not to, but that she φ's *no matter what she tries to do*. Another way to put the point is that the CA wrongly predicts that ⌜S cannot but *not* φ⌝ and ⌜S cannot but φ⌝ are consistent.

4.3   Counterexamples

Finally, the CA faces a number of related counterexamples.

First, there are cases in which the CA predicts that ⌜S can φ⌝ is true when it's false.[15] Suppose that John is planning to go to a movie alone. He has no special commitment to go—he has simply decided to. Ann asks John out to dinner; he replies:

(22)   I'm sorry, I'm not able to go; I'm going to a movie.

There is a prominent reading on which (22) is true. But if John *tried* to have dinner with Ann, he would succeed. So the CA cannot predict the true reading of (22). Importantly, 'able to' is abilitative on this reading, not deontic or bouletic: the retraction data associated with (22) differ from the retraction data we would expect if it were a deontic or bouletic modal. When someone makes a deontic or bouletic claim, rejoining with a claim about abilities feels like a non-sequitur. But in the present case, pointing out that John really does have an ability to go to dinner feels like a natural and effective response (as in (30) below), not a non-sequitur.

Second, there are cases in which the CA predicts ⌜S can φ⌝ is false when it's true.[16] Suppose Jones is a skilled golfer with an easy shot to make. Matt says:

(23)   Jones is able make this shot right now.

Matt has intuitively said something true. Now suppose Jones takes the shot and misses. We still judge Matt to have said something true. Afterwards, we can truly say

(24)   Jones (was able to/had the ability to) make that shot at that time.[17]

---

[15]Chisholm (1964), Lehrer (1968), Thomason (2005). Cf. cases with a phobic or comatose agent.

[16]From Austin (1961).

[17]There is also a false reading of (24), brought out when 'was able to' has perfective aspect (see Bhatt (1999) a.o.), but all that matters for our purposes is that there is a true reading, brought out when 'was able to' has imperfective aspect. Thanks to Nilanjan Das and Raphaël Turcotte for data in Hindi, Bengali, and French.

Yet given how f_c is defined, the closest world where Jones tries to make the shot is the actual world. Since Jones misses in the actual world, the CA thus wrongly predicts that (23) is false. Another way to put this point is that

(25)   Jones is able to make this shot right now, though if he tries of course he might miss.

is felicitous in some cases (as in this one), but is predicted by the CA to be always infelicitous.[18]
      Finally, there are cases in which an agent can do something, but not if she tries to do it:[19]

(26)   David can breathe normally for the next five minutes.

is true, but if David tried to breathe normally, he would end up breathing abnormally.
      From a technical point of view, these cases are easy to respond to: simply change the selection function relevant for ability ascriptions so that it selects worlds in a way that matches our intuitions.[20] The problem with this response is that it uncouples the CA from the analysis of conditionals, and thus from our intuitions about conditionals and similarity in general.[21] Without an intuitive characterization of the altered selection function, the resulting theory is not particularly predictive or explanatory. We believe we can do better.

5   Our Proposal

Examples like those discussed in the last section have been taken by many to refute the CA.[22] But we think that the CA is on the right track. It rightly captures the *hypothetical* nature of abilities: whether you are able to perform a particular action depends in some way on what happens under relevant alternate circumstances. Our account of ability modals aims to preserve this insight, but avoid the problems discussed in the last section by adding a layer of quantification over a contextually supplied set of actions to the meaning of 'can'.
      Let $\mathcal{A}_{c,S}$ be a set of actions that are—in a sense to be precisified—*practically available* to an agent S in a context c. With f_c the selection function as above, we propose:

(27)   **Our Proposal**: $[\![S \text{ can } \varphi]\!]^{c,w} = 1$ iff $\exists A \in \mathcal{A}_{c,S} : [\![\varphi(S)]\!]^{c,f_c(w,try(S,A))} = 1$.

I.e. ⌜S can φ⌝ is true just in case there is some contextually salient action A such that the closest world where S tries to do A is a world where S does φ.[23]
      At a first pass, we may assume that in many cases, if an ability ascription has the form ⌜S can φ⌝, then $\mathcal{A}_{c,S} = \{\varphi, \bar{\varphi}\}$. In those cases the predictions of our account come very close to those of the CA. Thus, e.g., if we make this assumption in evaluating

(28)   Louis can go for a swim in the MIT pool tonight.

---

[18]We can make the inverse point as well. Suppose that young Susie by chance hit a bullseye; there is a reading on which 'Susie was able to make the bullseye in this game' remains false (a reading easiest to bring out with imperfective morphology), contrary to the predictions of the CA.

[19]See Vranas (2010) for discussion.

[20]Thomason (2005) suggests a response along these lines in response to (22).

[21]It is a non-negotiable property of similarity that nothing is more similar to something than itself, a thesis the altered selection function would have to abandon.

[22]See Austin (1961), Lehrer (1968), van Inwagen (1983).

[23]Chisholm (1964) makes a similar suggestion. As far as we know his suggestion hasn't been taken up in the subsequent literature, perhaps because he himself sketches a fairly serious objection to it; our account, however, avoids that objection by restricting the set of actions we quantify over.

then the action *going for a swim in the MIT pool tonight* is included in $\mathcal{A}_{c,Louis}$; thus provided that Louis can go swimming tonight if he tries to do so, we predict that (28) is true.

$\mathcal{A}_{c,S}$ won't always be set this way, however. We argue that flexibility in how this parameter is set by context crucially allows us to avoid the problems discussed in the last section: to avoid the counterexamples and make plausible predictions about the dual and monotonicity of 'can'.

5.1 Improved Predictions

The most important benefit of our view is that it avoids the counterexamples discussed in Section 4.3. Consider first cases in which ⌜S can φ⌝ is intuitively false, even though it is true that S would φ if S tried. Recall John, who says

(29)   I'm not able to go [to dinner]; I'm going to a movie.

Unlike the CA, our account can predict that (29) has a prominent true reading. To be sure, if John *tried* to go to dinner, he'd succeed. But, on our proposal, this does not guarantee the truth of 'John can go to dinner' at a context c: the action *meeting Ann for dinner* (or something much like it) must also be in $\mathcal{A}_{c,John}$. That is, it must be treated as *practically available* to John in a relevant sense in the context of utterance. If this condition is not met, then there is no action in $\mathcal{A}_{c,John}$ such that trying to do it guarantees that John meets Ann for dinner, and so (29) is true.

And indeed, we argue that this condition is not met in this case. An action may be treated as unavailable in a context for a wide variety of reasons. In this case, the fact that the agent has decided against the action suffices to rule it out. In other cases, an action may be treated as unavailable because it takes place in the past relative to the time of evaluation; because it violates certain rules that the agent in question is bound by; or because the agent is psychologically or physiologically unable to even consider it (as if John were phobic or comatose). Different notions of availability are at work in different contexts. Changes in the set $\mathcal{A}_{c,S}$ reflect changes in the relevant notion of availability.

Circularity threatens here. Among other things, facts about S's *abilities* might influence the constituents of $\mathcal{A}_{c,S}$. If we were aiming to give a reductive analysis of ability, we would want to identify the constituents of $\mathcal{A}_{c,S}$ without reference to abilities. But, like most projects of semantic analysis, our present aim is elucidation rather than reductive analysis (though we don't want to suggest that such a reductive analysis is impossible in the present framework).

One way to test the plausibility of our account is to see whether insisting on the availability of the action *meeting Ann for dinner* can modulate intuitions in this case. Suppose Ann replies:

(30)   Of course you *can* meet me—just skip the movie and come to dinner!

It seems that Ann has said something true: John *can* meet her for dinner. We hypothesize that, in responding this way, Ann ensures that $\mathcal{A}_{c,John}$ include *meeting Ann for dinner*, and so (30) comes out true.

$\mathcal{A}_{c,S}$ also plays a crucial role in responding to cases in which ⌜S can φ⌝ is intuitively true, even though it is false that S would φ if S tried. Recall the golf case. We said that

(31)   Jones can make this shot.

is intuitively true. How can we predict this? Suppose Jones aimed to the left of the pin; had he aimed to the right, he would have made the shot. Let the action *aiming to the right* be in $\mathcal{A}_{c,Jones}$. Then we predict that 'Jones can make this shot' is true even though Jones actually misses. This looks intuitive: indeed, we say of Jones

(32)   Well, he *could* have made the shot, if he had only aimed to the right.

This move can be applied quite broadly. We often ascribe abilities to agents even when they are not certain to succeed at a given action should they try, and even in cases where they fail when they in fact try.[24]

A worry about over-generation arises. Bob, a lousy chess player, is playing white in a chess match against Kasparov. Let $\mathcal{A}_{c,Bob}$ include, for all possible sequences of moves for white in the game, the action of taking that sequence of moves. One of these is such that, given Kasparov's course of play, if Bob completes that sequence of moves, he will beat Kasparov. It would seem to follow that

(33)   Bob can beat Kasparov in this match.

is true. But (33) is false (on its prominent reading).[25] To predict this, we propose to place certain default constraints on what goes into $\mathcal{A}_{c,S}$—roughly, that an action is in $\mathcal{A}_{c,S}$ only if the agent in question has the right kind of *epistemic* access to that action: perhaps, only if she knows that that action is a way for her to do φ. There is some sequence of moves that would beat Kasparov. But Bob doesn't *know which* sequence it is. This means that it is not in $\mathcal{A}_{c,Bob}$, which explains why we judge (33) false.[26]

Finally, appeal to $\mathcal{A}_{c,S}$ lets us explain why 'David can breathe normally for the next five minutes', and sentences like it, are true: there is *something* relevant (let us suppose) such that if David tries to do *that*, he breathes normally (say, *working on a problem set*).

The notion of practical availability that determines $\mathcal{A}_{c,S}$ is determined by context. As with other quantificational structures in natural language, change the context, and you change the domain of quantification, and thus the relevant notion of availability. Much more needs to be said to flesh out a complete meta-semantic story about how $\mathcal{A}_{c,S}$ is determined, and the plausibility of our view will turn on the ultimate success of such a project. But we take the present discussion to show that whether or not ⌜S is able to φ⌝ is taken to be true in a context c depends not just on what would happen if S tried to φ, but also on which ways, if any, of trying to φ are treated as practically available to S in c.

5.2   The Dual

Our approach makes good predictions about the meaning of the dual of 'can':

(34)   **Dual (Our Proposal)**: $[\![\text{S cannot but } \varphi]\!]^{c,w} = 1$ iff $\forall A \in \mathcal{A}_{c,S} : [\![\varphi(S)]\!]^{c,f_c(w,try(S,A))} = 1$.

Informally: for every action A available to S in c, S does φ in the closest world in which S tries to do A. In other words, no matter what S tries to do (among the actions we are treating as practically available in c), S ends up doing φ. Consider:

(35)   Ginger cannot but eat another cookie right now.

(35) says that Ginger is *compelled* to eat another cookie: *no matter what she tries to do*, she'll end up eating another one. This is precisely what we predict. And unlike the CA, our account of the dual rightly predicts that ⌜S cannot but *not* φ⌝ and ⌜S cannot but φ⌝ are inconsistent, provided that $\mathcal{A}_{c,S}$ is non-empty, a condition that, plausibly, is satisfied in most cases.

---

[24]One might be tempted to adopt a *graded* proposal to accommodate this kind of case. We do not see a fruitful way to spell out a view like that, though, and the present response seems satisfactory to us.

[25]Though making salient the sequence of moves in question can make prominent a true reading, as we predict.

[26]This also explains why flukey successes don't always merit ability ascriptions; see Footnote 18.

5.3 Monotonicity

Finally, 'can' is upward monotone on our account, matching intuitions and NPI-licensing.

6 Comparison and Conclusion

Our account is less of a departure from the orthodox Kratzerian approach to natural language modality than the CA. In keeping with orthodoxy, we say that 'can' denotes an existential quantifier. We depart from orthodoxy, however, in saying that 'can' quantifies over a set of actions (i.e. properties), not a set of worlds. That said, a number of recent proposals have departed from orthodoxy in a similar way, and converged on structurally similar accounts of natural language modals. Our account of ability modals differs from these because of the role that the selection function plays in the view. But it shares with them a central structural feature: quantification over sets of properties.[27]

It is too early to decide whether this approach will supplant orthodoxy, however, so it is also worth asking whether our view can be reframed within the orthodox framework. The answer is 'yes'. For any agent S, just let the modal base take any world w to $\varnothing$ and let the ordering source take any world w to $\{\{w'\} : \exists A \in \mathcal{A}_{c,S}(w' = f_c(w, try(S, A)))\}$. This account makes the same predictions as ours.[28] Should we adopt it?

We think not. If there were conclusive evidence in favor of the orthodox account for all other modals, then it would make sense to formulate our view in its terms. But short of that, we think there is no reason to go that way, because our approach fits far more naturally with a meta-semantic story.[29] A meta-semantics for both our view and its recasting in the orthodox framework must account for how $\mathcal{A}_{c,S}$ and f_c are determined, since both views refer to these parameters. But the orthodox recasting of our account *also* makes reference to (what looks to us like) a highly gerrymandered ordering source and an empty modal base. It is hard to see how this ordering source could play any role in an explanatory meta-semantics. So we think there is reason simply to dispense with it, as our account does.

We should also ask whether there are other elements of semantic machinery that can 'see' the differences between the semantic entries and thus distinguish them on more straightforward predictive grounds. We think there are; in particular, research into epistemic modals has suggested that embedded epistemic modals quantify over their *local contexts*.[30] Under this assumption, the behavior of epistemic modals embedded under practical modals may help decide between the two views under consideration.

The plausibility of our semantics will ultimately turn on this and a number of other open questions. In addition to those highlighted above, these include the question of whether we need to encode a non-accidental connection between S and φ in order for $\ulcorner$S can $\varphi\urcorner$ to come out true; how to account for graded ability ascriptions like 'Jones can easily make this shot'; and how to account for actuality implications that arise from the interaction of ability modals and aspect.

---

[27]See Yalcin (2012) and Willer (2013) on attitude verbs; Cariani et al. (2013) and Cariani (2013) on deontic modals; and Villalta (2008) on verbs of desire. Officially the quantification in these proposals is over sets of propositions, but this is a superficial modeling difference—propositions are just zero-place properties.

[28]*Modulo* embedded epistemic modals; see below.

[29]This is the same kind of choice point that will decide the parallel question in other cases, as between, for instance, the approach of Cariani et al. (2013) versus Kratzer's approach.

[30]Veltman (1996), Stephenson (2007), Yalcin (2007), Klinedinst and Rothschild (2012), Stalnaker (2014) a.o.

References

Aristotle (2002). *Nicomachean Ethics*. Oxford University Press, Oxford.

Austin, J. L. (1961). Ifs and cans. In *Philosophical Papers*, pages 151–180. Oxford University Press, London.

Bhatt, R. (1999). Ability modals and their actuality entailments. In Shahin, K., Blake, S., and Kim, E.-S., editors, *WCCFL 17 Proceedings*, volume 17, pages 74–87.

Brown, M. A. (1988). On the logic of ability. *Journal of Philosophical Logic*, 17:1–26.

Cariani, F. (2013). 'Ought' and resolution semantics. *Nous*, 47(3):534–558.

Cariani, F., Kaufmann, M., and Kaufmann, S. (2013). Deliberative modality under epistemic uncertainty. *Linguistics and Philosophy*, 36:225–259.

Chisholm, R. M. (1964). J. L. Austin's philosophical papers. *Mind*, 73(289):1–26.

Cross, C. B. (1986). 'Can' and the logic of ability. *Philosophical Studies*, 50(1):53–64.

von Fintel, K. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148.

von Fintel, K. and Iatridou, S. (2005). What to do if you want to go to Harlem: Anankastic conditionals and related matters. MS, MIT.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy*, 66(23):829–839.

Giannakidou, A. (2001). The meaning of free choice. *Linguistics and Philosophy*, 24:659–735.

Giannakidou, A. and Staraki, E. (2012). Ability, action, and causation: from pure ability to force. In Mari, A., Beyssade, C., and Prete, F. D., editors, *Genericity*. Oxford University Press.

Hackl, M. (1998). On the semantics of 'ability attributions'. Master's thesis, MIT.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Oxford University Press, Oxford.

Kenny, A. (1976). Human abilities and dynamic modalities. In Manninen and Tuomela, editors, *Essays on Explanation and Understanding*, pages 209–232. D. Reidel.

Klinedinst, N. and Rothschild, D. (2012). Connectives without truth tables. *Natural Language Semantics*, 20:137–175.

Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, 1(337-355).

Kratzer, A. (1981). The notional category of modality. In Eikmeyer, H. and Rieser, H., editors, *Words, Worlds, and Contexts: New Approaches in Word Semantics*. de Gruyter.

Ladusaw, W. (1979). *Polarity Sensitivity as Inherent Scope Relation*. PhD thesis, University of Texas at Austin.

Lehrer, K. (1968). Cans without ifs. *Analysis*, 29(1):29–32.

Lehrer, K. (1976). 'Can' in theory and practice: A possible worlds analysis. In Brand, M. and Walton, D., editors, *Action Theory*, pages 241–270. D. Reidel.

Moore, G. (1912). *Ethics*. Williams and Norgate, London.

Portner, P. (2009). *Modality*. Oxford University Press, New York.

Setiya, K. (2007). Cognitivism about instrumental reason. *Ethics*, 117:649–673.

Stalnaker, R. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory*, pages 98–112. Oxford: Blackwell.

Stalnaker, R. (1981). A defense of conditional excluded middle. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Ifs: Conditionals, Beliefs, Decision, Chance, and Time*, pages 87–105. D. Reidel.

Stalnaker, R. (2014). *Context*. Oxford University Press.

Stephenson, T. (2007). A parallel account of epistemic modals and predicates of personal taste. In Puig-Waldmueller, editor, *Proceedings of Sinn and Bedeutung*, volume 11, pages 583–97.

Thomason, R. H. (2005). Ability, action, and context. MS.

van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press, New York.

Veltman, F. (1996). Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261.

Vetter, B. (2013). 'Can' without possible worlds: Semantics for anti-Humeans. *Philosophers' Imprint*, 13(16):1–27.

Villalta, E. (2008). Mood and gradability: an investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, 31:467–522.

Vranas, P. (2010). What time travelers may be able to do. *Philosophical Studies*, 150:115–121.

Willer, M. (2013). Dynamics of epistemic modality. *Philosophical Review*, 122(1):45–92.

Williams, B. (1982). Practical necessity. In *Moral Luck*, pages 124–131. Cambridge University Press.

Yalcin, S. (2007). Epistemic modals. *Mind*, 116(464):983–1026.

Yalcin, S. (2012). Context probabilism. In *Proceedings of the 18th Amsterdam Colloquium*, pages 12–21.

The imperfective in subjunctive conditionals: fake or real aspect? *

Fabienne Martin

University of Stuttgart
`fabienne.martin@ling.uni-stuttgart.de`

Abstract

This paper aims to provide a 'real aspect' approach of the 'fake' imperfective in subjunctive conditionals and a new account of the (non)-cancellability of the counterfactual inference in SCs, largely based on Ippolito 2013. It is argued that PAST and PRES above MODAL in conditionals compete the same way as in non-modal stative sentences, see Altshuler and Schwarzschild 2012. On this view, the counterfactual inference of SCs, when cancellable, is nothing else than the cessation implicature routinely triggered by past stative sentences.

1 Background and goals

This paper is dedicated to the interpretation of tense and aspect morphology in **subjunctive conditionals** (**SC**s). SCs are often built by using one or two additional layer(s) of past tense morphology on top of the regular tense morphology found in the corresponding **indicative conditional** (**IC**). Additional layer(s) of tense morphology characteristic of SCs are said to be 'fake', because they do not locate events in time. I will call '+ 1 past' SCs (resp. '+ 2 past SCs') those SCs that add one (resp. two) layer(s) of past tense morphology on top of the tense morphology locating events in time, see (1)-(3).

(1) If John runs the marathon next spring, he will win. IC

(2) If John ran$_{+1 \; PAST}$ the marathon next spring, he would$_{+1 \; PAST}$ win. '+1 past' SC

(3) If John had run$_{+2 \; PAST}$ the marathon next spring, he would have won$_{+2 \; PAST}$. '+2 past' SC

In several languages like e.g. French, Greek, Italian or Hindi, the first additional layer of past tense is realized (a.o.) with the past imperfective aspect (IMP), cf. Iatridou 2000, 2010; see e.g. (4). Note that the French *conditionnel* (COND) found in the consequent of SCs combines the imperfective morphology *-ai-* with the future morphology *-r-*, cf. Iatridou 2000.

(4) *Si John courait$_{+1 \; PAST}$ le marathon demain, il gagnerait$_{+1 \; PAST}$.*
     If John run-IMP.1SG the marathon tomorrow, he win-COND.1

     'If John ran the marathon tomorrow, he would win.'

Why is it that IMP makes this contribution to **counterfactuality** (**CF**) rather than the perfective (PFV) (in the relevant languages)? As Iatridou 2010 observes, this crosslinguistic generalization can be explained

---

in two ways. According to the 'fake aspect' approach, IMP *qua* aspect marker makes no semantic contribution to CF.[1] According to the 'real aspect' approach, IMP makes a substantial semantic contribution in SCs through the *aspectual* properties by which IMP differs from PFV. As Iatridou 2010, 14 notices, such an approach has not been much explored yet: "[We] have found no obvious way to extend [the meaning of IMP] to cover [counterfactual conditionals]".

The goal of this paper is twofold. Firstly, it aims to provide a 'real aspect' approach of the 'fake' imperfective in SCs, on the basis of French data. Secondly, it offers a new account of the (non)-cancellability of the CF inference in SCs, given the semantic contribution of aspectual operators (i.e. IMP) in these conditionals. I will argue that in languages like French, IMP is interpreted in SCs the same way as outside conditionals. However, I do not claim that the imperfective semantics is a necessary component of SCs. In fact, there is cross-linguistic evidence that it is not (see Halpert and Karawani 2012, Bjorkman and Halpert 2013, Karawani 2014). Besides, even in French, the imperfective semantics is perhaps not involved in *any* use of IMP, included *outside* conditionals (see section 6). Also, note that the 'real *aspect*' approach of IMP is independent from the 'real *past*' approach of the same morphology, according to which the additional layer of *past* in SCs is a real past. I defend both approaches here (IMP is both a real imperfective and a real past in SCs), but several authors defend the latter without committing about the former.

The paper is structured as follows. In sections 2-4, I present Ippolito's 2013 analysis and its shortcomings. Sections 5 and 6 present my proposal.

2 One vs. two additional layers of ('fake') past

2.1 The 'real past' approach of SCs

In this paper, I adopt the 'real past' approach of standard PSCs, see a.o. Dahl 1997, Ippolito 2003, 2013, Arregui 2005, Romero 2014. According to this view, the 'fake' layers of past in SCs are in fact real insofar they express temporal precedence like in the regular use of the past. However, the additional past morphology is not supposed to be interpreted within the 'bare' conditional (the structure consisting of the modal operator, the *if*-clause acting as its restriction and the consequent acting as its nuclear scope), but rather outside the *if-clause* and contributes to the interpretation of the modal:

(5) PAST[MODAL [PAST][PAST]]
     '*fake*'   *real*

The intuition behind this idea is that we evaluate SCs as if we returned to a past time —Dahl's 1997 'choice point' — at which it was still possible that the antecedent would come true, and looked at possible futures with respect to that past. Among these approaches, I adopt Ippolito's 2013 framework because it explicitly differentiates between cancellable and noncancellable CF.

2.2 Ippolito's typology of SCs

Ippolito 2013 assumes that what I call '+1 past' SCs are modal structures under a universal present perfect, cf. (6a). What I call '+2 past' SCs are modal structures under a universal past perfect, cf. (6b).

(6) a. PRES[PERF[$\forall_\subseteq$[WOLL[SIM[HIST p]][q]]]]   '+1 past' SCs

[1]For instance, it has been argued that IMP is chosen over PFV because (i) IMP is a cross-linguistically default aspect (Iatridou 2010), (ii) IMP is compatible with other ingredients that prove to be necessary to CF (and not because IMP is necessary to CF), see Halpert and Karawani 2012, or (iii) IMP is aspectually underspecified (it can have both perfective and imperfective interpretations), while PFV only has perfective interpretations (see Bjorkman and Halpert 2013).

b. PAST[PERF[$\forall_\subseteq$[WOLL[SIM[HIST p]][q]]]]   '+2 past' SCs

In both cases, when PERF combines with $\forall$, we obtain a 'perfect interval' t', such that for all subintervals t'' of t', the conditional proposition is true at t''. The right boundary of the perfect interval is the utterance time (UT) for '+1 past' SCs and a contextually salient past time for '+2 past' SCs. The (very much simplified) truth conditions for the '+1 past' SC (2) are in (7), and those for the '+2 past' SC (3) are in (8) (For lack of space, I have to refer to Ippolito 2013 for the details of the analysis).

(2)   If John ran the marathon next spring, he would win.   '+1 past' SC

(7)   true if $\exists t'$ such that the right boundary of t'=UT, and $\forall t'' \subseteq t'$, it is the case that all possible
       worlds historically accessible from the actual world at t'' maximally similar to the actual world
       and where John will run the marathon next spring are worlds where he will win.

(3)   If John had run the Marathon next spring, he would have won.   '+2 past' SC

(8)   true if $\exists t'$ whose right boundary is a salient past time, such that $\forall t'' \subseteq t'$, it is the case that all
       possible worlds historically accessible from the actual world at t'' maximally similar to the
       actual world and where John will run the marathon next spring are worlds where he will win.

The temporal schemata Ippolito attributes to '+1 past' SCs includes a first past component because the perfect interval extends before the UT, cf. (9). The one attributed to '+2 past' SCs has a second past component that shifts this right boundary to a past time, cf. (10).

(9)   Temporal structure for '+1 past' SCs
       PAST$_1$ [MODAL[PRES/PAST$_2$ p] [PRES/PAST$_2$ q]]

(10)   Temporal structure for '+2 past' SCs
       PAST$_1$ [PAST$_2$ [MODAL [PRES/PAST$_3$ p] [PRES/PAST$_3$ q]]]

On the other hand, MODAL is in the scope of PRES in ICs:

(11)   Temporal structure for ICs
       PRES [MODAL[PRES/PAST$_1$ p] [PRES/PAST$_1$ q]]

Note that when a '+2 past' SC describes events in the past, we add two layers of 'fake' past (above MODAL) to one layer of real past (below MODAL), cf. (10). Thus, we in principle end up with three layers of past. However, neither standard English nor standard French has a regular form that expresses three pasts within the same clause. I therefore assume (like Ippolito 2013 and Iatridou 2000 for English) that SCs about past events are expressed by the same form (a double past), whenever they instantiate a '+1 past' or a '+2 past' SC (see Iatridou 2000, 252, fn. 26, who analyzes the latter case as an instance of haplology).

3   CF is cancellable in '+1 past' SCs only

As is well-known, the inference of CF of many SCs can be canceled. SCs *à la* Anderson 1951 are of this type, cf. (12).

(12)   If John had taken arsenic, he would have shown exactly the symptoms that he has now.
       $\not\succ$ John did not take arsenic.

Ippolito 2013 proposes an important generalization, that I will dub 'Ippolito's generalization':

(13) *Ippolito's generalization*: CF is cancellable in '+1 past' SCs, but not in '+2 past' SCs.

Ippolito's first piece of evidence for (13) is provided by SCs with three layers of past found in some American and English dialects. In SCs of this type, two (fake) layers of past surface on top of regular past tense morphology, cf. (14).

(14)   a. If he knew she was coming, he stayed home.   IC

         b. If he hadd-*a* known$_{+2\ PAST}$ she was coming, he would-*a* stayed$_{+2\ PAST}$ home. '+ 2 past' SC

Crucially, the inference of CF of SCs like (14b) is *not* cancellable, see Dancygier and Sweetser 2005, Biezma et al. 2013, Ippolito 2013. This is illustrated in Biezma et al.'s 2013 example (15).

(15)   #If Jones had've taken$_{+2\ PAST}$ arsenic, he would have shown$_{+2\ PAST}$ exactly those symptoms that he in fact shows (so, he probably took arsenic).

Ippolito's second piece of evidence for (13) deals with SCs about future events built with a pluperfect. These SCs are unambiguously '+2 past' SCs, since none of the two layers of past locates the event within time. And as Ippolito observes, these SCs resist Anderson-like attempts to cancel the inference of CF, see her slightly modified example (16).

(16)   If Charlie had gone to Boston by train tomorrow, # Lucy would have found in his pocket the ticket that she in fact found. So, he will go to Boston by train tomorrow.

4   Ippolito's account for Ippolito's generalization

4.1   Cancellability of the CF inference in '+1 past' SCs

Ippolito 2013 accounts for the cancellability of the CF inference in '+1 past' SCs in a straightforward way. She argues that the 'No Empty Restriction' requirement ('The restriction of a quantifier cannot be empty') accompanying the modal is only a pragmatic constraint designed to avoid vacuously true assertions. This means, in practice, that it is only required that there be *some* subinterval *t''* of *t'* when at least some antecedent-world is historically accessible. Therefore, this subinterval *t''* *can* be the UT (i.e. the right boundary of *t'*), but does not have to. In case *t''*≠UT *only*, the antecedent is counterfactual.

One problem of this account, as Ippolito 2013 herself recognizes, is that languages like Italian (and French) do *not* have a present perfect in the English sense (whose right boundary is given by UT). Why would they build such a perfect just in SCs?

4.2   Noncancellability of the CF inference in '+2 past' SCs

Ippolito's 2013 account for the noncancellability of the CF inference in '+2 past' SCs is more complicated. Briefly, she argues that this inference is conveyed by antipresupposition, and assumes that "unlike scalar conversational implicatures, antipresuppositions cannot be suspended or canceled" (p. 92). Before presenting her argument into more detail, let me briefly recall what an antipresupposition is, through the lexical scale <*both, all*>. On this scale, *both* is 'stronger' than *all* in that it carries the stronger presupposition that the domain of quantification has only two members. The oddity of (18a) is due to the fact that the maximize presupposition principle (Heim 1991, Chemla 2008 a.o.) is not respected.

(17) *Maximize Presupposition:* Among a set of alternatives, use the felicitous sentence with the strongest presupposition.

(18) a.#Mary broke all her arms.   b. Mary took all her children to the park.

(19) Peter believes that I am against the embargo.

A felicitous occurrence of the sentence with the weaker presupposition (see (18b)) is said to convey, by antipresupposition, that the speaker believes that the strongest presupposition of the competing sentence is not fulfilled. For instance, (18b) conveys by antipresupposition that the speaker believes that it is not true that Mary has two children. The same way, (19) conveys that the speaker believes that the stronger presupposition of the competing verb *know* is not fulfilled (i.e., (19) conveys that the speaker believes it to be false that he is against the embargo).

Ippolito 2013 argues that '+ 1 past' SCs and '+ 2 past' SCs compete the same way, with the former triggering the strongest presupposition. Let me illustrate through (20) which presuppositions Ippolito 2013 assumes to be triggered by '+1 past' SCs.

(20) John died. #If he wrote a book, it would be a success. (Ippolito 2013)

Ippolito's first assumption is that the predicate *write a book* triggers an existence presupposition, namely, that its subject's referent exists at event time. For instance, in (20), the antecedent p presupposes that John is alive now. She further assumes that '+1 past' SCs require that the presuppositions of the antecedent p ($ps(p)$) be compatible with the set of worlds historically accessible at UT. For instance, the SC in (20) requires that 'John is alive now' is a possibility at UT. But this enters in contradiction with the given context, hence the problem of (20). Observe, now, that the oddity of (20) vanishes in the corresponding '+ 2 past' SC, see (21):

(21) John died. If he had written a book, he would have been a success.

For Ippolito, this is because the presupposition triggered by '+2 past' SCs is *weaker*: those only require that $ps(p)$ be compatible with the set of worlds historically accessible *at some past time*. Since (21) only requires that it was a possibility *in the past* that John is alive now, the problem vanishes.

According to this analysis, the presupposition of '+1 past' SCs is stronger than the one of the corresponding '+2 past' SC: The set of possibilities shrinks over time; therefore, being compatible with the set of worlds historically accessible at UT — as required by '+1 past' SCs – entails being compatible with the set of worlds historically accessible at some past time preceding UT – as required by '+2 past' SCs. But the reverse is not true.

Therefore, choosing a '+2 past' SC — the weakest alternative — conveys by antipresupposition that the speaker believes that the presupposition of the '+ 1 past' SC — the strongest alternative — is not satisfied. For instance, (21 conveys that $ps(p)$ ('John is alive now') is not a possibility at UT.[2]

The account just sketched faces two important problems. Firstly, Ippolito's assumption that antipresuppositions are not cancellable seems to rely on the view that antipresuppositions have a presuppositional essence. But this is debatable; Chemla 2008 argues explicitly against this view, and describes cases where antipresuppositions are suspended when the assumption of authority and/or competence is not fulfilled. In fact, antipresuppositions seem to be cancellable even when the assumptions of authority and competence are fulfilled, see e.g. (22).

(22) Mary believes that I am against the embargo. Well, she is totally right!

The second problem of Ippolito's account concerns her claim that the presuppositions of the antecedent of '+1 past' SCs always have to be compatible with the set of worlds historically accessible at UT. Sentences (23) and (24) are perfectly acceptable although 'Agatha Christie/Granny is alive now' (the presupposition of existence of the antecedent) is not a possibility at UT.

---

[2]Several details are omitted here: Chemla's 2008 *competence assumption* and *authority assumption* are also needed in order to obtain this result.

(23) If Agatha Christie wrote a detective novel today, she would make use of the possibilities for criminals offered by new technologies.

(24) Granny died already ten years ago! But if she were in the kitchen with us right now, she would sing this song out loud.

To be sure, a contrast marker like *but* seems required in (24). But this is routinely the case when a defeasible inference is canceled. In fact, to my ears, even (20) becomes much better once such a contrastive marker is introduced.

5   Proposal for '+1 past' SCs

In this section, I propose a new account for the cancellability of the CF inference in '+ 1 past' SCs. I start with two assumptions. Firstly, I assume that '+1 past' SCs are modals embedded under PAST rather than under the present perfect. Ippolito's structure (6a) is minimally modified as in (25).

(25)   PAST[$\forall_{\subseteq}$[WOLL[SIM[HIST p]][q]]]   '+1 past' SCs

Secondly, I adopt Ferreira's 2014 proposal that in conditionals, MODAL is a stativizer. Consequently, the aspect above MODAL attaches to a stative predicate. The running time ($\sigma(s)$) of the described state is the interval during which the bare conditional is true, see the modified truth-conditions (26) for (2).

(26)   true if $\exists s$ with $t' = \sigma(s)$ such that t' is a past interval, and $\forall t'' \subseteq t'$, it is the case that all possible worlds historically accessible from the actual world at t'' maximally similar to the actual world and where John will run the marathon next spring are worlds where he will win.

Next, I follow Ippolito 2013 or Leahy 2011, 2015 on the view that in some cases at least, the CF inference arises from the competition between two forms.[3] However, differently from them, I argue that this CF inference arises from the competition between *('fake') tenses* above MODAL, namely PAST (in '+ 1 past' SCs) and PRES (in the corresponding ICs). More precisely, the idea is that PAST and PRES above MODAL in conditionals compete the same way as in non-modal stative sentences, see Altshuler and Schwarzschild 2012. On this view, the CF inference of '+1 past' SCs is nothing else than the cessation implicature routinely triggered by past stative sentences. This is the reason why it is generally cancellable. Besides, according to this analysis, in languages like French, IMP makes the same semantic contribution above MODAL in conditionals as in stative sentences outside conditionals.

Before spelling out the proposal further, let me first briefly summarize Altshuler and Schwarzschild's 2012 account of past vs. present stative sentences and see how it applies to languages like French, that forces to choose between two past morphologies in stative sentences, too.

5.1   Past stative sentences and cessation implicature

Altshuler and Schwarzschild 2012 argue that for stative sentences, PRES and PAST are scalar alternatives, and claim that a stative PRES-ϕ sentence asymmetrically entails ($\rightarrow$) its PAST-ϕ alternative. To see this, assume a context where a little boy named Scotty has just been brought to the hospital. Dr. Spock is talking to him, when the nurse walks in and ask: 'How is he doing?'. In that context, (27a) entails (27b), but not the reverse.

---

[3]Leahy 2011, 2015 also accounts for the CF inference of SCs through the competition between ICs and SCs. My account differs from his on two points: (i) he does not link the competition between ICs and SCs to the competition between PRES and PAST outside conditionals; (ii) he exclusively deals with '+2 past' SCs, for he considers that '+1 past' SCs are not counterfactual (and argues consequently against Iatridou's 2000 point that '+1 past' SCs about the future present the antecedent as less likely as the corresponding ICs). See the contrast (41)-(42) against this view that ICs and '+1 past' SCs' do not clearly differ in terms of their antecedent falsity inference.

(27) a. Scotty is anxious.   b. → Scotty was anxious.

From this, Altshuler and Schwarzschild 2012 derive the well-known observation that a PAST-ϕ stative sentence (the weaker statement) often implicates ($\rightsquigarrow$) the negation of the stronger PRES-ϕ alternative:

(28) a. Scotty was anxious.   b. $\rightsquigarrow \neg$(Scotty is anxious).

This is what they call the *cessation implicature*, cf. (29).

(29) Cessation implicature: the utterance of a past stative sentence implicates that no state of the kind described currently holds.

5.2   IMP vs. PVF in stative sentences

Languages like French force one to choose between IMP (the *imparfait*) and PFV (the *passé composé*) for past stative sentences too.[4] I argue that in these languages, PRES in stative sentences competes with IMP, but not with PFV. Let us see why.

For cessation to be *implicated* rather than *entailed*, the denoted state has to be potentially non-maximal, i.e. potentially included within some larger state of the same nature, cf. Bary 2009. Importantly, non-maximality can be obtained with IMP, but not with PFV. For perfective operators differ from imperfective ones in that they impose a 'maximal part requirement', which is satisfied if a VP-event culminates or ceases to develop in the actual world. This is the essence of Altshuler's 2014 (slightly modified) hypothesis (30), that builds on insight from Koenig and Muansuwan 2001 and Filip 2008.

(30) Hypothesis about (im)perfective operators (Altshuler 2014)

    a. An operator is *imperfective* if it requires *a part* of an event in the extension of the VP that it combines with, but this part needs not be maximal.

    b. An operator is *perfective* if it requires a *maximal* part of an event in the extension of the VP that it combines with.

Thus, only imperfective stative sentences can *implicate* cessation. Perfective ones *entail* it (because of the maximal part requirement). For instance, (31), with IMP, implicates (32); (33) shows that this implicature can be canceled.

(31) *Scotty était   anxieux.*
      Scotty be-IMP.3SG anxious
      'Scotty was anxious.'

(32) $\rightsquigarrow \neg$(Scotty is anxious).   (implicature)

(33) *Scotty était   anxieux et   l'est toujours.*
      Scotty be-IMP.3SG anxious and it is still
      'Scotty was anxious and still is.'

On the other hand, (34), with PFV, entails that Scotty was not anxious afterwards (If Scotty is (again) anxious in UT, we necessary deal with *two* different fits of anxiety). Hence the difficulty to cancel the inference with *toujours* 'still', cf. (36).[5]

---

[4]Given its restricted use, I do not give examples with the *passé simple*, but the pattern is exactly the same as for the *passé composé*.

[5]Note that *à nouveau* 'again' solves the problem raised by *toujours* 'still' because it satisfies the cessation *entailment* triggered by PFV stative sentences.

(i.)   Scotty a été-PFV.3SG anxieux et l'est à nouveau. 'Pierre was anxious and again is.'

(34) *Scotty   a été   anxieux.*
     Scotty   be-PFV   anxious
     'Scotty was anxious.'

(35) → ¬(Scotty is anxious).   (entailment)

(36) *#Pierre a été   anxieux et   l'est toujours.*
     Scotty   be-PFV.3SG anxious and it is   still
     'Scotty was anxious and still is.'

Note that the cessation inference does not seem to be cancellable via an explicit statement of ignorance concerning the present (differently from what happens in Tlingit, see Cable 2015):

(37) *Il   a   été   malade ce   matin.   #Je ne   sais pas s'il   l'est toujours.*
     He   has been sick   this morning I   NEG   know NEG if he it is   still
     'He was sick this morning. I don't know whether he still is.'

5.3   PAST compete with PRES above MODAL in SCs too

In languages like French, the cancellability of the CF inference of '+1 past SCs' is a direct consequence of the fact that the 'fake' past (above MODAL) is spelled-out with IMP. Given the imperfective semantics of IMP, and the competition between IMP and PRES, a '+1 past' SC *implicates* that the past state during which at least some antecedent-world is accessible does not hold anymore at UT. For instance, (38) implicates (39) for the same reason that (31) implicates (32). Since the CF inference (39) *is* the (cessation) implicature, it is defeasible.

(38) *Si Scotty était   anxieux maintenant, il   nous appellerait.*
     if Scotty be-IMP.3SG anxious now   he   us   call-COND.1
     'If Scotty were anxious now, he would call us.'

(39) ⤳ ¬(the state during which at least some 'Scotty-is-anxious-now'-world is historically
     accessible is holding now)

On this view, in '+1 past' SCs, IMP is therefore 'real' in that it has the same semantic contribution as outside conditionals. In the following subsections, I present two arguments in favour of this analysis.

5.3.1   Asymetrical entailment between ICs and SCs

The first argument is the observation that ICs (with PRES over MODAL) asymmetrically entail the '+1 past' SC alternative (with PAST over MODAL): While (40) entails (38), the reverse does not hold.

(40) *Si Scotty est   anxieux maintenant, il   nous appellera.*
     if Scotty be-PRES.3SG anxious now   he   us   call-FUT
     'If Scotty is anxious now, he will call us.'

5.3.2 No competition with the corresponding IC

My second argument has to do with '+1 past' SCs that do not compete with their stronger alternative (the corresponding IC), because the context makes the antecedent obviously false. Take e.g. the IC (41).

(41) John est mort. #Mais s'il écrit-PRES un roman aujourd'hui, ce sera-FUT un succès.
John is dead. But if he writes a novel today, it will be a success.

(42) John est mort. Mais s'il écrivait-IMP. un roman aujourd'hui, ce serait-COND.1 un succès. (# Ça va peut-être arriver!)
John is dead. But if he wrote a novel today, it would be a success. (# Maybe it will happen!)

In that case, the CF inference of the corresponding '+1 past' SC (42) is not expected to be a cancellable implicature. For the competing stronger statement — the IC (41) — could not be uttered to begin with, precisely because p is taken to be CF. And as the oddity of the continuation parenthesis in (42) shows, the CF inference is indeed noncancellable in (42), although we deal with a '+1 past' SC.

6   Proposal for '+2 past' SCs

For '+2 past' SCs, I keep Ippolito's analysis (6b): the bare conditional is embedded under a past perfect, introducing ⟦PAST⟧ and ⟦PERFECT⟧. The past interval t' output by this tense is perfect (bounded), and its right boundary is in the past. I account for why CF is not cancellable with '+2 past' SCs simply by assuming that antecedent-worlds are accessible *only* at (some subinterval t'' of) the denoted interval t'. Since t' never includes UT with '+2 past' SCs, these SCs are necessarily CF.

Note that under this analysis, although IMP is morphologically present in the past perfect in languages like French, the imperfective semantics does not play a crucial role in the interpretation of the 'high' past perfect, like I argued it is the case in '+1 past' SCs. But in fact, even outside conditionals, it is still a matter of debate whether the French *plus que parfait* is semantically imperfective (Schaden 2007). If the *plus que parfait* is not imperfective outside conditionals, there is no reason to expect it to be so above MODAL in SCs — it might be that IMP appears here because combining a perfect with an imperfective is the only unmarked way to build a double past in languages like French, within and outside conditionals. If, indeed, the *plus que parfait* turns out *not* to be an imperfective, the imperfective semantics is not a necessary ingredient of SCs, even in languages like French.

References

Altshuler, D. (2014). A typology of partitive aspectual operators. *Natural Language and Linguistic Theory*, 32:735–775.

Altshuler, D. and Schwarzschild, R. (2012). Moment of change, cessation implicatures and simultaneous readings. In Chemla, E., Homer, V., and Winterstein, G., editors, *Proceedings of Sinn und Bedeutung 17*, pages 45–62.

Anderson, A. (1951). A note on subjunctive and counterfactual conditionals. *Analysis*, 12:35–38.

Arregui, A. (2005). *On the Accessibility of Possible Worlds: The Role of Tense and Aspect*. PhD thesis, University of Massachusetts at Amherst.

Bary, C. (2009). *Aspect in Ancient Greek: A Semantic Analysis of the Aorist and Imperfective*. PhD thesis, Radboud Universiteit Nijmegen.

Biezma, M., Carnie, A., and Siddiqi, D. (2013). Counterfactuality in non-standard subjunctive conditionals. In Huang, H.-L., Poole, E., and Rysling, A., editors, *Proceedings of NELS 43, vol. 1*, pages 25–36. GSLA, University of Massachussetts, Amherst, MA.

Bjorkman, B. and Halpert, C. (2013). In search of (im)perfection: the illusion of counterfactual aspect. In Keine, S. and Sloggett, S., editors, *Proceedings of NELS 42, vol. 1*, pages 67–78. GSLA Publications.

Cable, S. (2015). The Tlingit Decessive and 'Discontinuus Past': The Curious Implicatures of Optional Past Tense. Manuscript, University of Massachusetts at Amherst.

Chemla, E. (2008). An epistemic step for anti-presupposition. *Journal of Semantics*, 25:141–173.

Dahl, O. (1997). The relation between past time reference and counterfactuality: A new look. In Athanasiadou, A. and Dirven, R., editors, *On Conditionals Again*, pages 97–114. John Benjamins, Amsterdam.

Dancygier, B. and Sweetser, E. (2005). *Mental Spaces in Grammar*. Cambridge University Press, Cambridge.

Ferreira, M. (2014). Displaced aspect in counterfactuals. Towards a more unified theory of imperfectivity. In Crnič, L. and Sauerland, U., editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 2, pages 147–164. MITWPL 71.

Filip, H. (2008). Events and maximalization. In Rothstein, S., editor, *Theorical and Crosslinguistic Approaches to the Semantics of Aspect*, pages 217–256. John Benjamins, Amsterdam.

Halpert, C. and Karawani, H. (2012). Aspect in counterfactuals from A(rabic) to Z(ulu). In Choi, J., Hogue, A., Punske, J., Tat, D., Schertz, J., and Trueman, A., editors, *Proceedings of the 29th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings Project, Somerville.

Heim, I. (1991). Artikel und Definitheit. In von Stechow, A. and Wunderlich, D., editors, *Semantik. Ein internationales Handbuch zur zeitgenössischen Forschung*, pages 487–535. de Gruyter, Berlin.

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31:231–270.

Iatridou, S. (2010). Some thoughts about the Imperfective in Counterfactuals. Ms, MIT.

Ippolito, M. (2003). Presuppositions and implicatures in counterfactuals. *Natural Language Semantics*, 11:145–186.

Ippolito, M. (2013). *Subjunctive Conditionals*. MIT Press, Cambridge.

Karawani, H. (2014). *The Real, the Fake, and the Fake Fake in Counterfactual Conditionals, Crosslinguistically*. PhD thesis, LOT.

Koenig, J.-P. and Muansuwan, N. (2001). How to end without never finishing. *Journal of Semantics*, 17:147–184.

Leahy, B. (2011). Presuppositions and antipresuppositions in conditionals. In Lutz, D., Ashton, N., and Chereches, A., editors, *Proceedings of SALT 21*, pages 257–274.

Leahy, B. (2015). Counterfactual antecedent falsity and embedded antipresuppositions. In Kimmelman, N., Korotkova, N., and Yanovich, I., editors, *Proceedings of Moscow Syntax and Semantics 2*, number 75, pages 69–84. MIT Working Papers in Linguistics.

Romero, M. (2014). Fake tense in counterfactuals: a temporal remoteness approach. In Crnič, L. and Sauerland, U., editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 2, pages 47–63. MITWPL 71.

Schaden, G. (2007). *La Sémantique du parfait. Etude des "temps surcomposés" dans un choix de langues germaniques et romanes*. PhD thesis, Paris 8.

Accommodation and the Strongest Meaning Hypothesis*

Clemens Mayr and Uli Sauerland

Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany
mayr@zas.gwz-berlin.de, uli@alum.mit.edu

Abstract

Quantifiers have been claimed to differ with respect to how they project presuppositions from their scope. Early work claimed that, while negative quantifiers project presuppositions universally, indefinites do so existentially. But the environment the quantifier occurs in matters as well (e.g. Fox [12]). This paper suggests a new account of presupposition projection from a quantifier scope. We propose that, despite appearances, presupposition projection from the scope of quantifiers is always universal following Heim [16]. Apparent deviations from universal projection, we explain by means of additional acontextual silent restriction of quantifiers. Specifically, the mechanism of intermediate accommodation can add the presupposition of a quantifier's scope to its restriction and thereby derive apparent existential projection. We propose furthermore that intermediate accommodation must be licensed by a version of the Strongest Meaning Hypothesis. Our account explains how the monotonicity properties of quantifiers and the environments they occur in determine their projection properties.

1 The puzzle of varying presuppositions in quantifier scopes

That presuppositions embedded in the scope of quantifiers are central to the problem of presupposition projection has been known since at least Karttunen [18] and Heim [16]. The general issue is as follows. For the non-quantificational (1), the presupposition trigger *both* arguably contributes the presupposition that John wrote two papers.

(1)   John is proud of both of his papers.

But what about the quantificational (2)? Assuming that its LF is something like the one in (3), it becomes clear that things are not as straightforward as they were in (1). We cannot simply say that (2) presupposes that x wrote two papers, as x is a variable bound by a λ-operator and thus co-varying with the choice of boys *none* quantifies over.

(2)   None of the ten boys is proud of both of his papers.

(3)   None of the ten boys $\lambda x[x$ is proud of both of x's papers]

The standard reply to this question is that the presupposition of (2) is itself quantificational. But quantificational in which sense? To this one finds different answers in the literature. Heim's [16] theory of presupposition projection derives a universal projection pattern for all presuppositions in the scope of quantifiers.[1] That is, (2) is predicted to presuppose that everyone of the ten boys wrote two papers. Intuitively, this seems adequate. Moreover, Chemla [7] provides experimental

---

[1]This result obtains independently of the dynamic system and the related notion of local context employed there. Schlenker's [24] static reconstruction of a system of presupposition projection using local contexts derives exactly the same result. Also see [23].

evidence for this.[2] His findings are also supported by Fox's [12] observation that some speakers judge statements with negative quantifiers as incoherent if preceded by a sentence requiring that only some of the ten boys wrote two papers, as in (4) modeled on his example (4). The reason would be that the universal presupposition of the second sentence clashes with the assertion of the first one.

(4)   #Half of the ten boys wrote two papers. But none of the ten boys is proud of both of his papers.

Indefinites as in (5), however, seemingly do not lead to an inference that all of the ten boys wrote two papers. Under a quantificational analysis, however, Heim [16] would derive exactly that as its presupposition.[3]

(5)   One of the ten boys is proud of both of his papers.

In addition, to different quantifiers projecting presuppositions differently from their scopes, Fox [12] has produced evidence suggesting that the environment in which the quantifier occurs in also affects the possible presuppositional inferences. As a solution to this issue of varying presupposition projection patterns, he has pursued an account couched in a Strong Kleene semantics (see also [21, 2, 14, 15, 11] a.o.). Such a system shows more flexibility in the presuppositions that it assigns to quantificational sentences than competing theories do. In a nutshell, there the presupposition of a sentence corresponds to the statement of what it takes for the sentence to be either true or false – i.e., for it to have a truth-value that is not the third one. Thus, it derives a disjunctive presupposition for all of the examples above, where one disjunct corresponds to what it takes for the sentence in question to be true and the other disjunct corresponds to what it takes for it to be false. Therefore, since (2) is just the negation of (5), they are predicted to have the same disjunctive presupposition: *'Either all of the ten boys wrote two papers or some boy wrote two papers and is proud of both of them'.*

This presupposition is indeed weaker than the one predicted by Heim [16] and thus at first glance looks more promising in light of the contrast noted. Notice, however, that the second disjunct of this presupposition is equivalent to the assertive component of (5). Fox notes that this fact causes a potential issue. Assuming with Stalnaker [28] that presuppositions of sentences are requirements which must be met by the context, and that moreover the assertion of a sentence should not be entailed by the context in order for it to be informative, it follows that no context can entail the second disjunct of the predicted presupposition for (5) if it is to be asserted. Rather the first disjunct should be entailed because only this one allows the assertive component to be informative. But then one again derives an existential inference for positive indefinites, contrary to our conclusion from above.[4] For (2) no such issue arises. Given its assertion, the context should entail the first disjunct of its predicted presupposition, namely that every boy wrote two papers. A context entailing the second disjunct would be incompatible with the assertion of (2).

Summarizing, a more flexible system of presupposition projection making use of Strong Kleene semantics such as Fox's [12] one is not fully satisfying to account for the contrast discussed above. Less flexible systems, however, are also not completely adequate. Heim's [16] universal projection pattern, on the other hand, is needed for at least negative quantifiers. On the other hand, something like Beaver's [3] weaker existential projection pattern is necessary for

---

[2]At this point, we haven't been able to consider the data presented by Zehr et al. [31].

[3]Heim [16] being aware that this prediction might be problematic therefore assumes that the indefinite in (5) is not quantificational at all thereby avoiding the unwelcome result.

[4]Fox [12] shows that there are indeed situations where this universal inference might be adequate. But the issue that this inference appears too strong in many cases, and in particular in (5), remains as far as we can see.

positive indefinites it seems. It is not clear how one general system for presupposition projection can derive both kinds of projection patterns in a principled way, though.

The present paper proposes the following solution to the puzzle. Despite appearances, presupposition projection from the scope of quantifiers is uniform. In particular, following Heim [16] and Schlenker [24] it is always universal. We argue that apparent deviations from this universal projection pattern are due to differences in accommodation. Specifically, weaker existential inferences arise if the quantifier is acontextually restricted by accommodating the presupposition from its scope – so called intermediate accommodation (see [29] but cf. [3] for an opposing view). The choice between such a restriction and a vacuous restriction leading to a universal inference is negotiated by a weak version of the Strongest Meaning Hypothesis. We show that such a system does not only account for the data reviewed above but makes further welcome predictions: first, it correctly accounts for the presuppositional inferences licensed by other quantificational determiners. Second, it accounts for why projection patterns seemingly switch in entailment-reversing contexts.

2 The proposal

2.1 Accommodation and the Strongest Meaning Hypothesis

Given a sentence S with quantifier Q in it, we assume with Heim [16] that Q projects presuppositions universally from its scope. Contrary to Heim we include indefinites in the category of Q as existential quantifiers. Now, such a sentence S with a presupposition trigger in the scope of Q is in principle multiply ambiguous – that is, S is ambiguous with regards to the restriction of Q. And more specifically, S is ambiguous between interpretations S_U and S_D, where U is a vacuous restrictor on Q – with U standing for 'Universe' – and D is the non-empty restrictor on Q arising from accommodation of the presupposition in the scope of Q. In other words, S is ambiguous between (6a) and (6b), where R is the overt restrictor of Q and R'_P its scope with presupposition P. U and D are given in boldface in the following to differentiate them from R.

(6)   a. $Qx[x \in R \cap \mathbf{U}][x \in R'_P]$ (S_U)
         b. $Qx[x \in R \cap \mathbf{D}][x \in R'_P]$ (S_D)

Given our assumption that presuppositions are generally projected universally from the scope of Q, the following holds: S_U, on the one hand, results in the strong universal presupposition (SUP) in (7a). S_D, on the other hand, results in the weak universal presupposition (WUP) in (7b). That is, SUP is an inference about all the individuals in R restricted by U – which amounts to all the individuals in R given that U is a vacuous restrictor – whereas WUP is a presupposition about all those individuals in R which are also in D. In case D does not include all the individuals that are in R, S_D with its WUP leads to an inference that is weaker than the one arising from S_U with its SUP. Crucially, though the projection mechanism delivers the same result in both cases. Thus, strictly speaking it is not correct to refer to the WUP as an existential presupposition, at least not in the sense that, for instance, Beaver [3] uses the term.

(7)   a. $\forall x[x \in R \cap \mathbf{U} \to x \in P] \equiv \forall x[x \in R \to x \in P]$ (SUP for S_U)
         b. $\forall x[x \in R \cap \mathbf{D} \to x \in P]$ (WUP for S_D)

Now, Dalrymple et al. [10] argued on the basis of ambiguity in reciprocal statements that out of the theoretically possible readings only the strongest reading compatible with the context and world-knowledge is available. They call this principle the Strongest Meaning Hypothesis. We make use of the weak version of the Strongest Meaning Hypothesis (wSMH) in (8). In

particular, according to wSMH reading S_D and with it WUP is only available if S_D itself is not tautologous and S_D is at least as strong as S_U.

(8)   *Weak version of the Strongest Meaning Hypothesis*
      Given sentence S that is ambiguous between readings S_U and S_D, choose S_D only if $S_D \neq \bot$, and S_D either asymmetrically Strawson-entails S_U or S_D is Strawson-equivalent to S_U. Otherwise choose S_U.

Notice that the wSMH must use von Fintel's [30] notion of Strawson-entailment for comparing the strength of the two readings: given two propositions p and q, p Strawson-entails q if and only if whenever p and the presupposition of q are true, q is true itself.[5] Given that S_U and S_D come with presuppositions, Strawson-entailment might hold even though classical entailment might not. Generally, Strawson-entailment is a weaker notion than classical entailment. Hence the name wSMH.

2.2   Application to universal and negative quantifiers and indefinites

Consider the contrast in (9) modeled on Fox's [12] example (3). While (9a) is acceptable as a continuation of the first sentence, (9b) is not.

(9)   Half of the ten boys wrote two papers. . . .

      a.   Furthermore, one of the ten boys is proud of both of his papers.
      b.   #Furthermore, every one of the ten boys is proud of both of his papers.

The present proposal derives this contrast. Let us assume for simplicity that assertion and presupposition are conjoined at the scope level with the presuppositional component being underlined in the following representations. Then S_U and S_D are as in (10) for the sentences in (9), differing only in the quantifiers. B stands for the set containing the ten boys. Again U is the vacuous restrictor. Therefore $B \cap U$ is just the set containing the ten boys. D, however, is non-vacuous. Since D is moreover required to accommodate the presupposition in the scope of the quantifier, D corresponds to the set of individuals who wrote two papers. Thus $B \cap D$ is the set containing those individuals in B who wrote two papers. As a consequence, $B \cap D \subseteq B \cap U$.

(10)   a. $Qx[x \in B \cap \mathbf{U}]$ [$\underline{x \text{ wrote two papers}}$ and is proud of both of his papers] (S_U)
       b. $Qx[x \in B \cap \mathbf{D}]$ [$\underline{x \text{ wrote two papers}}$ and is proud of both of his papers] (S_D)

Now, which of the readings in (10) does the wSMH choose for the sentences in (9)? Note that while *every* is Strawson-antitone on its restrictor, *one* is Strawson-isotone on its restrictor. In other words, the restrictor of *every* is an entailment-reversing environment, but the one of *one* is not.[6] Thus *every* allows inferences from sets to subsets in its restrictor. Therefore S_U asymmetrically Strawson-entails S_D for (9b), which can be shown by the paraphrases for S_U and S_D in (11). As a consequence of this, wSMH selects S_U, and a SUP inference is predicted for (9b): each of the ten boys wrote two papers. This clashes with the requirement of the preceding sentence. The result is deviance.

---

[5]More formally, $p \in D_{\langle s,t \rangle}$ Strawson-entails $q \in D_{\langle s,t \rangle}$ iff in all worlds w such that $p(w) = 1$ and $q(w) \neq \#$, $q(w) = 1$.
[6]We use the lattice theoretic terms *isotone* and *antitone* of Birkhoff [4] rather than the terms $\pm$*affective* or *up-/downward entailing*, but with the same meaning. I.e. Given $f \in D_{\langle \sigma, \tau \rangle}$ and $x, y \in D_\sigma$ such that x Strawson-entails y, f is Strawson-isotone iff $f(x)$ Strawson-entails $f(y)$, and f is Strawson-antitone iff $f(y)$ Strawson-entails $f(x)$.

(11)   Everyone of $B \cap \mathbf{U}$ <u>wrote 2 papers</u> and is proud of both of his papers. (S_U)
        $\Rightarrow_S$
        Everyone of $B \cap \mathbf{D}$ <u>wrote 2 papers</u> and is proud of both of his papers. (S_D)

One, on the other hand, only allows inferences from sets to their supersets in its restrictor. Hence, S_D Strawson-entails S_U for (9a), and in fact, the readings are Strawson-equivalent. Again, this can be clearly seen by the paraphrases in (12). Consequently, S_D is selected by wSMH. That is, we predict the WUP inference that one of the ten boys wrote two papers for (9a), which makes it a coherent continuation of the preceding sentence.

(12)   One of $B \cap \mathbf{D}$ <u>wrote 2 papers</u> and is proud of both of his papers. (S_D)
        $\Leftrightarrow_S$
        One of $B \cap \mathbf{U}$ <u>wrote 2 papers</u> and is proud of both of his papers. (S_U)

Finally, the degraded status of (13), repeated from (4) above, is also expected. Negative quantifiers like other negative expressions are Strawson-antitone, and in particular, they are Strawson-antitone on their restrictor. (14) gives the paraphrases for the S_U and S_D readings. Crucially, these correspond to universal statements with negation at the scope level – where negation, of course, does not affect the presuppositional component. As can be seen there is asymmetric Strawson-entailment from S_U to S_D. This means that we predict that S_U is selected by wSMH for the second sentence in (13) in parallel to the case with the universal quantifier. From this the SUP inference that each of the ten boys wrote two papers follows, which accounts for the deviance of (13).

(13)   #Half of the ten boys wrote two papers. But none of the ten boys is proud of both of his papers.

(14)   Each of $B \cap \mathbf{U}$ <u>wrote 2 papers</u> and isn't proud of both of his papers. (S_U)
        $\Rightarrow_S$
        Each of $B \cap \mathbf{D}$ <u>wrote two papers</u> and isn't proud of both of his papers. (S_D)

3   Predictions and extension to other quantifiers

3.1   Embedded Projectors

By default the wSMH applies at the sentence level (e.g. [26]). A prediction of this is that the projection properties of quantifiers can change when embedded in an antitone environment such as the antecedent of a conditional or under the predicate *doubt*. It appears that this prediction – with related data already discussed by Fox [12] from a different theoretical perspective – is correct, as the following examples illustrate. We have shown that (15a) is acceptable in the context of the preceding sentence due to its WUP inference. We now, add to this that (15b) and (15c) are somewhat degraded relative to (15a) when uttered in that same context. We submit that this is due to the antitone property of the antecedent of the conditional and *doubt* leading to the selection of S_U by wSMH and thereby a conflicting SUP inference. (16) shows the reverse behavior for negative quantifiers, which on their own trigger a problematic SUP inference but when themselves embedded in an antitone environment can trigger an unproblematic WUP inference. In the following, we indicate whether S_U or S_D is chosen by wSMH for a given example by notating the respective restrictor as a subscript on the quantifier.

(15)   Five of the ten boys wrote two papers. . . .

        a.   One$_D$ of the ten boys is proud of both of his papers.

b. #If one$_U$ of the ten boys is proud of both of his papers, he will submit them.
c. #I doubt that any$_U$ of the ten boys is proud of both of his papers.

(16)   Five of the ten boys wrote two papers. ...

a. #None$_U$ of the ten boys is proud of both of his papers.
b.   If none$_D$ of the ten boys is proud of both of his papers, no one will submit one.
c.   I doubt that none$_U$ of the ten boys is proud of both of his papers.

We add to this that while entailment reversal can trigger a reversal in the reading selected by wSMH and while this indeed might be the preferred option, this need not happen. In particular, Sauerland [22] argues that the Strongest Meaning Hypothesis can apply locally. So, if the wSMH applies at the level of the antecedent of the conditional in the (b)- and (c)-examples above, the antitone property is immaterial and the same prediction as in the (a)-examples is made.

3.2   Modified Numerals

Chemla [7] shows experimentally that modified numerals like *more than n NP*, *exactly n NP*, and *fewer than n NP* all behave more similarly to indefinites than they do to universal or negative quantifiers with regards to the presuppositional inferences they allow for. Let us therefore consider the predictions of the present proposal for modified numerals.

More than n NP is Strawson-isotone on its restrictor. Therefore generally S_D rather than S_U is the preferred option by wSMH. But which D does it choose? In the case of (17), for instance, no D with a cardinality lower than four can be selected by wSMH. The reason for this is that the wSHM does not take into account contradictions, and restricting the set of ten boys to a set with fewer than four boys in it contradicts the semantic contribution of *more than three*. Thus a D with a cardinality of at least four is chosen by the wSMH. As the paraphrase of the resulting S_D of (17) in (18) shows, this indeed Strawson-entails the S_U reading and is in fact Strawson-equivalent to it. The result is a WUP inference for (17) saying everyone in a group of at least four boys out of the ten boys wrote two papers.

(17)   More than three$_D$ of the ten boys are proud of both of their papers.

(18)   More than 3 of $B \cap \mathbf{D}$ wrote 2 papers and are proud of both of their papers.   (S_D)
       $\Leftrightarrow_S$
       More than 3 of $B \cap \mathbf{U}$ wrote 2 papers and are proud of both of their papers.   (S_U)

The remarks about *more than three NP* in (17) carry over to the non-monotonic *exactly three NP* in (19): its S_D is Strawson-equivalent to its S_U as their paraphrases in (20) make apparent. Once again wSMH selects S_D and we get a WUP inference for (19). This time it says that everyone out of a group of exactly three boys out of the ten boys wrote two papers.

(19)   Exactly three$_D$ of the ten boys are proud of both of their papers.

(20)   Exactly 3 of $B \cap \mathbf{D}$ wrote 2 papers and are proud of both of their papers.   (S_D)
       $\Leftrightarrow_S$
       Exactly 3 of $B \cap \mathbf{U}$ wrote 2 papers and are proud of both of their papers.   (S_U)

While our predictions for *more than three NP* and *exactly three NP* align with Chemla's [7] experimental findings, this is not straightforwardly the case for *fewer than three NP*. The latter is Strawson-antitone on its restrictor. As the paraphrases of S_U and S_D in (22) for the sentence in (21) show, a situation parallel to the one with negative indefinites obtains. wSMH is predicted to select S_U. That is, a SUP inference that each of the ten boys wrote two papers is derived.

(21)   Fewer than three$_D$ of the ten boys are proud of both of their papers.

(22)   For each group of 3 of $B \cap \mathbf{U}$ each <u>wrote 2 papers</u> and isn't proud of both. (S_U)
$\Rightarrow_S$
For each group of 3 of $B \cap \mathbf{D}$ each <u>wrote 2 papers</u> and isn't proud of both. (S_D)

However, Chemla [7] shows an SUP inference to be inadequate for sentences like (21). Rather a weaker inference is observed. Now, Krifka [20] and Gajewski [13] independently argued that modified numerals like *fewer than n NP* can have an existential scalar implicature. That is, (21) is claimed to standardly lead to a strengthened interpretation as in (23). But since (23) is a non-monotonic environment, this affects which reading the wSMH selects. Parallel to *exactly three NP* above, S_U and S_D become Strawson-equivalent. Therefore wSMH selects S_D, and we are after all able to correctly predict a WUP inference for (21) according to which everyone of some group with one or two boys wrote two papers.

(23)   'Fewer than three of the ten boys are proud of both of their papers, but some of the ten boys are proud of both of their papers.'

3.3   *Only*

Note that (19) above contrasts sharply with the minimally differing (24) with *only three* instead of *exactly three*. (19), on the one hand, licenses the WUP inference saying that each boy out of a group of exactly three boys out of the ten boys wrote two papers. (24), on the other hand, seemingly licenses the stronger inference that all of the ten boys wrote two papers. This is somewhat surprising given that (24) seemingly entails that exactly three of the ten boys are proud of both of their papers.

(24)   Only three$_U$ of the 10 boys are proud of both of their papers.

Note, however, that under Horn's [17] semantics for *only* its literal meaning consists of both a presuppositional and an assertive component. This view has become standard, and we therefore adopt it. Furthermore for concreteness and simplicity let us assume with Büring and Hartmann [5] that *only* in (24) is a sentential operator as in (25). Nothing hinges on this, though.

(25)   only [three of the ten boys are proud of both of their papers]

For present purposes it then suffices to assume that *only* in (24) contributes the presupposition that the prejacent is true and asserts that every alternative with a cardinal higher than *three* is false.[7] This derives for S_U and S_D the presuppositional and assertive components in (26). Now, recall that the wSMH makes use of Strawson-entailment rather than classical entailment. Thus when checking whether S_U Strawson-entails S_D, the presupposition of the latter is assumed to be true.[8] As a consequence, strictly speaking only the assertive components of S_U and S_D matter for the wSMH. And since *only* is Strawson-antitone in its assertive component as the asymmetric entailment relation in (26) shows, wSMH selects S_U. This means the present account predicts a SUP inference for (24) saying that all of the ten boys wrote two papers.

(26)   3 of $B \cap \mathbf{U}$ <u>wrote 2 papers</u> and are proud of both.   (prsp. S_U)
For each group of 4 of $B \cap \mathbf{U}$ each <u>wrote 2 papers</u> and isn't proud of both. (asrt. S_U)
$\Rightarrow_S$
3 of $B \cap \mathbf{D}$ <u>wrote 2 papers</u> and are proud of both.   (prsp. S_D)
For each group of 4 of $B \cap \mathbf{D}$ each <u>wrote 2 papers</u> and isn't proud of both. (asrt. S_D)

---

[7]All of this is rather uncontroversial. For discussion on how the compositional semantics of *only* is to derive these inferences, we refer the reader to Coppock and Beaver [9] and references therein.
[8]In fact, the presuppositional components of S_U and S_D are Strawson-equivalent.

The fact that the present system accounts for the contrast in presuppositional inferences between *exactly n NP* and *only n NP* is strong evidence in its favor.[9]

3.4 Proportional quantifiers

Consider next the proportional quantifier *most of the NP*. In contrast to non-monotonic *exactly three NP*, it appears to lead to a presuppositional inference for (27) that all of the ten boys wrote two papers.

(27) Most_U of the ten boys are proud of both of their papers.

Even though *most of the NP* is Strawson non-monotone, the wSMH does not allow for restriction by D. This is due to its proportionality. As (28) shows, the S_U reading asymmetrically entails the S_D reading. We thereby predict an SUP inference for (28) according to which all of the ten boys smoke.

(28)   Most of $B \cap \mathbf{U}$ <u>wrote 2 papers</u> and are proud of both of their papers. (S_U)
        $\Rightarrow_S$
        Most of $B \cap \mathbf{D}$ <u>wrote 2 papers</u> and are proud of both of their papers. (S_D)

Notice that if D did not accommodate the presupposition in the scope of *most of the NP* but denoted a simple subset of U, there would not be any entailment between S_U and S_D. As a consequence, an S_D reading would be possible for (28). But this could lead to an unwanted inference that, for instance, most of a group of four boys out of the ten are proud of both of their papers. By accommodation we avoid this problematic prediction.[10]

3.5 Anaphoric Triggers

Charlow [6] argues so-called strong presupposition trigger like *also* trigger a universal presuppositional inference even from the scope of indefinites. He cites the infelicity of the sequence in (29) as evidence for this claim. That is, the second sentence in (29) is said to lead to the inference that all of the hundred boys smoked something other than Marlboro in the past, which is incompatible with the preceding sentences.

(29)   Just half of these 100 boys smoked in the past. They have all smoked Nelson.
        #Unfortunately, some_U of these 100 boys have also smoked Marlboro.

Now it is well-known that strong triggers to not easily allow for accommodation. Chemla and Schlenker [8] give the example in (30), to substantiate this point (also see [25, 1, 19]): if the presupposition of the strong trigger *too* – that the speaker gave a black eye to someone other than Susie – could be locally accommodated under *but*, Bill's reply should be felicitous. The fact that it is not suggests that *too* does not allow local accommodation.

(30)   Teacher: Johnny claims that you gave him a black eye. Is this true?
        Bill: I don't know, but #if I give Susie a black eye too, they'll be twins.

Now, recall that in the present system restriction by D always involves accommodation of the presupposition in the scope of the quantifier. But if accommodation of the presupposition of *also* in (29) is not allowed, for the reasons just discussed, then the S_U reading with the universal SUP that all of the hundred boys have smoked something other than Marlboro in the past is predicted. Thus we predict the infelicity of (29).

---

[9]Spector and Sudo [27] presumably also derive the correct result for *only n NP*. It is currently not clear to us, however, what their predictions for *exactly n NP* are.

[10]We thank Yasutada Sudo (p.c.) for raising this issue.

4 Conclusion

We showed that the wSMH viewed as a restriction on when presuppositions can be accommodated in the restrictors of quantifiers predicts the core data of presupposition projection from the scopes of quantifiers. While this solution is formally similar to the one pursued by Fox [12] in that they both make presupposition projection dependent on both the quantifier and the overall linguistic context, our account seems preferable for two reasons: first, it has fewer difficulties with indefinites than Fox's, and second it has wider empirical coverage.

References

[1] Dorit Abusch. Lexical alternatives as a source of pragmatic presupposition. In Brendan Jackson, editor, *Proceedings of SALT XII*, pages 1–19, Cornell University, Ithaca, NY, 2002. CLC Publications.

[2] David Beaver and Emiel Krahmer. A partial account of presupposition projection. *Journal of Logic, Language and Information*, 10:147–182, 2001.

[3] David I. Beaver. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications, Stanford University, 2001.

[4] Garrett Birkhoff. *Lattice Theory*. American Mathematical Soc., 1940.

[5] Daniel Büring and Katharina Hartmann. The syntax and semantics of focus-sensitive particles in German. *Natural Language and Linguistic Theory*, 19:229–281, 2001.

[6] Simon Charlow. "Strong" predicative presuppositional objects. In *Proceedings of ESSLLI*, 2009.

[7] Emmanuel Chemla. Presuppositions of quantified sentences: experimental data. *Natural Language Semantics*, 17(4):299–340, 2009.

[8] Emmanuel Chemla and Philippe Schlenker. Incremental vs symmetric accounts of presupposition projection: an experimental approach. *Natural Language Semantics*, 20(2):177–226, 2012.

[9] Elizabeth Coppock and David Beaver. Principles of the exclusive muddle. *Journal of Semantics*, 31:371–432, 2014.

[10] Mary Dalrymple, Makoto Kanazawa, Yookyung Kim, Sam Mchombo, and Stanley Peters. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy*, 21:159–210, 1998.

[11] Danny Fox. Two short notes on Schlenker's theory of presupposition projection. *Theoretical Linguistics*, 34:237–252, 2008.

[12] Danny Fox. Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening. In Ivano Caponigro and Carlo Cecchetto, editors, *From Grammar to Meaning: The Spontaneous Logicality of Language*, pages 201–232. Cambridge University Press, Cambridge, 2013.

[13] Jon R. Gajewski. Licensing strong NPIs. *Natural Language Semantics*, 19(2):109–148, 2011.

[14] Benjamin George. Predicting presupposition projection: some alternatives in the Strong Kleene tradition. ms. UCLA, 2008.

[15] Benjamin George. A new case for an old logic: reviving Strong Kleene approaches to presupposition projection. ms. UCLA, 2010.

[16] Irene Heim. On the projection problem for presuppositions. In Daniel P. Flickinger, editor, *Proceedings of WCCFL 2*, pages 114–125, Stanford University, Stanford, California, 1983. CSLI Publications.

[17] Laurence R. Horn. A presuppositional analysis of *only* and *even*. In *Papers from the 5th annual meeting of the Chicago Linguistics Society*, pages 98–108, Chicago, 1969. Chicago Linguistics Society.

[18] Lauri Karttunen. Some observations on factivity. *Papers in Linguistics*, 4(1):55–69, 1971.

[19] Nathan Klinedinst. Totally hardcore semantic presuppositions. In A. Capone, F. L. Piparo, and M. Carapezza, editors, *Perspectives on Pragmatics and Philosophy*. Springer, 2012.

[20] Manfred Krifka. At least some determiners aren't determiners. In K. Turner, editor, *The semantics/pragmatics interface from different points of view*, pages 257–291. Elsevier Science B.V., Amsterdam, 1999.

[21] Stanley Peters. A truth-conditional formulation of Karttunen's account of presupposition. *Synthese*, 40:301–316, 1979.

[22] Uli Sauerland. Where does the Strongest Meaning Hypothesis apply? *Snippets*, 25:13–14, 2012.

[23] Philippe Schlenker. Be articulate! A pragmatic theory of presupposition projection. *Theoretical Linguistics*, 34:157–212, 2008.

[24] Philippe Schlenker. Local contexts. *Semantics and Pragmatics*, 2:1–78, 2009.

[25] Mandy Simons. On the conversational basis of some presuppositions. In Rachel Hastings, Brendan Jackson, and Zsofia Zvolenszky, editors, *Proceedings of SALT XI*, pages 431–448, Cornell University, Ithaca, NY, 2001. CLC Publications.

[26] Benjamin Spector. Global positive polarity items and obligatory exhaustification. ms. Institut Jean Nicod, 2013.

[27] Benjamin Spector and Yasutada Sudo. Presupposed ignorance and exhaustification: How scalar implicatures and presuppositions interact. ms. Institut Jean Nicod, CNRS-ENS-EHESS University College London, 2014.

[28] Robert C. Stalnaker. Assertion. *Syntax and Semantics*, 9, 1978.

[29] Rob van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377, 1992.

[30] Kai von Fintel. NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics*, 16:97–148, 1999.

[31] Jérémy Zehr, Cory Bill, Lyn Tieu, Jacopo Romoli, and Florian Schwarz. Experimental evidence for existential presupposition projection from *none*. In *Amsterdam Colloquium 2015*, 2015.

Particle responses to negated assertions: Preference patterns for German *ja* and *nein* *

A. Marlijn Meijer, Berry Claus, Sophie Repp and Manfred Krifka

{`meijeram`|`berry.claus`|`sophie.repp`|`krifka`}`@hu-berlin.de`

Abstract

The present study focusses on German polarity particles as responses to negative assertions. In such responses, the German response particles *ja* and *nein* are not used complementarily − similarly to their roughly corresponding English counterparts *yes* and *no*. Rather it seems that both *ja* and *nein* can be used to affirm a negative antecedent (e.g. A: *Jim doesn't snore.* B: *Ja./Nein.* (=He doesn't snore)). In a series of acceptability judgement experiments, we tested the predictions of two recent theoretical accounts (Krifka 2013 and Roelofsen and Farkas 2015). For affirming responses to negative assertions, both accounts predict a default preference for *nein* over *ja*. However, our experimental results revealed two subgroups of participants. A majority of approx. 70% showed a preference for *ja* over *nein*. The other subgroup showed a preference for *nein* over *ja*. To explain this finding, we consider modified versions of both accounts.

1 Introduction

The use and interpretation of response particles such as English *yes* and *no* is clear-cut only in responses to non-negative antecedents. Matters are different with negative antecedents. It has been claimed that both *yes* and *no* can be used in affirming responses to negative assertions, see (1). Kramer and Rawlins have called this phenomenon 'negative neutralization' (Kramer and Rawlins 2011), as the meaning of *yes* and *no* seems to be collapsed.

(1)   A   John doesn't snore.
      B.i   No, he doesn't.   B.ii   Yes, he doesn't.

However, experimental evidence suggests that there is a preference for *no* over *yes* in responses to negative assertions. Brasoveanu, Farkas and Roelofsen (2013) have shown that *no* is preferred over *yes* in affirming responses to negative assertions. Kramer and Rawlins (2012) have shown the same preferences for affirming responses to negative questions.

The German particles *ja* and *nein* also are complementary in responses to non-negative antecedents and they display an unclarity in responses to negative antecedents (see (2)). It is assumed that both *ja* and *nein* can be used in affirming responses to antecedents with sentential negation (Blühdorn 2012: 386). Unlike English, however, German has a third, specialized response particle, in addition to *ja* and *nein*: *doch*. This particle is used in responses to negative antecedents and indicates that a previous assertion is not true (see (2)).[1]

[1]In some situations, *doch* can also be used in response to a non-negative assertion or question. For details, see Karagjosova (2006).

(2) A   Jim schnarcht nicht.
         'Jim doesn't snore.'
     B.i Ja./Nein.   (= He doesn't snore.)
     B.ii Doch.   (= He does snore.)

Up to now, there have not been any experimental studies investigating the German response particle system. There are several recent theoretical accounts of response particle systems, e.g. Holmberg (2013), Kramer and Rawlins (2012), Krifka (2013) and Roelofsen and Farkas (2015), two of which make predictions for preference patterns of the German response particles with competing analyses: Krifka's (2013) anaphor account and Roelofsen & Farkas' (2015) syntactic-semantic feature model. The present paper presents a series of experiments juxtaposing these two accounts.

2   Theoretical Background

2.1   Anaphor account (Krifka 2013)

Krifka argues that polarity particles are anaphors that refer back to salient propositions in the context. He proposes that *ja* targets a proposition and asserts it and *nein* targets a proposition and asserts its negation. Furthermore, Krifka argues that antecedents containing sentential negation introduce two propositional discourse referents (propDR): the negated proposition ($\bar{p}_{\text{DR}}$) and its non-negated counterpart (p_{DR}), see (3).

(3) $\llbracket$ [Jim $[_{\bar{p}_{\text{DR}}}$ t$_{\text{Jim}}$ doesn't $[_{p_{\text{DR}}}$ t$_{\text{Jim}}$ snore]]]$\rrbracket$ = $\neg$ snore(jim)
     $\bar{p}_{\text{DR}}$ (negated propDR): $\neg$ snore(jim)
     p_{DR} (positive propDR): snore(jim)

Support for this assumption comes from other propositional anaphors, e.g. *that*. In response to a negative assertion like the one in (4), B can use *that* to refer to the negated (B.i) or the non-negated part (B.ii) of A's utterance.

(4) A Two plus two isn't five.   B.i Everyone knows that$_{\bar{p}_{\text{DR}}}$.
     [$\neg$[two plus two is five]$_{p_{\text{DR}}}$]$_{\bar{p}_{\text{DR}}}$ B.ii That$_{p_{\text{DR}}}$ would be a contradiction. (Krifka 2013)

Crucially, according to Krifka, the propDRs differ in saliency. In a neutral context (the default), the non-negated proposition is assumed to be salient because negated propositions are usually uttered in contexts where the non-negated proposition is under discussion and thus is salient already. Krifka suggests that speakers prefer reference to the salient propDR since salient discourse referents in general are targeted more readily by anaphors than non-salient discourse referents (e.g. Gundel et al. 1993). Thus, he predicts that for affirming responses to negative antecedents, *nein* should be preferred over *ja* as a default: *nein* targets the salient p_{DR} and asserts its negation; *ja* targets the non-salient $\bar{p}_{\text{DR}}$ and asserts it. However, the preference pattern should differ from the default in contexts in which the $\bar{p}_{\text{DR}}$ (rather than p_{DR}) is salient, e.g. when the antecedent is preceded by a negative question as shown in Krifka's example for English in (5).

(5) A   Which of the mountains on this list did Reinhold Messner not climb?
     B   Well.. he did not climb Cotopaxí in Ecuador.
     A   Yes./No.   (Krifka 2013:14)

For (5) Krifka assumes that the negative question preceding the antecedent renders the $\bar{p}_{\text{DR}}$ salient. This should result in a preference for *ja* over *nein*, because *ja* targets the salient $\bar{p}_{\text{DR}}$ whereas *nein* targets the non-salient p_{DR}.

For *doch*, i.e. the specialized particle for rejecting a negative antecedent, Krifka assumes that it comes with the presupposition that there are two salient propDRs and one is the negation of the other; *doch* targets and asserts p_{DR}. When its presupposition is satisfied, *doch* blocks the use of *yes* which could also reject a negative antecedent by targeting and asserting p_{DR}.

To sum up, Krifka predicts that for affirming negated propositions in default contexts *nein* is preferred over *ja*, due to the salience of p_{DR} that *nein* targets and negates. This pattern is predicted to be reversed in negative contexts. For rejecting responses to negative antecedents, Krifka predicts that *doch* blocks *ja*, whereas *nein* is not blocked by *doch*, although it is dispreferred.

2.2 Feature account (Roelofsen and Farkas 2015)

Roelofsen and Farkas (2015, henceforth R&F) propose a semantic-syntactic feature account, in which polarity particles encode absolute and relative polarity features. The particles head a polarity phrase and take a (possibly elided) response clause as an argument (e.g. *he does(n't)* in (7)). The absolute features, [+] and [−], pertain to the polarity of the response clause. [+] presupposes that the response clause polarity is positive, [−] presupposes that it is negative. The relative features, [AGREE] and [REVERSE], encode the relation between the response clause and its antecedent. [AGREE] presupposes that the polarity of the response clause agrees with the polarity of the antecedent whereas [REVERSE] presupposes that the response clause reverses the polarity of the antecedent.

For the English particles *yes* and *no*, R&F propose the following realization potentials:

(6)   *Realization potential of English particles*
     a.   [AGREE] and [+] are realized by *yes*
     b.   [REVERSE] and [−] are realized by *no*   (R&F 2015)

Thus, [AGREE,+] must be realized by *yes* and [REVERSE,−] by *no*. These are the two complementary uses of *yes* and *no* in response to non-negative assertions. However, both particles are possible candidates in responding to a negative assertion, or in terms of R&F, for the realization of [AGREE,−] or [REVERSE,+]. This is illustrated in (7).

(7)   A   John doesn't snore.
        B.i   Yes/No, he doesn't.   [AGREE,−]   B.ii Yes/No, he does.   [REVERSE,+]

German is like English insofar as *ja* can realize [AGREE] and [+], whereas *nein* can realize [REVERSE] and [−], see (8). The specialized particle *doch* realizes the feature combination [REVERSE,+].

(8)   *Realization potential of German particles*
     a.   [AGREE] and [+] are realized by *ja*
     b.   [REVERSE] and [−] are realized by *nein*
     c.   [REVERSE,+] is realized by *doch*

The polarity features differ regarding their markedness. R&F assume that the more marked a feature is, the higher its realization need. Of the absolute features, [−] is more marked than [+]. Of the relative features, [REVERSE] is more marked than [AGREE]. [AGREE] and

[+] form a natural class, as do [REVERSE] and [−]; these classes correspond to the use of *ja* and *nein* in non-negative contexts. Furthermore, R&F note that [+] is contrastive when it co-occurs with [REVERSE]; this makes [REVERSE,+], which is realized by *doch*, a highly marked feature combination. Due to the higher realization need of marked particles, R&F predict that in affirming responses to negative assertions ([AGREE,−]), German speakers prefer *nein* over *ja* because *nein* realizes the marked feature [−] and *ja* the unmarked feature [AGREE]. For rejecting responses to negative assertions ([REVERSE,+]), R&F predict that only *doch* can be used. The availability of this specialized particle blocks both *ja* and *nein*.

Summarizing, R&F predict that in affirming negative antecedents *nein* is preferred over *ja*, due to the markedness of [−]. In rejecting negative antecedents R&F expect *doch* to be preferred over both *ja* and *nein*, due to the markedness of the feature combination it realizes. On this account, no effect of context is expected.

3   Experimental study

The goal of the present study was to gain insight into the preference patterns for German response particles. For affirming responses, Krifka predicts a default preference for *nein* over *ja* and a reversed preference pattern for contexts in which $\bar{p}_{\mathrm{DR}}$ is salient. In contrast, R&F predict a general preference for *nein* over *ja* without contextual modulation. For rejecting responses, Krifka predicts a preference for *nein* over *ja*, based on the assumption that *ja* is blocked due to the presence of *doch* in the system, whereas *nein* is not blocked but dispreferred. R&F predict no difference in (dis)preference as they assume that both *ja* and *nein* are blocked by *doch*. We tested these predictions in a series of acceptability-judgement experiments. Participants were presented with short dialogues, in which one speaker made a negative assertion, which the other responded to with a response particle. This response was rated by the participants. Every dialogue was introduced by a scene-setting passage in which the context was manipulated to render either p_{DR} or $\bar{p}_{\mathrm{DR}}$ salient, in order to reveal potential context effects.

3.1   Experiment 1

3.1.1   Method

Particpants. 48 students of Humboldt-Universität zu Berlin participated in the experiment. All participants were native speakers of German. They received payment for their participation.

Materials. There were 48 experimental items and 16 fillers. The dialogues in the experimental items were preceded by a scene-setting passage, which introduced two interlocutors and specified the dialogue's context, i.e. what the two interlocutors were talking about. The dialogue's context was conveyed by an embedded question with either positive or negative polarity, intended to make p_{DR} or $\bar{p}_{\mathrm{DR}}$ salient. The dialogue consisted in a negative assertion and a response to it. The response comprised a *ja* or *nein* and a follow-up phrase, which made clear whether the antecedent assertion was rejected (positive response clause polarity) or affirmed (negative response clause polarity). Table 1 shows an example item, translated to English. The fillers were similar to the experimental items, apart from having a positive antecedent assertion.
All negative versions of the context sentence, antecedent, and response clause contained the adverb *noch* ('yet'). All positive versions contained the adverb *schon* ('already'). To encourage the participants to read the scene-setting passages and dialogues carefully, each item was followed by a statement, which participants had to verify and which pertained to the content of the scene-setting passage or dialogue.

Table 1: Sample of the experimental items in Experiment 1, translated from German

Setting:	A couple of weeks ago Hildegard and Ludwig asked their gardener to redesign the back garden of their holiday home. Now they are chatting about
Positive Context: salient p_{DR}	what the gardener has done already./
Negative Context: salient $\bar{p}_{DR}$	what the gardener hasn't done yet.
Antecedent:	The gardener hasn't sown the lawn yet.
Rejecting response:	Yes/No, he has sown the lawn already./
Affirming response:	Yes/No, he hasn't sown the lawn yet.

Design and Procedure. Experiment 1 had a 2x2x2 within-subject design with the factors CONTEXT POLARITY (positive/negative), RESPONSE PARTICLE (*ja/nein*), and RESPONSE CLAUSE POLARITY (positive/negative). The 48 experimental items were assigned to eight sets. Sets and conditions were counterbalanced across participants. Experimental and filler items were presented to the participants in six different orders. Participants had to rate the naturalness and suitability of the *ja/nein*-response with regard to the context and dialogue on a scale from 1 ('very bad') to 7 ('very good').

3.1.2 Results and Discussion

All analyses reported in this paper were conducted by linear mixed-effects modelling with backward model selection. The final model for the data of Experiment 1 included a random intercept for participants and a random participant slope of RESPONSE POLARITY, RESPONSE PARTICLE, and their interaction. It indicated significant effects of CONTEXT (b=0.20, SE=0.04, t=4.65), of RESPONSE POLARITY (b=1.72, SE=0.17, t=9.96), and of RESPONSE PARTICLE (b=-0.97, SE=0.18, t=-5.49). These effects were qualified by significant interactions of CONTEXT with RESPONSE POLARITY (b=0.30, SE=0.08, t=3.42), and of RESPONSE POLARITY with RESPONSE PARTICLE (b=4.52, SE=0.49, t=9.32). To unpack the interactions separate analyses for the two RESPONSE POLARITY conditions were conducted. The model for the 'positive response clause' conditions (i.e. rejecting responses) did not reveal a significant effect of CONTEXT (b=0.05, SE=0.06, t=0.92). The effect of RESPONSE PARTICLE was significant (b=-3.23, SE=0.30, t=-10.70). As displayed in Table 2, *ja* received lower ratings than *nein*. The model for the 'negative response clause' conditions (i.e. affirming responses) indicated a significant effect of CONTEXT (b=0.34, SE=0.06, t=5.44); ratings were lower in the 'positive context' conditions than in the 'negative context' conditions (see Table 2). Moreover, there was a significant effect of RESPONSE PARTICLE (b=1.29, SE=0.30, t=4.34) with higher ratings for *ja* than for *nein* (see Table 2).

Table 2: Mean Ratings in Experiment 1 (rating scale from 1 ('very bad') to 7 ('very good'))

Response Polarity	Positive Context			Negative Context		
	ja	*nein*	*95% CI*	*ja*	*nein*	*95% CI*
Positive (rejecting)	2.16	5.24	±0.45	2.06	5.44	±0.42
Negative (affirming)	5.91	4.30	±0.43	6.26	4.96	±0.43

Note. The 95% *CIs* are within-subject confidence intervals (Masson and Loftus 2003) associated with the particle effect in the respective context and response polarity condition.

The results for rejecting responses, with low ratings for *ja* and significantly higher ratings for *nein*, suggest that only *ja* but not *nein* is blocked by *doch*. A further experiment, including *doch* as an additional level of the factor RESPONSE PARTICLE, yielded significantly higher ratings for *doch* (M=6.76) compared with *nein* (M=3.84) and *ja* (M=1.81), and replicated the significant difference between *nein* and *ja*, thereby suggesting that the finding of Experiment 1 did not rest upon the absence of *doch* in the experimental situation. The results for affirming responses were neither consistent with Krifka's predictions nor with those by R&F. Against both accounts, the ratings indicate an overall preference for *ja* over *nein* rather than for *nein* over *ja*, and against Krifka's account, the preference pattern was not modulated by the context manipulation.

3.2 Experiment 2

Experiment 2 investigated if the unpredicted results found for the affirming conditions of Experiment 1 could be replicated for bare particle responses.

3.2.1 Method

Participants. 26 students of Humboldt-Universität zu Berlin participated in this experiment. The data of two participants were excluded from analysis as they did not perform significantly better than chance in the verification task or did not follow the instructions. As in Experiment 1, participants received a payment for participating.

Materials. The responses in Experiment 2 consisted of bare particles only. To make clear whether a bare *ja* or *nein* should be taken as an affirming response, the items contained information on the knowledge of the answering person in the scene-setting passage. For the experimental items, this information indicated that the knowledge of the second person was consistent with the negative utterance by the first person. For the example item of Experiment 1, shown in Table 1, the additional information in Experiment 2 would be *In the morning, Ludwig ran into the gardener, who told him that he can only sow the lawn in a couple of days, due to the weather*. A further modification of the material was motivated by the overall lower ratings in the 'positive context' conditions for the affirming responses in Experiment 1, which may suggest that the dialogues in the 'positive context' conditions were generally perceived as less coherent. In Experiment 2, the positive context (which had been intended to induce a salient p_{DR}) was replaced with a 'neutral' context (e.g. for the sample item of Experiment 1: *They are talking about the gardener and the redesigning of their garden*), in which the p_{DR} was assumed to be salient by default. In total, the materials comprised 24 experimental items and 40 fillers. Sixteen fillers had a positive antecedent. The remaining 24 fillers had a negative antecedent followed by a rejecting response.

Design & Procedure. Experiment 1 employed a 2x2 within-subject design, with the factors CONTEXT (neutral/negative) and RESPONSE PARTICLE (*ja*/*nein*). The procedure was the same as in Experiment 1.

3.2.2 Results and Discussion

The final model for the data of Experiment 2 included a random intercept for participants and a random participant slope of RESPONSE PARTICLE. It revealed a significant effect of RESPONSE PARTICLE (b=1.67, SE=0.48, t=3.45). As Table 3 shows, *ja* received higher ratings than *nein*. Model comparison neither yielded a better fit for a model including the factor CONTEXT nor for a model including the interaction of RESPONSE PARTICLE and CONTEXT.

The results replicate the unexpected finding for the affirming responses of Experiment 1 and extend it to bare particles. Inconsistent with both R&F and Krifka, the ratings obtained in Experiment 2 suggest that bare *ja* is preferred over bare *nein* without contextual modulation.

Table 3: Mean Ratings in Experiment 2 (rating scale from 1 ('very bad') to 7 ('very good'))

Neutral Context			Negative Context		
ja	*nein*	**95% *CI***	*ja*	*nein*	**95% *CI***
5.83	4.31	±0.71	5.99	4.17	±0.75

Note. The 95% *CIs* are within-subject confidence intervals (Masson and Loftus 2003) associated with the particle effect in the respective context condition.

4   General Discussion

Our study aimed at investigating preference patterns for the German particles *ja* and *nein* in responses to negative assertions. For rejecting responses, the results from two experiments (Experiment 1 and the additional experiment including *doch*) indicate that *nein* is preferred over *ja*, suggesting that *ja* but not *nein* is blocked by *doch*. For affirming responses, the results of Experiment 1 and 2 revealed an overall preference of *ja* over *nein* without contextual modulation. This finding is inconsistent with the predictions of both Krifka and R&F.

However, a closer data inspection, i.e. comparing each participants mean rating for *ja* with the mean rating for *nein*, revealed differences among participants. A majority of approximately 70% of the participants of Experiment 1 and 2 showed the unpredicted pattern of a higher mean rating for *ja* compared with *nein*. Approximately 30% of the participants displayed a different pattern, with a higher mean rating for *nein* compared with *ja* for all but one participants in this subgroup. The two figures below illustrate the individual differences between the ratings for *ja* and *nein*; they show the difference score per participant (calculated by subtracting the mean rating in the *nein* condition from the mean rating in the *ja* condition after z-value transformation per participant). A positive difference score indicates a higher mean rating for *ja* compared with *nein* and a negative difference score indicates the reverse pattern.
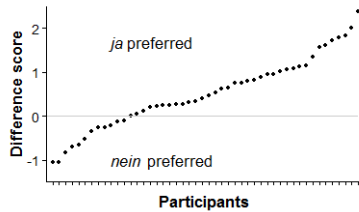
Figure 1: Difference scores between ratings for *ja* and *nein* for each participant in Expt. 1
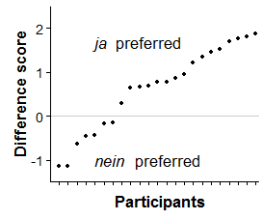
Figure 2: Difference scores between ratings for *ja* and *nein* for each participant in Expt. 2

As can be seen from Figures 1 and 2, both the positive and negative difference scores have some variability.[2] Yet, most of the participants have a fairly large difference score, indicating that they have a clear preference for either *ja* or *nein*. Thus, these participants fall into two subgroups: a *ja*-group and a *nein*-group. Remarkably, the two groups also differ in the ratings

---

[2]For some participants, the difference score was close to (or equal to) zero. This could reflect a lack of a pronounced preference for either of the two response particles. However, it should be kept in mind that the difference scores result from acceptability judgments rather than from production data. When judging acceptability, participants might allow for variation that they are used to in commonplace conversational settings, like other speakers having different preferences. Hence, the difference scores obtained with the present method may underestimate preference strength. Indeed, in a pilot production study, each participant showed a clear preference, with the majority preferring *ja* and a notable minority preferring *nein* in affirming responses to negative antecedents, similar to the main finding of the present experiments.

for *nein* as a rejecting response. Overall, participants with a preference for *ja* as an affirming response rated rejecting responses with *nein* higher than did participants with a preference for *nein* as an affirming response (M=5.00 vs 3.47), with the difference being especially pronounced with bare particles (M=4.49 vs 2.01). This suggests that *nein* may be a suitable alternative to *doch* for the *ja*-group but not for the *nein*-group.

In the following, we will discuss how the two groups can be accounted for in the frameworks of R&F and Krifka.

Accounting for the two groups in R&F's framework. One obvious way to account for the two groups in the framework of R&F it is to assume that the *ja*-group and *nein*-group differ in the feature realization potential of *ja* and *nein*. The *ja*-group prefers *ja* over *nein* in affirming responses to negated antecedents. Participants in this group seem to apply a truth-based response strategy (Jones 1999), with *ja* signaling the truth and *nein* the falsity of the antecedent. This could be captured by the two relative polarity features [AGREE] and [REVERSE]. Thus, in the *ja*-group *ja* realizes [AGREE] and *nein* realizes [REVERSE] (see Table 4). In contrast, participants in the *nein*-group have a preference of *nein* over *ja* in affirming responses to negated antecedents. The *nein*-group seems to apply a polarity-based strategy (Jones 1999), with *ja* signaling that the response clause has positive polarity and *nein* signaling a negative response clause polarity. Hence, in the *nein*-group *ja* realizes the absolute polarity feature [+] and *nein* realizes [−] (see Table 4).

Table 4: Realization potentials of *ja*, *nein* and *doch* for the two groups

Particle	*ja*-group	*nein*-group
ja	[AGREE]	[+]
nein	[REVERSE]	[−]
doch	[REVERSE,+]	[REVERSE,+]

Due to the presence of a third form, the specialized particle *doch*, the response systems for the two groups can neither be purely truth-based nor polarity-based. Thus, for both groups *doch* realizes the feature combination [REVERSE,+]. However, there is a general issue with this account. Upon closer scrutiny, it turns out that the absolute polarity features are problematic because they only impose a presupposition on the polarity of the response clause, but not its meaning: roughly, the response clause must denote a proposition that is 'highlighted', i.e. a propositional discourse referent being introduced by a preceding utterance (R&F). This means that there is no restriction on the meaning of the response clause in the case of absolute polarity features. Hence, a bare *no* in response to a positive assertion like *Jim snores* could be taken to mean e.g. *No, Mary doesn't know*, if the proposition *Mary doesn't know* was part of the context. To fix this problem, one could define the absolute polarity features with an additional condition such that the proposition p denoted by the response clause must either be the complement of, or semantically identical to a salient proposition q in the context. Yet, such a step would render the absolute polarity features relative.

Accounting for the two groups in Krifka's framework. Recall that the core of Krifka's proposal is that negated propositions introduce two propDRs in the discourse: p_{DR} and $\bar{p}_{DR}$. Furthermore, *ja* asserts the propDR it targets and *nein* asserts the negation of the propDR it targets. In this framework, the two groups can be accounted for by assuming that they differ in which propDR *ja* and *nein* target in the case of a negative antecedent, i.e. when both the p_{DR} and $\bar{p}_{DR}$ are available: we propose that the *ja*-group prefers $\bar{p}_{DR}$, whereas for the *nein*-group both p_{DR} and $\bar{p}_{DR}$ are equally good candidates. The difference to Krifka's original proposal is

that there is no default preference to target p_{DR} and no context modulation.

Ja-group. For the *ja*-group, we assume that $\bar{p}_{\text{DR}}$ is more salient than p_{DR}. This may be due to $\bar{p}_{\text{DR}}$ being introduced by a non-embedded constituent, whereas p_{DR} is introduced by an embedded constituent. Evidence for acceptability differences between embedded and non-embedded material stems from a study by Gordon et al. (1999). This study indicates that after processing utterances like the one in (9), the DR of *Bill's aunt* is more easily accessible than the DR of *Bill*.[3]

(9)   Bill's aunt owns a lake house.

As a result of the preference for targeting $\bar{p}_{\text{DR}}$, the particle *ja* unambiguously asserts $\bar{p}_{\text{DR}}$ and *nein* asserts the negation of $\bar{p}_{\text{DR}}$, see Table 5. We will deal with the use of *doch* after discussing the *nein*-group.

Table 5: Meaning of the response particles for negative antecedents in the *ja*-group

Particle	Reference	Meaning	Presupposition
ja	$\bar{p}_{\text{DR}}$	$\bar{p}_{\text{DR}}$	
nein	$\bar{p}_{\text{DR}}$	$\neg\bar{p}_{\text{DR}} \equiv p$	
doch	$\bar{p}_{\text{DR}}$	$\neg\bar{p}_{\text{DR}} \equiv p$	$\bar{p}_{\text{DR}}$ is available

Nein-group. For this group, we assume that both available propDRs, p_{DR} and $\bar{p}_{\text{DR}}$, do not differ in saliency. As a consequence, the use of *ja* in responses to negative assertions is ambiguous, because it can target and assert both p_{DR} and $\bar{p}_{\text{DR}}$. We suggest that the *nein*-group therefore avoids the use of *ja*, see Table 6. As for *nein*, its use would in principle be ambiguous as well: *nein* can assert either the negation of p_{DR} or of $\bar{p}_{\text{DR}}$. However, these two options are cognitively asymmetrical, since the latter one involves double negation. Therefore, we assume that speakers of the *nein*-group prefer targeting p_{DR}, as shown in Table 6.

Table 6: Meaning of the response particles for negative antecedents in the *nein*-group

Particle	Reference	Meaning	Presupposition	Use
ja	p_{DR}	p_{DR}		Avoid
	$\bar{p}_{\text{DR}}$	$\bar{p}_{\text{DR}}$		Avoid
nein	p_{DR}	$\neg p_{\text{DR}}$		✓
	$\bar{p}_{\text{DR}}$	$\neg\bar{p}_{\text{DR}} \equiv p$		Avoid
doch	$\bar{p}_{\text{DR}}$	$\neg\bar{p}_{\text{DR}} \equiv p$	$\bar{p}_{\text{DR}}$ is available	✓

Regarding the meaning of *doch*, we assume that the two groups do not differ. We propose that *doch* targets $\bar{p}_{\text{DR}}$ and asserts its negation; thus, *doch* presupposes that $\bar{p}_{\text{DR}}$ is available. For rejecting negated propositions, *doch* is favored over other particles with the same meaning, due to Maximize Presupposition (Heim 1991). Thus, for the *ja*-group *nein* is dispreferred to *doch*, in rejecting responses to negative antecedents. However, as our findings suggest, speakers from this group judge *nein* as quite acceptable in such conditions.

Conclusion. The main result of our experimental study is the finding of two subgroups of participants, differing in the preference patterns for the German response particles *ja* and *nein* as affirming responses to negative assertions. To account for the two groups we discussed

---

[3]We thank Massimo Poesio for pointing out this reference.

possible modifications of the R&F and Krifka frameworks. Regarding the R&F framework, there is a theory-internal issue with the notion of absolute polarity features, as outline before. As for the Krifka framework, it is an empirical task to evaluate the validity of the modified assumptions on saliency, that are proposed in this paper.

References

[1] Hardarik Blühdorn. *Negation im Deutschen*. Narr Verlag, 2012.

[2] Adrian Brasoveanu, Donka Farkas, and Floris Roelofsen. N-words and sentential negation: evidence from polarity particles and VP-ellipsis. *Semantics and Pragmatics*, 6(7), 2013.

[3] Peter C. Gordon, Randall Hendrick, Kerry Ledoux, and Chin Lung Yang. Processing of reference and the structure of language: An analysis of complex noun phrases. *Language and Cognitive Processes*, 14(4):353–379, 1999.

[4] Jeanette K. Gundel, Nancy Hedberg, and Ron Zacharski. Cognitive status and the form of referring expressions in discourse. *Language*, 69:274–307.

[5] Irene Heim. Artikel und Definitheit. *Semantik: ein internationales Handbuch der zeitgenössischen Forschung*, pages 487–535, 1991.

[6] Anders Holmberg. The syntax of answers to polar questions in English and Swedish. *Lingua*, 128:31–50, 2013.

[7] Bob Morris Jones. *The Welsh answering system*, volume 120. Walter de Gruyter, 1999.

[8] Elena Karagjosova. The German response particle 'doch' as a case of constrastive focus. In *Proceedings of the SPRIK Conference. Explicit and implicit information in text. Information structure across languages*, 2006.

[9] Ruth Kramer and Kyle Rawlins. Polarity particles: an ellipsis account. In *Proceedings of the 39th Annual Meeting of the North East Linguistic Society (Volume 1)*, 2011.

[10] Ruth Kramer and Kyle Rawlins. An ellipsis approach to polarity particles across languages. Presentation at Workshop on the syntax of 'yes' and 'no', Newcastle, 2012.

[11] Manfred Krifka. Response particles as propositional anaphors. In *Proceedings of the 23rd Semantics and Linguistic Theory Conference*, May 2013.

[12] Micheal E.J. Masson and Geoffrey R. Loftus. Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57:203–220, 2003.

[13] Floris Roelofsen and Donka Farkas. Polarity particle responses as a window onto the interpretation of questionas and assertions. *Language*, 91(2):359–414, 2015.

A Predicativist Semantics of Modals based on Modal Objects

Friederike Moltmann
CNRS-IHPST and NYU
fmoltmann@univ-paris1.fr

Abstract

I will outline a novel 'predicativist' semantics of modals in natural language, by focusing not on possible worlds and quantifiers ranging over them, but on what I will call modal objects and their satisfaction conditions.

1 The standard view and the predicativist view of modals

On the standard view, modal expressions, modal verbs and adjectives, act semantically as quantifiers ranging over a (contextually restricted) set of possible worlds. Modals of necessity on the standard view represent universal quantifiers ranging over worlds, whereas modals of possibility represent existential quantifiers ranging over world. Thus the logical forms of (1a) and (1b) will be as in (2a) and (2b) respective:

(1) a. John may leave.
    b. $\exists w\, (w \in f(w_o)\ \&\ [John\ leave]^w = \text{true})$
(2) a. John must leave.
    b. $\forall w(w \in f(w_o) \rightarrow [John\ leave]^w = \text{true})$

Following Kratzer(1977) , the different readings of modal expressions (dentic, epistemic circumstantial etc) are generally accounted for by restricting the set of worlds to a contextually given modal base and ordering source. For motivating the present approach to modals, it is first useful to take into consideration not just modal auxiliaries such as may and must, but the full range of modal predicates in English:

(3) a. might, may, must, should   (modal auxiliaries)
    b. ought to, need to, have to   (modal verbs)
    c. is possible that, is necessary that, is able to, is capable of   (modal adjectives)

There are two semantic issues regarding modal predicates that the standard account does not address, but that bear on the present approach:

[1] What do nominalizations of modal predicates describe?
[2] What if any is the Davidsonian (event) argument of modal predicates?

The present view is that the answers to 1 and 2 will shed light on the semantics of modal verbs and motivate the view that modals are predicates of 'modal objects', the Davidsonian arguments of modal predicates.

What are modal objects? Modal objects are the sorts of entities we would refer to with nominalizations of modal predicates, that is, entities we refer to as an 'obligation' as a 'permission', as a possibility', a 'necessity' and 'an ability'. Some modal verbs, in particular modal auxiliaries do not come with nominalizations that could serve to refer to modal objects. Nonetheless they will involve modal objects in their semantics. I take modal objects to be the implicit, Davidsonian arguments of modal predicates. Modal objects as implicit argument of modal predicates are the things adverbials are predicated of. However, modal objects are not events or states. They differ in their properties from the latter, in particular in being able to in having satisfaction conditions and thus satisfiers (or violators).[1]. An obligation can be fulfilled and an invitation accepted, but not so for a state or event. Modal objects are in that respect on a par with certain cognitive and illocutionary products, in the sense of Twardowski's (1911) distinction between actions and products (see also Moltmann 2013, Chap. 4 and Moltmann 2014, to appear). Cognitive products are the nonenduring products of cognitive act. Thus a judgment is the cognitive product of an act of judging and a decision is the cognitive product of an act of deciding. Similarly, a command is the illocutionary product of an act of commanding, and a promise the illocutionary product of an act of promising. A command can be complied with, but not the act of commanding and a promise can be broken, but the act of promising. Some modal objects are also products of illocutionary acts. Thus an obligation may be the product of an act of commanding or an act of promising, and a permission the product of an act of offering. Such modal objects thus are modal products. Modal products also include laws, the products of acts of declaration or passing. Note that laws are entities not tied to particular nominalizations. Not all modal objects are the products of intentional acts, though. Some modal objects exist without the intentionality of an agent, for example abilities and perhaps metaphysical and logical necessities and possibilities.

The semantic proposal then is that (4a), with a modal of necessity, has roughly the logical form of (4b), namely as in (4c):

(4) a. John needs to leave.
  b. John has a need to leave.
  c. $\exists d(need(d)$ & $[John\ to\ leave](d))$

A modal of possibility as in (5a) leads to the very same logical form, namely as in (5c), which is also roughly, the logical form of (5b):

(5) a. John is permitted to leave.
  b. John has a permission to leave.
  c. $\exists d(is\ permitted(d, John)$ & $[John\ to\ leave](d))$

In fact, there are even syntactic arguments in favor of the logical form in (4c) and in (5c). Thus, Harves / Kayne (2012) argue that the sentence (14a) is derived from (14b), with the noun *need* being prior in the linguistic derivation to the verb *need*.

The logical form of modal sentences with modal auxiliaries will similarly be as below:

(6) a. John must help.
  b. $\exists e(must(e)$ & $[John\ help](e))$

---

[1] The ability of modal objects to have satisfaction conditions is reflected in natural in the *by-* locution, as below:

 (1) a. John's need was fulfilled by having X be done.
    b. The obligation was met by doing X.

(7) a. John may leave.

    b. ∃e(may(e) & [*John leave*](e))

2 Truthmaker semantics for modals

How can clauses act predicates of modal objects? On the present view they do so by specifying the truthmaker and falsifiers or better satisfiers and violators of the modal object in roughly the sense Kit Fine's (2912, 1914, to appear) truthmaker semantics. Central on Fine truthmaker semantics is the notion of exact truthmaking or satisfaction, which, for Fine, holds between an action or situation and a sentence S iff s is wholly relevant for the truth of S. I will apply the notion of exact truthmaking/satisfaction also to modal objects (as well as cognitive and illocutionary products). Thus, satisfiers of a modal object are situations or actions fulfilling or complying with the condition associated with the modal object, and as such they should be exact satisfiers of the modal object, in the sense of Fine's exact truthmaking relation. Violators of modal objects are situations or actions incompatible with or contravening the condition associated with the modal object. Thus, John's obligation to help has as satisfiers actions of John helping and as violators actions of refraining to help. What distinguishes John's obligation to leave from John's permission to leave is that the former has satisfiers and violators, whereas the latter has only satisfiers, namely actions of John of leaving. That is, permission only enable things, but do not exclude anything, whereas obligations enable some things and exclude others. The difference between modal predicates of necessity and of possibility thus is not traced to a difference in logical form, as universal and existential quantification over worlds, as on the standard view. Rather it consists entirely in a difference in the satisfaction conditions of the two sorts of modal objects, that is, in the nature of the two sorts of modal objects themselves.

Not only modal objects have satisfiers and violators, also attitudinal objects may and that can account for relevant semantic connections between attitude verbs and modal verbs.

Thus, a modal object of obligation that results from an act of requesting may share its satisfiers with the request, the illocutionary product resulting from the act of requesting.

On Fine's (2012, 2014, to appear) truthmaker semantics, a sentence S has as its meaning a pair consisting of a set of (exact) truthmakers (situations or actions wholly relevant for the truth of S) and a set of exact falsemakers (or violators). The exact truth making relation applies to more complex sentences as below, which are fairly standard:

(8) a. s ⊩ P *and* Q iff for some s' and s'', s = sum(s', s'') and s' ⊩ P and s'' ⊩ Q.

    b. s ⊩ P *or* Q iff s ⊩ P or s ⊩ Q

    c. s ⊩ ∃x S iff s ⊩ S [x/t] for some term t

What is new on Fine's truthmaker semantics is the relation of (exact) falsemaking ⫣ (s ⫣ S: S is false in virtue of s). The falsemaking relation is involved in the truthmaking condition on negative sentences:

(9) s ⊩ *not* S iff s ⫣ S.

That is, a situation s is an exact truthmaker of *not* S in case s is a falsemaker of S.

Imperatives have as their meaning a pair consisting of a set of actions complying with the imperative and a set of actions contravening the imperative (Fine, to appear).

On the present approach, the relation of exact satisfaction and violation will also be a relation between situations or actions and modal objects. Modal objects of necessity and modal objects of possibility differ in whether they have violators. Modal objects of necessity have both satisfiers and violators, but modal objects of possibility have only satisfiers and lack violators. An obligation can be satisfied by actions and it can be violated by actions. By contrast, a permission, an invitation, or an offer only sets up options: actions of satisfying or 'taking up' the permission, invitation, or offer. The predicative denotation of a

clausal subject or complement (or a prejacent) S can then be given as a property of (modal) objects pred([S]), as below, where pos(S) is the set of verifiers of S and neg(S) the set of falsifiers of S:

(10) pred([S]) = λd[∀s(s ⊩ d → s ∈ pos(S)) & ∀s(s ⊪ d → s ∈ neg(S))]

That is, the predicative denotation of a sentence S is the property that holds of a modal object d just in case all exact satisfiers of d are exact truthmakers of S and all exact violators of d are exact falsemakers of S. (10) is suited to characterize both modal objects of necessity (the second conjuncts applying non-vacuously) and modal objects of possibility (the second conjunct applying vacuously).

The duality of modals of necessity and of possibility (must ¬ S ↔ ¬ may S) is straightforwardly accounted for once the existential quantifiers in the logical form of modal sentences are allowed to quantify over a contextually highly restricted domain. A modal product whose satisfiers make S false and whose violators make S true is not a modal product that has only satisfiers that make S true, and vice versa. Suppose the domain contains just one relevant modal object, the obligation for John to help. Then *John is obliged to help* implies *John is not permitted not to help* because the modal object of obligation just is not a modal object that has only satisfiers, namely actions of John's not helping. Similarly, if there is just one relevant modal object of the permission for John to leave the house, *John is allowed to leave the* house implies *John is not obliged not to leave the house* because then the modal object of John's permission is not a modal object that has as satisfiers actions of not leaving the house and as violators actions of leaving the house. The contextual restriction also takes care of Kratzer's (1977) insight that modals generally have a highly context-dependent interpretation.

The predicativist account applies particularly well to deontic modals, since we have a good sense of obligations and permissions as objects having particular sorts of satisfiers and violators. But of course the predicativist account is meant to apply to all modals. Application to other modalities. Going through the various sorts of (readings) of modals must await other occasions. Here I just give some indications of how the account extends. First, the account has a plausible application to ability modals:

 (11) John is able to / can walk.

Abilities are modal objects, though not modal products, the products of illocutionary ir cognitive acts. There is a good intuitive sense in which of an ability has what should take the role of satisfiers, namely its physical manifestations. Abilities are like permissions and not like obligations in that they only have satisfiers and not violators.

The account should also extend to epistemic modals, though the modal objects here are somewhat less straightforward. It appears there may be a variety of ways of introducing modal objects for epistemic modals. For example, a modal object for epistemic *must* maybe 'generated' by a piece of evidence and assigned situations supported by the evidence as satisfiers and situations excluded by the evidence as falsifiers. The same piece of evidence may alternatively generate a modal object of possibility, an object assigned only satisfiers, situations supported by the evidence. Instead of a particular piece of evidence, a set of accepted facts or the 'common ground' may generate modals objects in that way.
Epistemic modal verbs have approximatively the following analysis:

(12) a. John must be at home.
     b. λe[∃d (dRe & must(d) & [*John be at home*](d))]

Here e is meant to be the speech event, and R a relation of 'close connection'

3 Connections between modals and propositional attitudes

As mentioned, not only modal objects have satisfiers and possibly violators, also attitudinal objects may do so. Thus requests have satisfiers and violators and invitations have only satisfiers. (13a) and (13b) will then have the same logical form, as in (14a) and in (14b) respectively, based on the predicative meaning of the complement clause in (14c):

(13) a. John asked Mary to come.
    b. John invited Mary to come.
(14) a. $\exists e(ask(e, John, Mary)$ & $[Mary\ come](product(e)))$
    b. $\exists e(invite(e, John, Mary)$ & $[Mary\ come](product(e)))$
    c. $[Mary\ come_{emb}] = \lambda d[\forall s(s \parallel d \rightarrow s \in pos(Mary\ come))$ & $\forall s(s \parallel d \rightarrow s \in neg(Mary\ come))]$

Moreover modal objects may share their satisfiers and possibly violators with attitudinal objects. In general, the modal product that may result from an illocutionary act shares its satisfiers and possibly verifiers with the illocutionary product resulting from that very same act. Thus sharing of satisfaction conditions explains the validity of the inferences below, on a suitable reading of *must* and *may*:

(15) a. <u>John asked Mary to leave.</u>
    Mary must leave.
    b. <u>John offered Mary to take an apple.</u>
    Mary may take an apple.

Modal products, though, differ in nature from the corresponding illocutionary products.
Illocutionary products can generally not endure past the time of the illocutionart act, but modal products can. This is reflected in the choice of tense with modal products, as opposed to illocutionary products as below:

(16) a. Yesterday, John promised to help (today).
    b. (Today) John's obligation is to help.
    c. John's promise was / ??? is to help.

The illocutionary product that is the product of John's act of promising will not last beyond the act of promising, but the resulting modal product, the obligation will. The illocutionary act produces both a (non-enduring) illocutionary product and an (enduring) modal object, and the illocutionary product and the modal object share exactly the same satisfiers (and violators).

The nouns *permission* and *offer* are actually ambiguous between describing illocutionary products and describing modal products, the latter being their meaning in the sentences below:

(16) d. John still has the permission to use the huse.
    e. The offer still stands.

The fact that modal objects may be the result of illocutionary acts explains certain cases of modal concord or 'harmonic' modals, as below:

(17) a. John requested that Mary *should* leave.
    b. John offered Mary that she *could* use the house.

The occurrence of the modal in such sentences is best seen as a performative use of a modal in an embedded context. In independent sentences, performative uses of modals as in (18a) and (19a) lead to statements roughly equivalent to the use of sentences with a performative use of an illocutionary verb – one of request as in (18b) with a modal of necessity and one of offer as in (19a), the two of which correspond to the two uses of the imperative, as in (18c) and (19c):

(18) a. You must leave.

b. I hereby ask that you leave.

c. Leave!

(19) a. You may take an apple.

b. I hereby offer you to take an apple.

c. Take an apple!

Independent sentences within the present view are considered predicates of illocutionary products or modal products produced by the illocutionary acts performed by uttering the sentence. An imperative will thus denote the property λd[[*leave!*](d)], to be predicated of the illocutionary product meant to be produced by the utterance of the sentence. This then allows the analysis of performatives as in (18a) and (19a) as roughly in (20a) and (21a), and similarly of the performative modals in (18b, 19b) as in (20b) and (21b):

(20) a. λe[ask(e, speaker) & [(addressee) *leave*](product(e))]

b. λd[must(d) & [(addressee) *leave*](d)]

(21) a. λe[offer(e, s) & [(addressee) *take an apple*](product(e))]

b. λd[may(d) & [(addressee) *take an apple*](d)]

Harmonic modals such as *should* in (17a) will then contribute to the meaning of the embedded sentence as below:

(22) [*that Mary should leave*] = λd[should(d) & [*Mary leave*](d)]

There will be two interpretations of illocutionary act reports. One of them involves predicating the clausal complement of the illocutionary product of the described illocutionary act, the Davidsonian event argument of request. The other involves predicating the clausal complement of the modal product of the illocutionary act that is the Davidsonian event argument. The latter is of course the one relevant for harmonic modals as in (17a), as in the analysis below:

(23) ∃e(request(e, John) & [*that Mary should leave*](modal-product(e)))

The very same account is applicable to harmonic modals of possibility such as could in (17b).

The predicativist account also has a nice application to modal concord, as below:

(24) a. John could possibly have missed the train.

b. John must obligatorily fill out the form.

c. John may optionally fill out the back of the form.

On the present view, modal adverbs just like modal verbs act as predicates of modal objects. Then (24b), on the modal concord reading, involve simple predication of two modal predicates of the same modal object, as below:

(25) ∃d(must(d) & obligatorily(d) & [*John fill out the form*](d))

Of course there is in principle another reading available, on which the modal adverb and the modal verb each introduce an existential quantifier ranging over modal objects.

The account immediately explains a constraint on modal concord, namely that the two modals be of the same sort (both need to be modals of necessity or modals of possibilities):

(26) a. ??? John must possibly have missed the plane.

b. ??? John may obligatorily fill out the form.

This constraint does not come out on the one alternative semantic account of modal concord offered in the literature within the standard view, namely according to which the two modals apply to the same modal base (Anand, P. / A. Brasoveanu 2010).

References

Anand, P. / A. Brasoveanu (2010): 'Modal Concord as Modal Modification'. In *Proceedings of Sinn and Bedeutung* 14, 19-36

Fine, K. (2012): 'Counterfactuals without Possible Worlds'. Journal of Philosophy 109 (3), 221-246

---------- (2014): 'Truth-Maker Semantics for Intuitionistic Logic'. Journal of Philosophical Logic 43, 2-3, pp 549-577

--------- (to appear): 'Truthmaker Semantics'. In Blackwell Philosophy of Language Handbook.

Harves, S. / R. Kayne (2012): 'Having 'need' and 'needing' have'. *Linguistic Inquiry* 43. 120–132.

Kratzer, A. (1977): What 'must' and 'can' must and can mean'. *Linguistics and Philosophy* 1, 335 – 315.

Moltmann, F. (2013): *Abstract Objects and the Semantics of Natural Language*. Oxford UP, Oxford.

--------------- (2014):'Propositions, Attitudinal Objects, and the Distinction between Actions and Products'. *Canadian Journal of Philosoph*y, supplementary volume on propositions, edited by G. Rattan and D. Hunter, 43.5-6, pp. 679-701.

--------------- (to appear): 'Cognitive Products and the Semantics and Attitude Verbs and Deontic Modals'. To appear in F. Moltmann / M. Textor (eds.).

Moltmann, F. / M. Textor (eds.) (to appear): *Act-Based Conceptions of Propositions: Contemporary and Historical Contributions.* Oxford UP, Oxford, to appear in 2016.

Twardowski, K. (1911): 'Actions and Products. Some Remarks on the Borderline of Psychology, Grammar, and Logic'. In J. Brandl/J. Wolenski (eds.): *Kazimierz Twardowski. On Actions, Products, and Other Topics in the Philosophy*. Rodopi, Amsterdam and Atlanta, 1999, 103-132. Reprinted in Moltmann / Textor (to appear).

His name is Socrates because that's what he's called:
A model-theoretic account of name-bearing

Patrick Munoz

University of Chicago
pjmunoz@uchicago.edu

Abstract

The notion of name-bearing, and its relation to the use of proper names to refer to individuals, has long remained mysterious in linguistics. In the present paper, I provide a novel semantics for proper names as variable expressions that enforce restrictions on their syntactic indexation, in order to give a precise model-theoretic definition of name-bearing that explicitly depends on the notion of reference to individuals using proper names.

1   Introduction

Semantic accounts of proper names traditionally propose that all literal, linguistically appropriate uses of a proper name to refer to an individual require that the individual bear the name. As Bach ([1]: 371) puts it:

> Socrates is called 'Socrates' because he has the property of bearing the name 'Socrates.'
> He is called 'Socrates' because that's his name.

This sounds trivial, but it harbors substantial theoretical commitments. Bach's dictum involves two distinct but related claims: first, that there is a name-bearing relation that holds between a name and an individual, and second, that this relation allows uses of that name to refer to that individual. The former is taken to be a necessary condition on the latter, and since the ability to refer to an individual using a name depends on the name-bearing relation, but not vice-versa, it must be possible to characterize name-bearing independently of reference using a name (see also [6]).

The consensus on this issue in the literature cuts across otherwise disparate approaches to the semantics of proper names.[1] Appeals to name-bearing relations established independently of reference are made at the level of predicative content, when names are treated either as predicates ([2], [5], [15], [19]) or as definite descriptions ([1], [6], [12], [14]); at the level either of indexical character ([17], [20]) or semantic use conditions ([18]), when they are treated as indexicals; and supplementarily in many referentialist approaches, which propose that proper names come to be associated with their conventional referents via procedures establishing name-bearing relations, and that they refer in particular utterances via causal chains of communication tracing back to these procedures ([13]).

Gray ([7]) has recently challenged this picture on both empirical and conceptual grounds. Empirically, he notes that many name-bearing relations are established as a result of speakers' referential habits, rather than vice-versa — these include cases of so-called reference transfer, like Evans' ([4]) famed example of 'Madagascar' coming to refer to an island off the coast of Africa rather than a portion of the continent's mainland. Conceptually, he casts doubt on the claim that the notion of name-bearing can be characterized independently of capacity for

---

[1] Outside this generalization lie older approaches to proper names that have not traditionally been taken up by linguists, such as classical descriptivism and the use of Quinean artificial name-predicates.

reference to begin with: to bear a name itself substantially consists in being able to be referred to using that name. The present proposal is an exercise in formally implementing the spirit of this claim.[2] It is in virtue of a name being able to refer to an individual that the individual bears the name. So Bach's dictum is inverted: His name is 'Socrates' because that's what he's called. My goal is to provide a precise model-theoretic account on these terms of what it means for an individual to bear a name. Because name-bearing is so closely bound to the semantic properties of proper names themselves, this requires a novel account of the semantics of proper names out of which the notion of name-bearing naturally falls.

In broad outline, the proposal is as follows. In line with a standard Kripkean referentialist semantics, proper names are rigidly designating referential expressions of extensional semantic type e. But in contrast with the traditional Millian view, proper names are formally not constant expressions, but variable expressions, as proposed by Cumming ([3]). Their semantic value is a function, often non-constant, from variable assignments to individuals, such that a proper name's referent in a given context is relative (i) to a contextually supplied assignment function, and (ii) to a syntactic referential index. In addition to this basic variabilist framework, the present account proposes two innovations. First, proper names enforce restrictions on which referential indices they can be tagged with: the interpretation function is defined on proper names only when the name is indexed with some member of a proper subset of the set of all possible indices. Second, the set of assignment functions available for interpreting expressions in a language does not necessarily include all formally definable functions from indices to individuals: rather, it includes only an (almost always proper) subset of these.

While the basic variabilist framework provides the core semantic contribution of proper names, these two proposed innovations work in tandem to characterize name-bearing. For example, suppose that the proper name 'John' can only be interpreted when it is syntactically indexed in a certain way, and call the indices that 'John' permits in this way, 'John-indices.' Further suppose that the assignment functions available in the language are a proper subset of all those that are formally definable. Then there might be, or might not be, available assignments that map John-indices to a certain individual. If there are, then that individual is named 'John,' and if not, then it isn't. In other words, given the restrictions that proper names have on the way they can be indexed, which individuals are named what is a function of which assignments are available in the language. An individual bears a name just in case for every index allowed by that name, there exists some assignment that maps that index to that individual. And this in turn happens only if that individual can be referred to using that name, since the individual that the index maps to is a possible referent of the name in some context. Thus, name-bearing depends on the capacity for a proper name to refer to an individual.

This approach allows for a formally definable notion of name-bearing, which until now has remained a black box in linguistics, with almost all authors taking the notion for granted, and at least one prominent attempt to explicate the notion explicitly despairing of giving a unified account of it ([6]). It also yields a robust way to characterize the semantics of different kinds of proper names. The class of proper names is not semantically uniform, as has been previously assumed, and the present approach allows the different kinds of proper names to be characterized in terms of the different kinds of syntactic indices that they permit. At the same time, it allows the class as a whole to be unified under a general characterization of what a proper name is — and this characterization is inherently tied to the notion of name-bearing described here. Section 2 fleshes out and expands the empirical domain of proper names; section

---

[2]The present account is however hugely different from, even incompatible with, Gray's own. Gray assumes that proper names are predicates, and I believe there is strong morphosyntactic evidence to the contrary; the issue outruns the scope of this paper.

3 provides a semantics of proper names and nam-bearing; section 4 briefly concludes.

2   Three kinds of proper names

As mentioned above, the class of proper names is not semantically uniform. I will introduce three distinct kinds of proper names here, though I do not intend this list to be exhaustive.[3]

First, a distinction must be drawn between *shared names* and *unique names.* The intuitive difference between these two is clear enough: shared names exist in a kind of linguistic reservoir, with the understanding that they can be used for any number of individuals, while unique names are tied to a single referent. Shared and unique names are not randomly distributed, but systematically occupy different sections of the lexicon. For example, personal names, both given and familial, tend to be shared ('John,' 'Smith'), while calendrical names, ('Thanksgiving,' 'Tuesday') location names ('Spain'), names of institutions ('The University of Chicago'), and titles of works ('*War and Peace*') tend to be unique.[4]

But beyond this superficial observation, shared and unique names are also semantically distinct. Different sorts of semantic competence are required in order to master the use of shared names versus unique names. It is possible to be fully semantically competent with a shared name without knowing of any particular individuals at all that bear the name. A learner of English knows all there is to know about the semantics of 'John' upon learning that it is a masculine shared name, so long as the learner is competent with shared names generally: it is a matter of indifference to the learner's competence whether he or she knows of any individual to whom 'John' can actually refer, since who bears the name, if anyone, is no part of the word's definition.

This is not the case with unique names, semantic competence with which requires knowledge of their referents. One simply cannot claim to know what 'California' means without knowing that it refers to a particular state. The semantics of a shared name makes no inherent mention of any particular individual, while the semantics of a unique name does — for unique names, the establishment of a name-bearing relation therefore coincides with the coinage of an expression, since the unique individual to which the name refers is hard-wired into its semantics. This difference will be reflected in the proposed semantics in section 3 below.

There is additionally a third kind of proper name, which I term *ephemeral names.* These are far more marginal than shared or unique names: extremely few are present in the English lexicon, and their characteristic 'ephemerality' comes in part from the fact that a large portion of their uses involve nonce-expressions. Some of them have been conventionalized, however, and English examples include 'Captain Obvious,' 'Mr. Right,' and 'Negative Nancy.' Nonce-expression ephemeral names are extraordinarily productive, especially when they include titles of address that signal their status as proper names, e.g. 'Mr. I Don't Know What Time It Is.'[5]

As with predicative expressions, the individuals to which ephemeral names are applicable are only those that bear a certain property: for example, the referent of 'Captain Obvious' is someone who points out the obvious. But these expressions behave morphosyntactically as

---

[3]To give an example of a class of expressions not considered here that might fairly be called proper names: some referential expressions obligatorily take on the form of relational nouns, whose referents must bear the relation denoted by said noun, preferably toward the speaker, such as 'Mom' and 'Teacher.'

[4]These observations about the division of kinds of proper names in the lexicon are only tendencies that serve to elucidate the distinction, and aren't meant to be universally binding within or across languages. In fact it is quite likely that languages differ as to how different parts of the onomasticon are divided between kinds of proper name: for example, the use of name signs in American Sign Language as reported e.g. in [16] seems to hint that there personal names are unique rather than shared.

[5]I owe this example to Itamar Francez.

proper names, and not as uncontroversial predicative expressions like count nouns. Like other proper names, they are referential expressions, and in languages like English that ordinarily prevent proper names from occurring with overt determiners such as the definite article in argument position (see [19]),[6] ephemeral names are bound by that same restriction. Thus ephemeral names pattern in this respect like uncontroversial proper names, and unlike common nouns.

(1)   John needs to be quiet.

(2)   *The John needs to be quiet.

(3)   Captain Obvious needs to be quiet.

(4)   *The Captain Obvious needs to be quiet.

(5)   *Cat needs to be quiet.

(6)   The cat needs to be quiet.

Morphosyntactically, all three of these kinds of proper names look to behave identically, and semantically they have a major feature in common, as will be shown in section 3.4. The task now is: (i), to formally spell out the semantics of each kind of proper name; (ii), to show how the semantics does justice to the empirical observations noted here; (iii), to demonstrate that in spite of their differences, these kinds of proper name deserve to be placed in a common semantic class; and (iv), to provide a formal notion of name-bearing that arises naturally as a result of the proposed semantics.

3   The semantics of proper names and name-bearing

To begin, I assume that a semantic model contains a set of contexts of utterance $\mathcal{C}$, a set of possible worlds $\mathcal{W}$, a set of assignment functions $\mathcal{G}$,[7] a domain of individuals $\mathcal{D}$, and a set of indices $\mathcal{V}$, which is the set of positive integers. The members of $\mathcal{G}$ are then partial functions $\mathcal{V} \nrightarrow \mathcal{D}$, i.e. mappings from indices to individuals. I also assume that for every index $i \in \mathcal{V}$, there is some assignment $g \in \mathcal{G}$ such that $g(i)$ is defined; that is, there are no vacuous indices that never map to anything.

Adopting a framework similar to [10] as adopted in [3], I further assume: (i), that the semantic value of an expression is a function from contexts of utterance to intensions, which are themselves functions from indices of evaluation (here, world-assignment pairs) to extensions; (ii), that contexts are ordered tuples containing several contextual parameter values; and (iii), that among these parameters for any context c is included a contextually provided world of evaluation $c_w \in \mathcal{W}$, and a contextually provided assignment function $c_g \in \mathcal{G}$. Here the first innovation mentioned in section 1 above comes into play: because an expression must be interpreted relative to some member of $\mathcal{G}$, and per the above $\mathcal{G}$ need not contain all formally definable partial functions from indices to individuals, it follows that not all such formally definable functions need be available for linguistic interpretation. That is to say, some *formally* definable assignment functions are *linguistically* impossible to use.

---

[6]Though see [9] for evidence that this restriction can be pragmatically relaxed in the appropriate context, where a name is treated predicatively as part of a definite description.

[7]I assume that the set of assignments is part of the model, because the model represents linguistic competence, and as will be clear below, I believe that knowledge of which formally definable functions $\mathcal{V} \nrightarrow \mathcal{D}$ are members of $\mathcal{G}$ is a kind of linguistic competence.

Proper names are referential variable expressions whose referents are determined relative to an assignment function g. I assume that the interpretation function operates on syntactic structures, and that all proper names are syntactically tagged by some referential index $i \in \mathcal{V}$.[8] The index then acts as an argument to g to determine the extension of the proper name. A basic template for the semantics of a proper name n indexed with i is thus as follows: this is simply a standard interpretation of variable expressions as might be found e.g. in a Traces and Pronouns rule of interpretation, as in [8].

(7) **Core variabilist semantics of a proper name n indexed with i**
$[\![n_i]\!]^{c,w,g} = g(i)$

For the present, I assume that g is always the contextually provided assignment c_g, though nothing in principle prevents expressions from shifting the local assignment, just as modal contexts shift the world of evaluation from c_w.[9] While on this treatment proper names are assignment-sensitive, they are not world-sensitive: the world of evaluation w here makes no non-trivial contribution to a proper name's extension. Proper names are thus rigidly designating in the manner of Kripke ([13]).[10] Finally, as g is contextually provided, and the semantic value of a proper name is a possibly a non-constant function from assignments to individuals, the extension of a proper name may vary from context to context: in this sense, proper names have a non-stable character and so are indexical expressions (though as will be shown in section 3.2, this is not true of unique names, whose extensions are invariant across contexts).

All kinds of proper names share this core semantics. But as stated in section 1 above, they also enforce restrictions on the way that they can be syntactically indexed. What these restrictions consist in constitutes the semantic difference between different kinds of proper names.

3.1 Shared names

Shared names can refer to any number of individuals. Further, there is no grammatical (that is, semantic) reason why shared names can refer to some individuals and not others: name-bearing relations holding between shared names and individuals are grammatically arbitrary, though they are not socially or historically arbitrary.[11] To reflect this, I introduce a restriction on the way in which shared names are capable of being syntactically indexed. Every shared name has associated with it some infinite proper subset of $\mathcal{V}$; tagging a shared name with an index outside of this subset renders the expression uninterpretable. To reflect the grammatical arbitrariness of shared name-bearing, the subset that the shared name allows is itself arbitrarily selected. That is, for each shared name, there is simply some set of indices that it allows, and there is no deeper model-theoretic reason why this set should include some indices and not others.

The semantic entry for a shared name is thus as follows, where s is the semantic type of an index of evaluation (a world-assignment pair), w_s and g_s are the world and assignment contained in the pair s, respectively, and $\mathcal{V}_n \subset \mathcal{V}$ is an infinite set of arbitrarily chosen indices.

---

[8]I remain agnostic here about the syntactic structure of NPs or DPs containing proper names, and how this affects the referential index. The present approach is intended to be compatible with any standard account that involves referential indices. It's then a further question whether it can be imported in spirit to e.g. a variable-free approach.

[9]Quantificational expressions would traditionally be treated as assignment-shifters in this sense, as would the result of Heim and Kratzer-style predicate abstraction. See [3] for an account of belief contexts as assignment-shifters.

[10]Though see section 3.3 below: ephemeral names are world-sensitive in some sense, though this does not threaten their rigidity.

[11]This simplifies to exclude grammatical factors like gender features, which I do not have space to discuss.

(8) **Semantics of a shared name n indexed with i**

    (a) $[\![n_i]\!]^c = \lambda s : i \in \mathcal{V}_n . g_s(i)$

    (b) $[\![n_i]\!]^{c,w,g} = g(i)$, if $i \in \mathcal{V}_n$; else undefined

    (c) $[\![John_1]\!]^{c,w,g} = g(1)$, if $1 \in \mathcal{V}_{John}$; else undefined

(8a) says that the intension of a proper name relative to a context c is a function from world-assignment pairs to individuals. In each case, the individual is $g_s(i)$, the value that results from applying the assignment function of the world-assignment pair to the syntactic index i. This is simply a recapitulation of (7) above. What is novel here is the domain restriction: the function returns $g_s(i)$ only if i is a member of the arbitrarily selected infinite set of indices, $\mathcal{V}_n$. If it isn't, then no value is returned: thus, n only permits interpretation if it is indexed using some member of $\mathcal{V}_n$. (8b) provides an equivalent formulation for the extension of n: relative to a context c, world w, and assignment g, the name's referent is $g(i)$, but only if $i \in \mathcal{V}_n$; otherwise, its value is undefined. (8c) shows an example of the extension of 'John,' indexed with 1.

Shared names require that they be indexed with a member of some arbitrarily selected proper subset of the set of indices. How does this guarantee that shared names can refer to any number of individuals, and moreover, that they can refer to some individuals and not others? The answer lies in the innovation introduced above, viz. that the set of assignments $\mathcal{G}$ need not contain all formally definable partial functions from indices to individuals, but might contain only a proper subset thereof. Suppose for instance that an occurrence of 'John' is indexed with 1, and that $1 \in \mathcal{V}_{John}$, so that the expression is interpretable. It might be, or it might not be, that there exist assignments $g \in \mathcal{G}$ that map 1 to some individual, say d_3, such that $g(1) = d_3$. If there are such assignments, then in some context, d_3 is a possible referent of 'John' indexed with 1, and if there are no such assignments, then it isn't such a possible referent. What this means is that the set of individuals to which a name is capable of referring, given some index, is determined by which functions $\mathcal{V} \nrightarrow \mathcal{D}$ are members of $\mathcal{G}$. The presence or absence of various assignments is what, given the semantic restriction that shared names have on their indexation, determines whether an individual can be referred to using an appropriately indexed shared name. How this ties into name-bearing will be shown in section 3.4 below.

The semantic entry in (8) makes no reference to any particular individual. This fits with what was said in section 2 above, that semantic competence with a shared name requires no knowledge of what bears the name, i.e. to which individuals it can possibly refer. The semantic entry for a shared name is minimal and formulaic: given the knowledge that some lexical item is a shared name, a learner is *ipso facto* equipped with its semantics (excluding complications like gender features). The only semantic difference between different shared names is that they allow indexation with possibly different sets of indices.

The division of labor in becoming competent with the use of a shared name is split in two. On the one hand, one must learn the minimal semantic entry, which yields semantic competence, and on the other, one must learn the status of $\mathcal{G}$, which yields competence in the matter of which individuals bear the name and which do not. There are therefore two corresponding types of incompetence possible with shared names — one might not know what the name means, or one might not know who bears the name.

On the present approach, shared names are not at all ambiguous for the fact that they have multiple possible referents. Referentialist approaches to the semantics of proper names often adopt the so-called ambiguity thesis, popularized in [10] and [11], according to which every instantiation of a shared name is in fact a separate lexical item with a separate semantics. Thus, if two individuals are named 'John,' then they literally bear different homophonous names, and the semantic contents of these names differ, since they have distinct referents. Here, shared

names are not ambiguous in this way: the referentialist approach to the semantics of proper names is preserved while maintaining a single, simple semantic entry for each shared name. When an individual comes (or ceases) to bear a shared name, lexical items are not created (or destroyed). Rather, the membership of $\mathcal{G}$ is altered.

3.2   Unique names

Unique names refer to a single individual, and semantic competence with a unique name requires knowing which individual that is. To capture this, I introduce a new kind of restriction on the indices that a proper name can allow. Unlike with shared names, however, this restriction is not model-theoretically arbitrary, since there is a reason that a unique name allows for some indices and not others: a unique name is inherently tied to a certain individual, and thus it allows only for indices that map only to that individual.

(9)   **Semantics of a unique name n referring to individual d indexed with i**

    (a) $[\![n_i]\!]^c = \lambda s : \forall g : i \in Dom(g)[g(i) = d].g_s(i)$

    (b) $[\![n_i]\!]^{c,w,g} = g(i)$, if for all g such that $g(i)$ is defined, $g(i) = d$; else undefined

    (c)   (where d_1 is Spain):
         $[\![Spain_1]\!]^{c,w,g} = g(1)$, if for all g such that $g(1)$ is defined, $g(1) = d_1$; else undefined

    The intension of a unique name carries a new kind of domain restriction: the function only returns a value if for all assignments in $\mathcal{G}$ that map i to some value, that value is always some single individual d. A unique name therefore always refers to the same individual, when it refers at all.

    In addition to being world-insensitive, and therefore rigidly designating, the extensions of unique names are assignment-insensitive, because their semantics is defined in such a way that no matter which g is chosen, the extension of the name will always be the same individual so long as the name is interpretable. Because the contribution that context makes to the extension of a proper name is to yield an assignment function, it follows that the extension of a unique name is context-invariant, and so unique names have a stable character and are non-indexical. In other words, the present proposal recapitulates the classical Millian semantics of proper names as constant expressions whose semantic contents are exhausted by their referents.

    But there are two important differences between the present proposal and classical Millianism. First, the Millian semantics applies only to unique names, and not to shared or ephemeral names, both of which are context-sensitive. The traditional Millian semantics is therefore correct so far as it goes, but is overly narrow, failing to recognize the semantic diversity of proper names. The second is that in spite of the fact that unique names are *in effect* constant expressions, they are *formally* still variable expressions, just variable expressions of a certain special kind. This constitutes a kind of reduction of constants to variables: the former are 'frozen' instances of the latter.

    The semantic entry for a unique name makes reference to some specific individual in the domain: it only allows for indexation with indices that map only to said individual. This recapitulates the fact that in order to be semantically competent with a unique name, one must know the individual to which it refers, unlike with shared names. To know the meaning of a unique name is essentially to know its referent, and this referent is encoded in the name's semantics. As stated in section 2 above, and as proposed by traditional referentialist accounts of proper names, the establishment of a name-bearing relation between a unique name and an individual coincides with the coinage of that name.

3.3 Ephemeral names

Ephemeral names can refer to an individual only if that individual bears some property in w, the world of evaluation relative to which the name is interpreted. In order to capture this, I introduce a new sort of restriction on the kind of index that a proper name allows: in order to be interpreted, an ephemeral name must be tagged with an index that always maps only to individuals that bear a certain property p in the world of evaluation w. The semantic entry for an ephemeral name is as follows, where a property p is a function from indices of evaluation to individuals to truth values.

(10) **Semantics of an ephemeral name n limited by property p indexed with i**

    (a) $[\![n_i]\!]^c = \lambda s : \forall g : i \in Dom(g)[p(s)(g(i)) = true].g_s(i)$

    (b) $[\![n_i]\!]^{c,w,g} = g(i)$, if for all g such that $g(i)$ is defined, $g(i)$ bears p in w; else undefined

    (c) $[\![Captain\ Obvious_1]\!]^{c,w,g} = g(1)$, if for all g such that $g(1)$ is defined, $g(1)$ points out the obvious w; else undefined

The only indices that an ephemeral name permits, relative to a world of evaluation w, are those that always map to an individual that bears a certain property in w so long as they map to anything at all. Like shared names, ephemeral names are thus context-dependent, but they are so in two ways: w is dependent on the context via c_w to fix the set of individuals to which the ephemeral name is capable of referring in that context, and g is provided by the context to determine which of these individuals is in fact being referred to. Ephemeral names are rigidly designating, but their sensitivity to w means that modal contexts can have an impact on which single referent they allow.[12] Ephemeral names do not 'stick' to any individuals in particular, but change their possibilities of reference as the properties that individuals bear change. In addition to their penchant for nonce-uses, this is what the 'ephemerality' of ephemeral names consists in, and as will be shown in the following section, it yields the result that ephemeral names, despite being grammatically proper names, nonetheless do not take part in name-bearing relations.

Semantic competence with an ephemeral name does not require knowing any particular individual to which it can refer, but only knowing which property constrains its indexation. The further task of knowing to which individuals it can refer is then accomplished not by knowledge of $\mathcal{G}$, as with shared names, but rather with knowledge of which individuals bear the relevant property in w.

3.4 Name-bearing

I have provided an account of the semantics of the three kinds of proper names introduced in section 2, along with an explanation of how these semantics recapitulate the empirical observations made there. As promised, these semantics allow for a unified model-theoretic notion of name-bearing.

Let the name-bearing relation $\mathcal{NR}$ be a set of ordered pairs $\langle n, d \rangle$, where n is a proper name and d is an individual in the domain — d then bears n just in case $\langle n, d \rangle \in \mathcal{NR}$. Name-bearing can then be characterized by how $\mathcal{NR}$ is defined, in the following way.[13]

---

[12]To see how ephemeral names can be world-sensitive, yet rigid, consider a sentence like, 'Julie thinks she'll meet Mr. Right.' The ephemeral name has both a shifted and an unshifted interpretation: 'Mr. Right' might refer, within the scope of the belief operator, to some one man who must be marriageable in each of Julie's doxastic alternatives, or, outside the scope of the operator, to some one man who is marriageable in the actual world, whom Julie independently believes she will meet.

[13]As defined here, the relation $\mathcal{NR}$ is determined independently of any possible world, and thus name-bearing relations are 'necessary' from the perspective of the language. What this means is that where counterfactuals

(11) **Definition of name-bearing**
For any proper name n and individual d, $\langle n, d \rangle \in \mathcal{NR}$ iff:

(a) there is a set of indices $\mathcal{I} \subset \mathcal{V}$ such that for all $i \in \mathcal{V}$, $i \in \mathcal{I}$ iff for all c, w and g for which $g(i)$ is defined, $\llbracket n_i \rrbracket^{c,w,g}$ is defined, and:

(b) for all $i \in \mathcal{I}$, there is some $g \in \mathcal{G}$ such that $g(i) = d$.

What this definition says is that an individual bears a proper name just in case (a) there is some set of indices consisting of all and only those that the proper name always allows, and (b) for each of these indices, there is some assignment available in the language that maps that index to the individual. In other words, an individual bears a proper name just in case that individual is always a potential referent of that name, depending on the selection of g. This the formal rendition of the inverted dictum: his name is 'Socrates' because that's what he's called.[14]

A consequence of the definition provided here is that ephemeral names do not participate in name-bearing relations. This is because in the case of any ephemeral name n, there are no indices i such that for all c, w and g such that $g(i)$ is defined, $\llbracket n_i \rrbracket^{c,w,g}$ is defined. The reason for this is that which indices are permitted by an ephemeral name is sensitive to w. For any ephemeral name n and index i, one can therefore find a w relative to which, for some c and g, $\llbracket n_i \rrbracket^{c,w,g}$ is undefined: simply select a w in which $g(i)$ does not bear p.[15] The point of name-bearing is that all of a proper name's always-allowed indices have some assignment that maps them to a certain individual; but the 'ephemerality' of ephemeral names guarantees precisely that there are no such always-allowed indices, and therefore that individuals do not bear ephemeral names in the way they bear shared and unique names.

I have shown the semantic variety of proper names, and demonstrated that their variety is the result of the types of semantic restrictions on their own indexation that they allow. But each of these different kinds of expression deserves to be collected under the common umbrella term, 'proper name,' because all of them place semantic restrictions on their indexation. Proper names can therefore be defined as a semantic class as follows.

(12) **Semantic definition of a proper name**
A proper name is an assignment-sensitive referential expression whose extension relative to some c, w, and g is defined only if the expression is syntactically indexed with a member of some proper subset of the total set of referential indices.

---

and so on are considered involving alternate name-bearing relations, these are counterfactual situations in which the *language* is different, and not the properties of the individual apart from the language — this reflects that name-bearing is a linguistically determined relation. From the language's perspective, one cannot find a possible world in which John does not bear the name 'John' for its own purposes of reference, but one can find a possible world in which English is such that, from the perspective of the language *as used in that possible world*, he does not.

[14]I have posited that in order for an individual to bear a name, for *every* index that a name always allows, there has to be some assignment mapping that index to that individual. The reasoning behind this is that it is a matter of indifference which specific index the proper name is tagged with, so long as it is one that the proper name always permits. It would be very strange, and formally there would be no sense in supposing, that an individual was named 'John,' in spite of the fact that some oddball John-index, e.g. $876 \in \mathcal{V}_{John}$, was incapable of mapping to that individual. The point is that the name itself, regardless of which specific index that it always allows is chosen, can refer to an individual. Note also that the definition of name-bearing provided here overgenerates in the sense that it allows for many formally definable name-bearing situations that natural languages seem not to make use of. This just means that natural language generally makes use only of a proper subset of possible name-bearing scenarios: and though I don't have space to talk about it here, the types of name-bearing exploited in natural languages can be characterized using a small set of simple possible operations on $\mathcal{G}$.

[15]Assuming that p is not metaphysically necessary, which I take to be true for all ephemeral names.

4 Conclusion

The foregoing approach provides a precise model-theoretic definition of name-bearing that reflects the idea that name-bearing is dependent on reference using a proper name. It opens a number of issues regarding the syntax and semantics of proper names that cannot be addressed here, including the relation between proper names and pronouns, proper names' status as R-expressions and their interaction with bound variable interpretations, and the sociolinguistic role of naming conventions. But I hope to have shown that name-bearing can be given a rigorous and intuitively plausible formal characterization.

References

[1] Kent Bach. What's in a name. *Australasian Journal of Philosophy*, 59(4):371–386, 1981.

[2] Tyler Burge. Reference and proper names. *Journal of Philosophy*, 70(14):425–439, 1973.

[3] Samuel Cumming. Variabilism. *Philosophical Review*, 117(4):525–554, 2008.

[4] Gareth Evans. The causal theory of names. *Aristotelian Society Supplementary Volume*, 41:187–208, 1973.

[5] Delia Graff Fara. Names are predicates. *Philosophical Review*, 124(1):59–117, 2015.

[6] Bart Geurts. Good news about the description theory of names. *Journal of Semantics*, 14(4):319–348, 1997.

[7] Aidan Gray. Name-bearing, reference, and circularity. *Philosophical Studies*, 171(2):207–231, 2014.

[8] Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell Publishing, Oxford, UK, 1998.

[9] Robin Jeshion. Referentialism and predicativism about proper names. *Erkenntnis*, 80:363–404, 2015.

[10] David Kaplan. Demonstratives. In John Perry, Joseph Almog, and Howard Wettstein, editors, *Themes from Kaplan*, pages 481–563. Oxford University Press, 1989.

[11] David Kaplan. Words. *Aristotelian Society Supplementary Volume*, 64:93–119, 1990.

[12] William Kneale. Modality *de re* and *de dicto*. In Ernst Nagel, Pat Suppes, and Alfred Tarski, editors, *Logic, Methodology, and Philosophy of Science*, pages 622–633. Stanford University Press, 1962.

[13] Saul Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, MA, 1980.

[14] Brian Loar. The semantics of singular terms. *Philosophical Studies*, 30(6):353–377, 1976.

[15] Ora Matushansky. On the linguistic complexity of proper names. *Linguistics and Philosophy*, 31(5):573–627, 2008.

[16] Anna Mindess. What name signs can tell us about Deaf culture. *Sign Language Studies*, 66(1):1–23, 1990.

[17] Michael Pelczar and Joseph Rainsbury. The indexical character of names. *Synthese*, 114(2):293–317, 1998.

[18] Dolf Rami. The use-conditional indexical conception of proper names. *Philosophical Studies*, 168(1):119–150, 2014.

[19] Clarence Sloat. Proper nouns in English. *Language*, 45(1):26–30, 1969.

[20] Heidi Tiedke. Proper names and their fictional uses. *Australasian Journal of Philosophy*, 89(4):707–726, 2011.

Resolving Quantity- and Informativeness-Implicature in Indefinite Reference

Till Poppels and Roger Levy [*]

Department of Linguistics
University of California, San Diego
{tpoppels,rlevy}@ucsd.edu

Abstract

A central challenge for all theories of conversational implicature (Grice, 1957, 1975) is characterizing the fundamental tension between Quantity (Q) implicature, in which utterance meaning is refined through exclusion of the meanings of alternative utterances, and Informativeness (I) implicature, in which utterance meaning is refined by strengthening to the prototypical case (Atlas & Levinson, 1981; Levinson, 2000). Here we report a large-scale experimental investigation of Q-I resolution in cases of semantically underspecified indefinite reference. We found strong support for five predictions, strengthening the case for recent rational speaker models of conversational implicature (Frank & Goodman, 2012; Degen, Franke, & Jäger, 2013): interpretational preferences were affected by (i) subjective prior probabilities (Informativeness), (ii) the polarity and (iii) the magnitude of utterance cost differentials (Quantity), (iv) the felicity conditions of indefinite NPs in English, and (v) the 'relatability' of X and Y.

1 Introduction

1.1 The phenomenon

In transitive sentences of the form 'The X V-ed a Y' the relationship between X and Y remains semantically unspecified and must be inferred (Horn, 1984):

(1) a. The man injured a child. $+>$[1] not his own child
     b. The man broke a finger. $+>$ his own finger

In these examples the indefinite reference to Y is ambiguous between a relational reading, where Y is X's own (as in 1b; we will be referring to this as the 'OWN interpretation'), and a non-relational reading, where Y does not belong to X (as in 1a; henceforth referred to as 'OTHER'S'). In the way this OWN/OTHER'S ambiguity is resolved, (1) illustrates the tension between two major constraints on language production that have been recognized as driving extrasemantic inferences since Grice (1957): Quantity and Informativeness. In Grice's original proposal, these two constraints were captured in the Maxim of Quantity, which instructs speakers to be brief, while also being informative. Subsequent efforts to reduce redundancy in the Gricean

---

[1]'$+>$' is short for 'implicates'.

Maxims, most notably by Levinson (2000) and Horn (1984), have singled out Quantity and Informativenss as key sources of Gricean inferences.[2]

The effect of Quantity can be witnessed in 1a, which is completely analogous to standard cases of Quantity (Q-) implicature: The OTHER'S interpretation of *The man injured **a** child* is Q-implicated through exclusion of the meaning of a similarly brief, yet unambiguous and thus more informative, alternative utterance for conveying OWN: *The man injured **his** child*. By the same reasoning we would expect an analogous Q-implicature from 1b, yet the opposite inference appears to arise. This has been argued to reflect the effect of Informativeness (I-) implicature (Atlas & Levinson, 1981), i.e. the strengthing of utterance meaning to the "stereotypical meaning, use, or situation" (Horn, 2004, p. 16).

Thus, the puzzle in 1 illustrates the fact that the tension between Quantity and Informativeness is resolved differently across different utterances. The present paper addresses *how* and *when* this Q/I tension is resolved in favor of Q (as in 1a) or I (as in 1b).

1.2   The Rational Speech Act model

At the center of the Gricean program is the Cooperative Principle, which licenses the assumption (among others) that the speaker's utterance of choice aims to strike an optimal balance between being maximally informative (I) and efficient (Q) compared to its alternatives. This core idea has been formalized in recent Bayesian (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013) and game-theoretic (Degen et al., 2013) models, in which interlocutors maintain probabilistic beliefs about each other's knowledge and communicative goals, and rely on these beliefs to reason iteratively about each other's choices. Here, we focus on the Rational Speech-Act (RSA) model, which models the interpretational preferences of a pragmatic listener L_1, whose speaker model (S_1) is based on a hypothetical literal listener (L_0) who interprets utterances (u) according to their literal semantics and a prior distribution over meanings (m):[3]

$$P_{L1}(m|u) \propto P_{S1}(u|m)P(m) \tag{i}$$

$$P_{S1}(u|m) \propto \exp(\lambda[\log(P_{L_0}(m|u)) - D(u)]) \tag{ii}$$

$$P_{L_0}(m|u) \propto \mathcal{L}(u,m)P(m) \tag{iii}$$

This iterative reasoning process grounds out in the mutually known 'lexicon' $\mathcal{L}$, which maps utterances to meanings that are consistent with it:

$$\mathcal{L}(u,m) = \begin{cases} 1 & \text{if } m \in [\![u]\!] \\ 0 & \text{otherwise} \end{cases} \tag{iv}$$

While L_0 and L_1 reason about potential meanings of a given utterance, S_1 models the tension between being optimally brief and informative: this 'Gricean speaker' chooses (softmax with

---

[2]Note that Horn refers to the Informativeness constraint as R, and that Levinson's taxonomy includes an additional component, M, which roughly maps onto Grice's Maxim of Manner. For present purposes, we restrict our discussion to Q and I, however, since it is the resolution of these two interpretational forces that we are looking to explain.

[3]This formalization goes back to Frank and Goodman (2012) and Goodman and Stuhlmüller (2013), although in the original proposal the literal listener does not use the prior distribution, so that $P_{L0}(m|u) \propto \mathcal{L}(u,m)$. In line with subsequent extensions of the model (Degen, Tessler, & Goodman, 2015; Lassiter & Goodman, 2015), we do attribute knowledge of the prior to L_0, in order to enable S_1 to capitalize on shared world knowledge when choosing the optimal utterance for conveying her meaning.

hardness parameter λ) between alternative utterances for expressing her intended meaning by weighing the cost $D(u)$ of each utterance u against the surprisal that her intended meaning m would have in the posterior distribution $P_{L_0}(m|u)$ of her listener model L_0 (Eq. ii). This is where alternative utterances exert what we will call their 'scalar pressure': since speakers generally choose utterances that are maximally informative about their intended meaning, every utterance that is *not* chosen by the speaker exerts some amount of interpretational pressure away from the meaning it encodes. Crucially, however, S_1 is also constrained by costs, which we assume are monotonically decreasing in utterance brevity. Thus, L_1 can explain away the speaker's choice not to use a particular alternative utterance if that utterance is more prolix than her utterance of choice. This means that when an utterance u is compatible with a meaning m for which a more precise utterance u' is also available, the less costly u' is the more scalar pressure it exerts against u being interpreted as m. This notion of 'scalar pressure' is central to the present study and we will appeal to it repeatedly in the remainder of the paper.

In order to apply the RSA model to the utterances such as 1, we assume a minimal lexicon that contains the (ambiguous) utterance in question along with its unambiguous alternatives— schematically represented as $\{a;\ self's;\ someone\ else's\}^4$—and a set of two possible meanings: OWN and OTHER'S. We further assume a cost function, which assigns production costs to utterances based on the number of syllables they comprise: $D(a) = D(self's) = 1$ and $D(someone\ else's) = 4.^5$ Table 1 summarizes these assumptions. Figure 1 graphically depicts the lexical meaning relationship among alternative utterances; the fainter depiction of *someone else's* reflects its higher cost, suggesting that it should exert less scalar pressure on a toward an OWN interpretation than *self's* does toward an OTHER'S interpretation. Throughout the present paper, we further assume that $\lambda = 1$, recovering a Luce choice rule (Luce, 1959), which has used to model human decisions across a diverse set of domains (Sutton & Barto, 1998). Greater values for λ produce more polarized production preferences, although they are qualitatively robust across different parameter settings.

utterance (u)	$\mathcal{L}(u,\ \text{OWN})$	$\mathcal{L}(u,\ \text{OTHER'S})$	$D(u)$
a	1	1	1
$self's$	1	0	1
$someone\ else's$	0	1	4

Table 1: Assumed 'lexicon' and utterance costs

Figure 1: Graphical depiction

With these assumptions in place, Figure 2 shows the model's posteriors at successive stages of the iterative reasoning process as a function of the prior distribution over meanings $P(m)$. Consider first the literal listener (left panel). Owing to the 'lexicon' just described, L_0's probability of OTHER'S is 0 given *self's* and 1 given *someone else's*, whereas the interpretation of the ambiguous utterance (solid line graph) is driven entirely by the prior, since it is (by definition) compatible with either meaning. Consequently, *self's* cannot be used to convey OTHER'S to L_0 (therefore, $P_{S1}(self's|\text{OTHER'S}) = 0$), leaving a to compete with *someone else's*, since both of these alternatives are literally compatible with S_1's intended meaning (top panel). In choosing between these two utterances, S_1 is drawn towards *someone else's* because it is guaranteed to be interpreted correctly (by L_0), and therefore maximally informative. However, the relative

---

[4] Throughout the paper we use the umbrella term '*self's*' to refer to the forms that convey OWN unambiguously, which include *his, her, its, our, your, their*, and *my*, all of which are monosyllabic.

[5] The monotonicity, not the exact numbers, of this cost profile is crucial for the qualitative predictions we address here. In principle, the psychological cost of an utterance is likely not reducible to its length, and exactly what determines production costs remains an open question. We will not address this issue here, although we do propose an extension to this cost function in Section 1.3.

brevity of a is attractive as well, and its informativeness increases with the prior probability of OTHER'S. Thus, S_1 will bother to use the unambiguous utterance to convey OTHER'S only when the prior probability of that meaning is below a certain threshold (in the present example the cross-over point is at $P(\text{OTHER'S}) = 0.05$. Crucially, this contrasts with the decision facing a speaker who wants to convey OWN, since in that case both literally compatible utterances are equally inexpensive ($D(a) = D(his) = 1$). In that scenario (bottom panel), the speaker will be drawn to the unambiguous utterance unless, of course, the $P(\text{OTHER'S}) = 0$, in which case the prior rules out OTHER'S, and a and $self$'s are equally informative, so that $P_{S1}(a|\text{OWN}) = P_{S1}(self$'s$|\text{OWN}) = .5$.
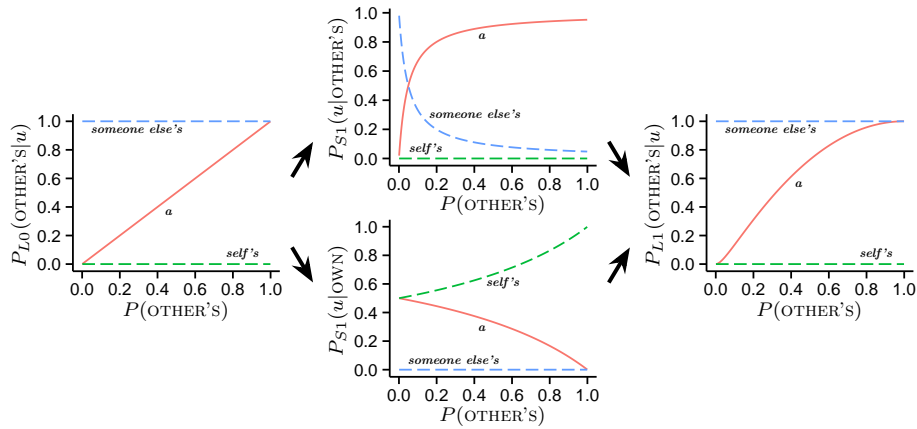
Figure 2: Posterior distribution of the literal listener (left), the 'Gricean speaker' (center), and the pragmatic listener (right). The speaker's intention is to convey OTHER'S in the top panel, and OWN in the bottom one. The solid line represents the ambiguous utterance.

The pragmatic listener (right panel) then attempts to infer S_1's intention by reverse-engineering the production process, and integrating S_1's posterior with prior probabilities. As a result, her interpretation of the ambiguous utterance is skewed towards OTHER'S relative to the prior. However, since the scalar pressure driving this Q-implicature is proportional to the Informativeness benefit of using $self$'s instead of a, the effect of Q can be overturned by I if the prior tends strongly towards OWN.

Note the RSA model closely captures traditional definitions of Quantity implicature as alternative-based inferences and Informativeness implicature as the strengthening of utterances to the stereotypical meaning. Crucially, however, it goes beyond these definitions by providing the necessary machinery for explaining the trade-off between these two interpretational forces in a principled way, by casting it as the prior-likelihood trade-off in Bayesian inference. As a result, the model is capable of generating precise predictions about Q/I resolution in sentences like 1, which are desribed in the next section.

1.3   Predictions

We conducted a large-scale forced-choice experiment to investigate interpretational preferences in sentences like (1). Using mixed logit regression, we test five predictions about the way the Q/I tension is resolved in these sentences. Three of these predictions follow directly from the RSA model and are illustrated in Figure 3. Two additional predictions are motivated independently and explore potential limitations of the RSA model in accounting for Informativeness-driven inferences.test
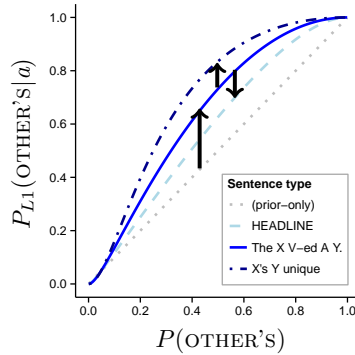
Figure 3: Predictions
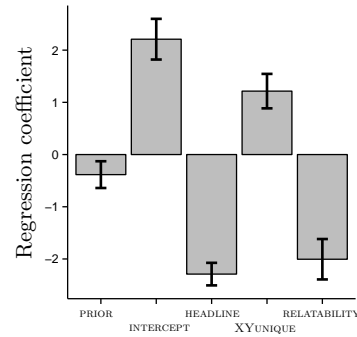
Figure 4: Regression coefficients and Standard Errors

Predictions 1 and 2: Event priors and baseline Q-implicature.   Since rational-speaker models generally assume that speakers will capitalize on their common ground with their addressee, mutual knowledge about the world represents the starting point for pragmatic reasoning according to these models. Within the RSA model, interpretations are partly determined by, and should therefore track, prior probabilities, which corresponds to the general monotonicity of the prior–posterior relationship in all lines of Figure 3. But relative to the prior, interpretations should be skewed towards OTHER'S due to the scalar pressure from *self's*: in Figure 3, all model predictions are above the $x = y$ line.

Prediction 3: Reduced effect of Q in headline cases.   Manipulating the utterance cost profile should modulate the effect of Q. In particular, if we reduce the cost of the ambiguous utterance then it should reduce overall scalar pressure and bring the L_1 posterior closer to the prior. Consider the following contrast:

(2)   a. The man injured a child.
       b. Man injured child.

The HEADLINE-like style of 2b affords an ambiguous utterance with no explicit determiner—we denote this option with $\emptyset$. If utterance costs are monotonic in length, then $\emptyset$ is cheaper than either unambiguous alternative: $D(\emptyset) < D(\textit{self's}) < D(\textit{someone else's})$. As illustrated in Figure 3, RSA predicts for this modified cost profile an overall weaker effect of Q-implicature in headline-like sentences. Since the original effect of Q was a skew toward OTHER'S interpretations, headline-like sentences should be OWN-skewed relative to their full-sentence counterparts.

Prediction 4: Non-uniqueness in indefinite reference. Hawkins (1991) argues that the use of the indefinite determiner in English is infelicitous whenever its reference is uniquely satisfied within the domain:

(3)   a.   # a brightest student
       b.   # a president of the United States

This generalization makes an interesting prediction about the sentences we are considering ('The X V-ed a Y'): in cases where the semantics of X and Y are such that a typical X has only one Y, the non-uniqueness requirement of the indefinite is violated on an OWN interpretation, rendering the indefinite reference infelicitous given that OWN is intended. If this felicity condition is mutually known and constrains the speaker's utterance choice, we would expect 4b to be less likely to receive an OWN interpretation than 4a:

(4)   a.   The man broke a finger.
       b.   The man broke a nose.

This prediction can be accommodated within the RSA model under the auxiliary assumption that D be made a 2-place function of u, m pairs (Jäger, 2012), such that not only utterance brevity but also 'felicity' of specific form–meaning pairs contribue to costs. Let the value of the new $D(u, m)$ be the value of $D(u)$ from Table 1 in all cases except for $u = a$ and $m = $ OWN in contexts where X and Y are such that a typical X possesses only a single Y (as in 4b), in which case $D(a, own) > D(self's, own)$. In these contexts, S_1 will be disincentivized from using the ambiguous utterance when aiming for an OWN interpretation, and correspondingly shifts some probability mass from OWN to OTHER'S in L_1's posterior.

Prediction 5: The real-world 'relatability' of event participants. The materials used in the present study include event descriptions that vary with respect to the 'relatability' of the event participants, exemplified by the following contrast, where *father* and *son* are intuitively more relatable than *man* and *child*:

(5)   a.   The man injured a child.
       b.   The father injured a son.

Precisely characterizing relatability is beyond the scope of this paper; we informally take it as an index of the degree to which the description of one NP brings into mind another referent with which the other NP could be identified. Relatability seems closely related to bridging (Clark, 1975), as in the following example, due to Kehler (in prep):

(6)   I almost bought a car today, but...
       a.   ... *the engine* was too noisy.
       b.   # ... *the TV* was blurry.
       c.   ## ... *the stapler* was broken.

The relationship between the car and the entity denoted by the italicized definite NP is semantically unspecified and must be inferred. That inference appears more natural, however, when the entity in question is highly relatable to the car (cf. *engine* in 6a) than when the two are less relatable (as in 6b and 6c). By analogy to such bridging inferences, one may expect that sentences with highly relatable NPs, such as (5b), tend more strongly towards an OWN interpretation, since on that reading the two entities *are* related. It is important to note, however, that the RSA model has no mechanism for predicting such effects at present. We will return to this issue in the Discussion.

2   Methods

Participants.   1885 native speakers of English were recruited via Amazon.com's Mechanical Turk to complete a single-trial forced-choice experiment. A UniqueTurker script[6] ensured that each participant could complete the survey only once.

Materials.   53 sentences of the form *The X V-ed a Y* were designed to vary widely with respect to prior OWN/OTHER'S probability, the relatability of X and Y, as well as the number of Ys a typical X possesses. Each of those X-V-Y sentences was matched with a corresponding newspaper headline-like version of the form *X V-ed Y*.

Procedure.   Participants were presented with a single-trial forced-choice task, which was implemented using the online survey platform Qualtrics. The answer choices identified OWN and OTHER'S readings of the sentences and were presented in a random order.

Prior norming.   The prior OWN/OTHER'S probability of each of the 53 items was estimated in a separate norming experiment. We recruited 885 participants from Mechanical Turk who were presented with five questions of the form 'How likely is an X to V his own Y compared to V-ing someone else's Y?', each corresponding to one of the 53 events described in the main experiment. Using a slider, participants chose between '100% likely to V his/her/its own Y' and '100% likely to V someone else's Y', describing OWN and OTHER'S events, respectively.

3   Results

The interpretational preferences measured in the main experiment ranged from 100% OWN to 100% OTHER'S, and estimated prior probabilities from 80.49% OWN to 88.59% OTHER'S. A logistic regression analysis with mixed effects tested whether interpretations varied as a function of prior probabilities, whether the sentence was presented as a full sentence or a headline-like version (HEADLINE), whether or not a typical X had only one Y (XYUNIQUE), and the RELATABILITY of X and Y. Values for XYUNIQUE and RELATABILITY were hand-coded by the first author. Prior estimates p from the norming experiment were scaled to range from 0 to 1, centered around 0.5, and logit-transformed (PRIOR $= \mathrm{logit}\left[\frac{0.8 \times p + 10}{100}\right]$). All other factors were treatment-coded categorical predictors and HEADLINE was added as a random by-item effect.[7] The results are overall consistent with our predictions (Fig. 4): The prior had a numerical effect in the predicted direction, although this trend did not reach significance ($\beta = -.38, p = .13$). The significant INTERCEPT indicates the predicted OTHER'S-skew ($\beta = 2.21, p < .001$), which was enhanced where X's Y was unique ($\beta = 1.22, p < .001$). HEADLINEs were significantly more likely to receive OWN interpretations than their full-sentence counterparts ($\beta = -2.29, p < .001$), and so were sentences with highly relatable nouns compared to those with less relatable nouns ($\beta = -2.01, p < .001$).

---

[6]uniqueturker.myleott.com

[7]The   full   formula   was   `response ∼ prior + XYunique + relatability + headline + (1 + headline|item)`.

4   Discussion

Support for the RSA model.   The overall OTHER'S-skew (as indicated by the significant negative intercept) reflects the predicted baseline effect of Q-implicature: because *self's* is lower-cost than *someone else's*, it exerts more scalar pressure on the interpretation of the ambiguous a, pushing the latter towards an OTHER'S interpretation. Furthermore, the within-item HEADLINE manipulation confirms RSA's prediction that by lowering the cost of the ambiguous utterance, scalar pressure on it and thus the OTHER'S-skewing Q effect would be ameliorated. Together, these results provide strong support for the RSA model.

Event priors.   Since we expected interpretations to track prior probabilities, it is worth speculating why the effect of the prior did not reach significance. Notice that on a strict interpretation of the RSA model as Bayesian inference, with the posterior distribution defined over communicative intentions, the corresponding priors should be defined over intentions as well. Our norming experiment, on the other hand, measured people's expectations about events, not event descriptions. While the probability of an event may be correlated with its probability of being mentioned, that correlation is likely to be noisy: extremely unlikely events may be more remarkable and therefore more likely to be talked about than highly predictable events. If this reasoning is correct, our prior estimates may not have the correct "currency," which may explain why they did not affect interpretations significantly. This contrast between event priors and 'intention priors' is mirrored in the definition traditionally given to I-implicature, as the strengthening of utterance meaning to the "stereotypical *use* (...) or *situation*" (Horn, 2004, p. 16; emphasis added).

Encoding felicity conditions as form-meaning costs.   The prediction that the interpretation of *The X V-ed a Y* would tend towards OTHER'S in cases where X has only one Y is supported by our data. This prediction was based on Hawkins (1991)'s observation that indefinite reference in English tends to be infelicitous if it is known to be uniquely satisfied. To derive this prediction from within the RSA architecture, we encoded the non-uniqueness requirement as a soft constraint on production preferences through an additional cost function that assigned an extra 'felicity cost' whenever the speaker's intention was to elicit an OWN interpretation and X's Y was unique. Encoding felicity conditions as form-meaning costs captures the intuition that these constraints are part of interlocutors' mutual knowledge about the language they use to communicate.

The effect of relatability: Q or I?   One may be tempted to explain the relatability effect by invoking *ad-hoc* scalar (Q-) implicature (Hirschberg, 1985) about referring expressions. On that view, highly relatable entities drive interpretations towards OWN because the speaker's use of referring expressions that are more specific than some of its scale mates (e.g. *father* and *son* as opposed to *man* and *child*) triggers a search for the purpose of that choice. Crucially, however, this explanation requires the auxiliary assumption that more specific referring expressions are in general more costly than their less specific competitors, which would have to be motivated independently in order to avoid circularity.[8]

   Instead, the prediction that relatability would drive interpretations towards OWN was motivated by analogy to bridging inferences (Clark, 1975), in which reference resolution is biased towards entities that are associated with previously mentioned entities (recall that 'the engine'

---

[8]In principle, such production costs could correspond to 'lexical retrieval effort' if, for example, basic-level category labels, such as *man*, are easier to retrieve from memory than alternatives like *father*.

in 6a is strengthened to 'the engine of that car'). According to Prince and Cole (1981), such bridging inferences arise when entities associated with a previously mentioned entity are more 'inferrable' than their competitors, such as the engine of the previously mentioned car compared to other engines in the world. If this analogy holds, the relatability effect in our data reflects an Informativeness implicature that is based on the interlocutors' mutual real-world knowledge about the relatability of event participants.

This raises an interesting point about the nature of Informativeness-driven inferences: since I-implicatures are based on shared real-world knowledge, their *content* may be determined by entirely non-intentionalist inference mechanisms for making sense of the world, rather than iterative reasoning between interlocutors. Consider in this context the following example, due to Cohen and Kehler (submitted):

(7)   The manager fired the employee who came in late every day last week.

Cohen and Kehler argue that the 'causal attribution' inference that this utterance invites— that the employee was fired *because* of being tardy—would arise in much the same way if the situation was perceived directly, rather than described linguistically: if you saw an employee who you know to be tardy getting fired, you may come to the same conclusion, although in that case there is no communicative intent to be inferred. Crucially, however, if the relevant information *is* conveyed intentionally (e.g. by uttering 7), the listener may reason that the speaker intended to elicit that inference, which Cohen and Kehler therefore call a 'conversational elicit-ure'. In that case, elicitures show a 'focal point effect' (Schelling, 1980), by which the inference is strengthened iteratively, although its content is fixed by ego-centric, non-intentionalist inference mechanisms. With respect to our X-V-Y, that would mean that the listener is drawn towards OWN when X and Y are highly relatable, and that tendency is strengthened by the assumption that the speaker must have anticipated that inference and wanted it to be drawn. Levinson appears to consider the possibility that Informativeness-based inferences may have non-intentionalist components when he likens them to 'inferences to the best explanation' in scientific contexts (cf. Thagard, 2000, 2007). Future work will be needed to further explore the role that semantic relatability, specifically, and non-intentionalist inferences, in general, play in language comprehension. As a first step towards that goal, future studies should aim to find reliable and fine-grained ways of quantifying relatability, such as corpus-based co-occurrence measures.

5   Conclusion

We set out to investigate the resolution of Quantity and Informativeness implicature in seman-tically underspecified cases of indefinite reference. We have demonstrated that the resolution of these antinomic interpretational forces involves several interacting factors, many of which follow straightforwardly from rational-speaker accounts of interpretation. Knowledge from var-ious domains, such as semantic, pragmatic, social, and real-world knowledge, plays a role in the inferential mechanism that drives the interpretation of utterances. We have argued that developing predictive theories that explain exactly how interpretational inferences are drawn is an important theoretical goal and suggested that a complete account may ultimately require a non-intentionalist component. Finding a place for such a component (if it exists) within rational-speaker accounts may extend their explanatory reach and further our understanding of the cognitive machinery behind language comprehension.

References

Atlas, J. D. & Levinson, S. C. (1981). *It*-clefts, informativeness, and logical form: radical pragmatics (revised standard version). *Radical pragmatics*, *1*, 1–61.

Clark, H. (1975). Bridging. In *Proceedings of the 1975 workshop on theoretical issues in natural language processing*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Cohen, J. & Kehler, A. (submitted). Conversational Elicitures.

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the annual meeting of the cognitive science society* (pp. 376–381).

Degen, J., Tessler, M. H., & Goodman, N. D. (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd. In *Preceedings of the 37th annual conference of the cognitive science society*.

Frank, M. C. & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998.

Goodman, N. D. & Stuhlmüller, A. (2013). Knowledge and implicature: modeling language understanding as social cognition. *5*(1), 173–184.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Grice, H. P. (1975). Logic and conversation. In P. Cole & G. Morgan (Eds.), *Syntax and semantics III* (pp. 41–58). London: Academic Press.

Hawkins, J. A. (1991). On (in)definite articles: implicatures and (un) grammaticality prediction. *Journal of Linguistics*, *27*(2), 405–442.

Hirschberg, J. (1985). A Theory of Scalar Implicature (Natural Languages, Pragmatics, Inference). *Dissertations available from ProQuest*. Retrieved from http://repository.upenn.edu/dissertations/AAI8603648

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: linguistic applications* (pp. 11–42). Washington: Georgetown University Press.

Horn, L. R. (2004). Implicature. In L. R. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 3–28). Blackwell Publishing Ltd.

Jäger, G. (2012). Game theory in semantics and pragmatics. In C. Maienborn, P. Portner, & K. von Heusinger (Eds.), *Sematics: an international handbook of natural language meaning* (Vol. 3, pp. 2487–2516). Berlin: de Gruyter.

Kehler, A. (in prep). Reference in Discourse. In S. Lappin & C. Fox (Eds.), *Handbook of contemporary semantics - second edition* (pp. 851–890). Wiley-Blackwell.

Lassiter, D. & Goodman, N. D. (2015). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*. Retrieved from http://web.stanford.edu/%7B~%7Ddanlass/Lassiter-Goodman-adjectival-vagueness-draft.pdf

Levinson, S. C. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. MIT Press.

Luce, R. D. (1959). Individual Choice Behavior a Theoretical Analysis.

Prince, E. F. & Cole, P. (1981). Toward a taxonomy of given-new information. (pp. 223–255).

Schelling, T. C. (1980). *The Strategy of Conflict*. Harvard University Press.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement learning*. MIT Press.

Thagard, P. (2000, November). *Coherence in thought and action*. MIT Press.

Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of Science*, *74*(1), 28–47.

On Reporting Attitudes: an Analysis of Desire Reports and their Reading-Establishing Scenarios[*]

Tillmann Pross

University of Stuttgart, Germany
prosstn@ims.uni-stuttgart.de

Abstract

I argue that conventional wisdom about attitude reports lumps together theorizing about the attitudes themselves with theorizing about the meaning of attitude reports. Instead of conflating these, I propose to treat them as distinct (although obviously related) topics. A central reason for this separation is the distinction between anaphoric constructions that are *internally* about the same object and anaphoric constructions that are *externally* about the same object. Anaphoric constructions that are internally about the same object reveal the structure of the attitudes themselves – e.g. that desires are parasitic on beliefs – whereas anaphoric constructions which are externally about the same object make explicit the structure of attitude reports. I correlate the distinction between the attitudes themselves and attitude reports with the asymmetry between the first- and third-person perspective on attitudes. On the basis of this asymmetry, I propose a semantics of desire reports according to which a desire report like *Adrian wants to buy a jacket like Malte's* is not ambiguous between different readings. Conventional wisdom readings are reconstructed as a third person's pragmatic strategies for justifying the truth of an attitude ascription. I conclude with an outlook on the transfer of the analysis proposed to reports involving other non-doxastic predicates like *wish* or *hope* and reports involving doxastic predicates such as *believe*.

1 Three concerns about the conventional wisdom analysis of attitude reports

Conventional wisdom has it that an attitude report like (1) is ambiguous between a range of so-called *readings* of (1) that appear when (1) is interpreted against a certain set of *background assumptions*.

(1)   (A reporter says:) Adrian wants to buy a jacket like Malte's.

Background assumptions – often called *scenarios* – for the identification of readings usually comprise (a) a doxastic specification of the descriptive content of the report, e.g. whether or not Adran knows Malte or his jacket and (b) a specification of the reality in which Adrian finds himself, e.g. whether or not Adrian stands in a causal relation of acquaintance to a specific jacket like Malte's. For example, a scenario in which Adrian has decided to buy a certain jacket but has no idea that the jacket he wants to buy is like Malte's gives rise to the so-called *de re* reading of (1). According to this reading, there is a jacket like Malte's which Adrian wants to buy. Or, in a scenario in which Adrian has not decided which jacket he wants to buy but he wants it to be like Malte's, (1) receives a *de dicto* interpretation, according to which Adrian wants to buy something that is a jacket like Malte's. Under the terms of conventional wisdom, the existence of readings is established by the difference that scenarios make to the interpretation of an attitude report. As (von Fintel and Heim, 2011, p. 85) put it: "[. . .] construct

---

scenarios that make one of the readings true and the other false. This establishes the existence of two readings.". Let me call such scenarios that entail the truth of an attitude report when it is taken in one way, but not when it is taken in another *reading-identifying scenarios*. Conventional wisdom also has it that the distinction between the *de re* and *de dicto* readings of an attitude report like (1) can be explicated in its logical form in terms of the scope relation between the modal verb *want* and the existentially quantified noun phrase *a jacket like Malte's*, as shown in (2).

(2) a. De re: $(\exists x)(\text{jacket}(x) \ \& \ \text{like-Malte's-jacket}(x) \ \& \ \text{wants}(\text{Adrian}, \text{buy}(\text{Adrian},x)))$
      b. De dicto: $\text{wants}(\text{Adrian}, (\exists x)(\text{jacket}(x) \ \& \ \text{like-Malte's-jacket}(x) \ \& \ \text{buy}(\text{Adrian},x)))$

There are three concerns about the conventional wisdom that the meaning of an attitude report finds expression in the readings it has and that the different readings can be captured as different scope relationships.

First, Fodor (1970) argued that there are more readings of (1) than just the *de dicto* and *de re* reading. For example, in a scenario in which Adrian has decided what kind of jacket he wants to buy but has no idea that the kind of jacket he wants is like Malte's jacket, (1) intuitively has a reading which neither the *de re* nor the *de dicto* reading in (2-a), (2-b) capture correctly. On the one hand, the existential quantifier introduced by *a jacket like Malte's* must be inside the scope of *wants*, since there is no one particular jacket that Adrian wants to buy. But that would bring the description *a jacket like Malte's* within the scope of *wants* and that seems wrong since Adrian is assumed to know nothing about Malte's jacket. But the noun phrase *a jacket like Malte's* cannot be both inside and outside the scope of the verb *want*. The literature has mainly focused on saving conventional wisdom from Fodor's problem by improving on the simple connection between scope and intensional status of a noun phrase that renders the *de dicto* and *de re* readings, see e.g. Romoli and Sudo (2009); Schwager (2009); Keshet and Schwarz (2014) for strikingly distinct proposals. But even if (improved theories of) scope and intensional status of a noun phrase could provide a semantic explanation of Fodor's paradox of scope, conventional wisdom is challenged by the other two problems, which are even more fundamental.

Second, we need an explanation why in first-person reports as in (3) there is no room for the traditional ambiguities, since (3) cannot be a true utterance by Adrian if Adrian isn't fully aware that what he wants to buy is a jacket like Malte's.

(3)   (Adrian says:) I want to buy a jacket like Malte's.

As the only difference between (3) and (1) is the grammatical subject of the report, conventional wisdom about the role of scope and intensional status of the indefinite noun phrase *a jacket like Malte's* cannot explain why (3) doesn't have the same variety of readings as (1). But if scope and intensional status of a noun phrase cannot account for why (1) allows for several readings while (3) only has one, then these are apparently not the right tools for dealing with these sentences.

Third, there is a difficulty with the conventional wisdom that the meaning of (1) (like that of more or less any sentence) is given by the range of its possible readings at least if these readings are associated with the scenarios that have been discussed in the literature. The problem here is how much an interpreter of (1) must know about the scenario which forces a certain reading upon it. The information that tells him to opt for one interpretation rather than another shouldn't give too much away. For otherwise the utterance of (1) couldn't teach the interpreter anything that he didn't already know from his perception of the scenario: an utterance of (1) would be uninformative for the interpreter in any situation where he understands enough about the scenario to be able to disambiguate (1) to its intended reading because every reading-identifying scenario of (1) entails the truth of (1).

These three problems with the conventional wisdom about the meaning of attitude reports are all symptoms, I want to argue, of an underlying mistake: that of lumping together theorizing about the attitudes themselves with theorizing about the meaning of attitude reports. What is needed is to separate the

analysis of the attitudes themselves from the semantics of their reports. The next section 2 implements this separation by distinguishing anaphoric constructions that refer back to what is the same entity from an internal point of view from those that corefer with their anaphoric antecedents from an external perspective. Anaphoric constructions that are internally about the same object make explicit the structure of the attitudes themselves – e.g. that desires are parasitic on beliefs whereas anaphoric constructions which are externally about the same object reveal the structure of attitude reports. Section 3 grounds the distinction between attitudes and attitude reports in the asymmetry between the first- and third-person perspective on attitudes. On the basis of this asymmetry, section 4 proposes a semantics of desire reports according to which (1) is not ambiguous between different readings but isolates the meaning that (1) has independently of the availability of reading-identifying scenarios. Conventional wisdom readings are reconstructed as third person strategies for justifying the truth of an attitude ascription in section 5. Section 6 concludes with a brief exploration of how the proposed analysis of desires and their reports can be extended to reports involving such non-doxastic predicates like *wish* or *hope* and reports involving doxastic predicates such as *believe*.

2   Intra- and interpersonal anaphora in attitude reports

Developing a line of thought that originated in Kamp (1985) with contrasts as in (4) (focused constituents in caps) Maier (2015) argues that non-doxastic attitudes such as desires are *parasitic* on their doxastic host attitudes.

(4)   a.   John believes that Mary will come. He hopes that SUE will come too.
        b.   *John hopes that Marry will come. He believes that SUE will come too.

Maier argues that the linguistic asymmetry in (4) "mirrors an underlying asymmetry in the logic of the attitudes themselves" (Maier, 2015, p. 209): "We only have desires relative to our beliefs" (Maier, 2015, p. 220). Semantically, the asymmetry between attitudes of different types manifests itself as the 'anaphoric' dependency of non-doxastic on doxastic attitudes. For example, if Adrian believes that the object in front of him is a jacket and decides that he wants to buy *it*, the object of his want is anaphorically dependent on the object he believes to be a jacket. In such cases, different attitudes of one and the same attitude holder are *internally* about the same object (adopting the terminology of Kamp et al. (2011)) in that the different attitudes are connected by *intrapersonal* anaphoric links. Attitudes can also be *externally* about the same object when "two expressions refer to the same real world entity, but where a particular speaker (or thinker) may be unaware of this, or alternatively where a speaker takes two expressions (or occurrences thereof) to refer to the same thing although in actuality they refer to distinct real world entities." (Kamp et al., 2011, p. 332). Attitudes that are externally about the same object are connected by *interpersonal* anaphoric links. Interpersonal anaphoric links are central to reading-identifying scenarios. Consider the scenarios for Fodor's puzzling reading in (5)-(7) (italics added and naming adopted to the running example of the paper).

(5)   Reading-identifying scenario for (1) (Romoli and Sudo, 2009, p. 427)
        Suppose a store sells some jackets that all look like Malte's and that Adrian does not know anything about Malte. Assume further that Adrian wants one of *those* jackets and any of *them* is an option.

(6)   Reading-identifying scenario for (1) (Keshet and Schwarz, 2014, p. 16)
        Adrian wants to buy either jacket A or jacket B, for instance, although he had not decided which yet. Both *jacket A* and *jacket B* actually happen to be jackets like Malte's, although Adrian may or may not know this.

(7)   Reading-identifying scenario for (1) (Schwager, 2009, p. 395)
        Malte and Adrian do not know each other. Adrian has seen a green Bench jacket in a catalogue
        and wants to buy one. Malte happens to own precisely *such* a green *Bench* jacket.

The background assumptions in (5)-(7) all serve the same purpose. They entail the truth of (1) while
rendering the *de dicto* and *de re* reading false. Thus, according to common wisdom, (5)-(7) establish
the existence of a further reading of (1) besides the *de dicto* and the *de re* reading. But although (5)-(7)
have been argued to motivate quite different logical forms for the reading identified, the presentation
of Adrian's attitude follows the same schema. Each scenario specifies the doxastic host of Adrian's
parasitic desire which does not include Adrian's awareness that what he wants to buy is a jacket like
Malte's. This description is the antecedent for the specification of the parasitic desire that is ascribed
to Adrian. It is important to note that the author of each of these scenarios invites us to consider (1)
in the light of their respective scenarios. More specifically, we are invited to interpret the infinitival
complement of *wants* in terms of what the scenario says about Adrian's information about the jacket
that he wants to get, and to agree with them that in the light of this information their specification of
Adrian's desire is the correct paraphrase of (1). It is important to note against this background that
in their own specification of Adrian's desire in the different scenarios the authors use an interpersonal
anaphoric expression that links the descriptive content of Adrian's desire to what is said about Adrian's
information[1]. In (5), the plural pronoun *those* makes explicit that the descriptive content of Adrian's
attitude is externally about the same object of reference that is identified by the description *a jacket like
Malte's*. In (6), the same anaphoric function is realized by the use of the proper names *jacket A* and
jacket B both inside and outside of Adrian's attitude. In (7), a combination of the pronoun *such* and
the proper name *Bench* for a kind of jackets is used to indicate how Adrian's own concept of what he
wants to buy is connected to the 'external' description that the authors have just provided of that target.
The crux of the use of such interpersonal anaphora in reading-establishing scenarios is the presentation
of the same object of reference from two distinct points of view. These observations about the central
role of anaphora in attitude reports generalize to the other conventional wisdom readings of (1). *De
re* reading-identifying scenarios require a singular interpersonal anaphoric specification and *de dicto*
reading-identifying scenarios involve an intrapersonal anaphor. Consider the textbook scenarios for *de
re* (8) and *de dicto* (9) (von Fintel and Heim, 2011, adopted from p. 100, italics added).

(8)   I am walking along Newbury Street with Adrian. Adrian sees a jacket in a display window and
        wants to buy *it*. He tells me so. I don't reveal that I have one just like *it*. I report (1) to you.

(9)   Adrian's desire is to buy some jacket or other which fulfills the description that *it* is just like
        mine.

The role of inter- and intrapersonal anaphora in attitudes and their reports is the fundamental phe-
nomenon to be explained by a semantic theory of the meaning of attitude reports. Fodor's reading
shows that scope relationship reflects some of the aspects of attitude reports that anaphoric construc-
tions can express, but not all. In fact, Fodor's puzzle is not the only case where anaphoric constructions
create apparent scope paradoxes, compare e.g. Geach (1967)'s Hob-Nob sentences. On the one hand,
the goal of this paper is to make the linguistic distinction between intra- and interpersonal anaphoric ex-
pressions as sharp as possible. On the other, the goal of this paper is also to show that only when taken
together, intra- and interpersonal anaphoric expressions provide the basis for the semantic analysis of
the attitudes themselves and their reports. To this end, in the next section, I explain the role of anaphora
in attitude reports in terms of the asymmetry of first- and third-person reports observed with (3) and (1).

---

[1]A similar observation about the role of interpersonal anaphora in *de re*-identifying scenarios is made in footnote 7 to chapter
8 of Brandom (1994)

3   The asymmetry of first- and third person perspective

(1) ascribes a desire to Adrian from a third-person perspective whereas (3) is a report from Adrian's own first-person perspective. Only the first-person perspective grants introspection of one's own mental states (i.e. the attitudes themselves). A third person has no direct access to another person's mental states. The asymmetry in the accessibility of mental states finds expression in the way the truth of an attitude report is justified given that a justfication of a claim is a set of observations that, when taken together entail the truth of the claim. The truth of a first-person report is justified just in virtue of the authority of the first person over what her state of mind is like. A rational justification for the self-ascription of a desire[2] as in (3) would make explicit how the reported desire is parasitic on a set of beliefs. For the case of (3), Adrian could justify his desire by grounding it in a description of the doxastic state which gave rise to the desire: maybe he saw a jacket like Malte's and liked it and *thus* wants to buy a jacket like the one he saw. Linguistically, the justification of a first-person report amounts to making explicit the intrapersonal anaphoric dependency of a parasitic attitude. Things are different for a third-person report like (1). Given that only the first-person has authority over her state of mind, (1) cannot be justified to be true just in virtue of the reporter saying so. Instead, (1) can be justified in two fundamentally different ways[3]. The simpler strategy to justify the truth of (1) is when Adrian told the reporter: "I want to buy a jacket like Malte's". In this case, the third-person reporter can simply reproduce Adrian's self-ascription without taking on any further responsibility for why (1) is true besides the fact that Adrian said so and, because Adrian is granted introspection of beliefs, Adrian could rationalize his desire by specifying the antecendet of the desire with an intrapersonal anaphoric construction. This justification of the truth of a third-person attitude report is called the *de dicto* strategy of justification: "the words which I, the speaker, am using to describe the attitudes content, are the same (at least as far as the relevant DP is concerned) as the words that the subject herself would use to express her attitude." (von Fintel and Heim, 2011, p. 84). Besides the delegation of responsibility to a first-person *de dicto* (i.e. according to what the first-person said), a third-person reporter can also justify (1) with an interpersonal anaphoric construction. The interpersonal justification of a desire ascription requires the reporter to be able to figure out the common object which the reportee considers to be relevant to his desire and to which the reporter herself can refer. The justification of the truth of an attitude report as being externally about the same object is called the *de re* justification: "[t]he term *de re* [...] indicates that there is a common object [...] whom I (the speaker) am talking about [...] and whom the attitude holder would be referring to if he were to express his attitude in his own words." (von Fintel and Heim, 2011, p. 84).

4   The attitudes themselves and reports of attitudes

In this section, I develop a semantic theory of desires themselves and their reports that allows to distinguish theorizing about the attitudes themselves from theorizing about attitude reports on the one hand and justification strategies for the truth of an attitude report (provided by reading-identifying scenarios) from the truth-conditions of an attitude report. Heim (1992) popularized a semantic analysis of desire reports based on Stalnaker's proposal for a qualified consequence condition for rational desires that deals with the problem of logical omniscience: "wanting something is preferring it to certain relevant alternatives, the relevant alternatives being those possibilities that the agent believes will be realized if he does not get what he want." (Stalnaker, 1984, p. 89). It is important to note that this characterization of desires is concerned with the logical properties of an agent's first-person perspective on his

---

[2]I use the term *rational* here in the sense of Davidson (1963)'s rationalization of actions.

[3]There is in principle a third justification strategy which I do not consider here, namely abduction from observed behaviour and conventional goals of behaviour to the best explanation. E.g. if Adrian prowls around the fridge continously, one is justified to infer that he has the desire to eat something.

attitudes, e.g. an agents beliefs about his beliefs or the ways in which an agent's beliefs may change in response to his experiences (Stalnaker, 1984, cp. p. 80). But the problem that arises for third-person reports is a more fundamental one. To a third person, another agent's state of mind is basically a black box, the inputs and outputs to which can be observed but not its internal workings. This is true in particular of the third-person perspective on an agent's beliefs (including those beliefs that a first person considers in her preferences). Instead of the introspection of beliefs that is granted by the first-person perspective, a third-person attitude report like (1) must be based on observations about the behaviour of the reportee that the reporter considers to be indicative of his state of mind. Stalnaker is one of many philosophers who argued that an agent's behaviour indicates that she desires that P if "[t]o desire that P is to be disposed to act in ways that would tend to bring it about that ϕ" (Stalnaker, 1984, p. 15), i.e. desires are pro-attitudes Davidson (1963) that are capable of playing a causal role in their induction of a tendency to act in a certain way. Consequently, (1) is not a report of Adrian's attitude itself (which is inaccessible to a third person) but presents a third person's assessment of Adrian's behaviour guided by justification strategies for the truth of (1). In other words, I argue that the truth-conditions of (1) do not reflect the conditions under which (1) is a true description of Adrian's attitudes themselves but that the truth-conditions of (1) reflect the conditions under which (1) is a true description of Adrian's behaviour. To deal with the distinction between the attitudes themselves and their reports, I propose that linguistic descriptions of the form *x wants ϕ* are to be interpreted as in (10), a simplified version of the action-theoretic analysis of desires proposed in Schroeder (2015).

(10)   A report of the form "x wants ϕ" is true if and only if x is disposed to take those actions which are likely to bring about a state of affairs of which the reporter is justified to believe that ϕ obtains.

The interpretation of linguistic descriptions of the form *x wants ϕ* in (10) separates the semantics of a desire report from the mental state of desiring itself. It captures that third-person reports are based on the observable realization of a desire – the behaviour in which it manifests itself – instead of the unobservable attitudes themselves. (10) thus provides the possibility to distinguish the first- and third-person perspective. In a first-person perspective report like (3), the reportee is identical with the reporter and thus the truth of (3) is justified by intrapersonal reference to the attitudes themselves, thereby making explicit the joint responsiblity of desire and belief for the actions undertaken by the reportee. Things are different for attitude reports in which the reportee is not identical with the reporter. Third person reports do not make a claim about Adrian's attitudes themselves – although Adrian himself certainly has a certain combination of desire and belief that makes him act in a certain way. Instead, (1) is true as long as the reporter is justified to claim that Adrian's actions bring about a state of affairs where ϕ obtains. To make more preceise the distinction of truth-conditions of an attitude report and justification strategies for the truth of an attitude report, it is helpful to formalize the behaviour-based account of desires in (10).

To the development of a formal account of desires as pro-attitudes, I take to be central what Stalnaker calls the future-looking nature of desires, i.e. the fact that the reporter of (1) or (3) claims that the reportee acts towards a future state of affairs at which the descriptive content of the attitude report eventually becomes true. A formal framework in which such an analysis can be captured has been proposed in Singh and Asher (1993). The basic idea – I refer the reader to Singh and Asher (1993) for full details – is to include in a standard model theory of the language of Discourse Representation Structures (DRS) (see Kamp et al. (2011)) a time structure $\mathscr{T}$ and two functions $\mathbf{B}$ and $\mathbf{D}$. $\mathscr{T} = \langle \mathbf{T}, < \rangle$, where $\mathbf{T}$ is a set of possible times and $<$ a partial order on $\mathbf{T}$. We may view $\mathscr{T}$ as a tree-like structure of times branching towards the future in which arc labels are basic actions. That is, the order $<$ on $\mathbf{T}$ is transitive, asymmetric and does not allow for merging of branches. We define a scenario S at a world w and time t as any maximal branch starting from t and $\mathbf{S}_{w,t}$ is the class of all scenarios at world w and

time t. **B** and **D** are functions defined for a set of 'agents' (a subset of the set of individuals in the model theory) to subsets of $\mathbf{S}_{w,t}$ which assign beliefs **B** and desires **D** to agents at different worlds and times (qua modelling). On the basis of these additional postulations, (Singh and Asher, 1993, p. 523) propose a semantics for intentions and beliefs, which I adopt in (11) to capture the difference between desires themselves (a set of scenarios) and desire ascriptions (beliefs about the goal of acting on a desire)[4]. (11) defines the interpretation of the DRS-predicate $DES(a,K)$, where the first argument is an agent and the second argument is a DRS K, as the set of scenarios S at t such that if the world w developed along any of them, the reporter b believes that the desire described with K would be realized at $t' > t$, i.e. that K has a verifying embedding g in M at w at t'.

(11)   Interpretation of $DES(a,K)$:
        $[\![DES(a,K)]\!]_{M,w,t,g} = \{S | \exists w,t : S \in \mathbf{S}_{w,t} \wedge (\exists t',f : t' \in S \wedge g_w \subset_U f_w \wedge M \vDash_{w,t',f} BEL(b,K))\}$,
        where a is the bearer of the desire and b the ascriber of the desire

(12)   Verifying embedding of $DES(a,K)$:
        $M \vDash_{w,t,g} DES(a,K)$ iff $[\![DES(a,K)]\!]_{M,w,t,g} \in \mathbf{D}(w,t,a)$

Singh and Asher (1993) also propose a definition of beliefs relative to scenarios as the set of scenarios S at whose initial world w and time t, the descriptive content K of the belief is true under the given embedding g, cp. (13). I do not relativize belief ascriptions to the reporter-reportee distinction here, but see section 6 for additional discussion.

(13)   Interpretation of $BEL(a,K)$
        $[\![BEL(a,K)]\!]_{M,w,t,g} = \{S | \exists w,t : S \in BEL(w,t,a) \wedge (\exists f : g_w \subset_U f_w \wedge M \vDash_{w,t,f} K))\}$

(14)   Verifying embedding of $BEL(a,K)$:
        $M \vDash_{w,t,g} BEL(a,K)$ iff $[\![BEL(a,K)]\!]_{M,w,t,g} \in \mathbf{B}(w,t,a)$

Given the model-theoretic background on the interpretation of desire ascriptions, consider (15), the semantic representation of the first-person report (3) in the representation formalism for attitudes developed in Maier (2015). DRS_1 labels (for the sake of presentation) the representation for the global belief state of Adrian in which the desire representation labelled with DRS_2 is embedded. (15) is true iff there is a set of scenarios which belong to Adrian's desire set such that if Adrian would act on any of these scenarios the reporter (i.e. Adrian) believes that DRS_2 eventually becomes true.

(15) $DRS_1 : \boxed{\begin{array}{l} i \\ \hline DES - DRS_2 : \boxed{\begin{array}{l} y,e \\ \hline e : buy(i,y) \\ like-Maltes-jacket(y) \end{array}} \end{array}}$

(16) $DRS_1 : \boxed{\begin{array}{l} x \\ Adrian(x) \\ \hline Att(x) : DRS_2 : \boxed{\begin{array}{l} DES - DRS_3 : \boxed{\begin{array}{l} y,e \\ \hline e : buy(x,y) \\ like-Maltes-jacket(y) \end{array}} \end{array}} \end{array}}$

The representation of (1) in (16) involves a subtle but important difference. In (1), the reportee is not identical with the reporter and thus the semantics in (11) predicts that (1) is true iff there is a set of

---

[4]The relevant difference between the semantics proposed in Singh and Asher (1993) for intentions and the version adopted here to model desires concerns the relativization of goal descriptions. To me, goal relativization seems to be the relevant linguistic difference between intentions and desires, given the fact that *Adrian intends to buy a jacket like Malte's* can only be a true description of Adrian's behaviour if Adrian is aware of the descriptive content of the intention ascription.

scenarios which belong to Adrian's desire set such that if Adrian would act on any of these scenarios the reporter (i.e. not Adrian) believes that DRS_3 eventually becomes true.

5   The justification of attitude reports

The representations in (15) and (16) capture the model-theoretic truth-conditions of attitude reports. But given what has been said so far, an equally important aspect of the interpretation of attitude reports is their justification. In the definition of the semantics of desire ascriptions in (12), I glossed over the role of justification in assuming that it is enough for a reporter of (1) or (3) to believe that what the reportee is disposed to do will bring about a state of affairs in which the reportee bought a jacket like Malte's. But belief seems to be too weak to provide an adequate characterization of the conditions under which an attitude report is justified. In this section I consider in more detail the strategies which a reporter of (1) can employ to justify the truth of her belief that Adrian wants to buy a jacket like Malte's, where justified true belief amounts to an answer to the question put to the reporter "How do you *know* that (1)?"

For (3), Adrian can justify his self-report by specifying the intrapersonally antecedent host of his desire to buy a jacket like Malte's. For example, Adrian could justify (3) by grounding his desire in the belief that jackets like Malte's are trendy and that trendy jackets are desirable. Because *de dicto* justification in terms of revealing the intrapersonal anaphoricity of the desire in (3) is the only way in which a rational agent can justify a self-report[5], we expect a difference between (3) and (1), for which several justification strategies are available. Besides the delegation of the responsibility for the justification of the truth of (1) *de dicto* to Adrian himself, the reporter of (1) can also justify the truth of (1) *de re*.

The textbook *de re* justification of (1) starts from the assumption that Adrian does not know Malte or his jacket but ignores the fact that this assumption itself is a third-person attitude report. From the point of view advanced in this paper, this is a mistake: it is interpersonal anaphoric constructions, not doxastic attitudes that are relevant to *de re* justification. Consider the following reasoning. Self-reports like (3) do not allow for *de re* justification because Adrian's report of his doxastic background with *I do not know Malte or his jacket* would be incompatible with (3). Consequently, a third person cannot challenge the justification of the truth of (3) by the justified assumption that Adrian does not know about Malte or his jacket (as this would require Adrian to contradict himself) and *de dicto* is the only justification strategy for (3). Similarly, considerations about doxastic states of Adrian are irrelevant when (3) is used as a *de dicto* justification of (1). But the same conclusion about the irrelevance of doxastic states of Adrian holds for the *de re* justification of (1). The crucial point about *de re* justification of (1) is not whether or not Adrian knows about Malte or his jacket but that it rests on an interpersonally anaphoric self-report such as *I want to buy it* (e.g., as in (8)). It is this self-report which the reporter of (1) considers as the *de dicto*-justified basis – i.e. that he is justified to say that *Adrian wants to buy it* – of his *de re* justification. In the *de dicto*-basis of *de re* justification, the anaphoric expression serves the identification of an object external to the reportee's state of mind[6] but does not hint at any particular description that the reportee associates with the antecedent of the anaphoric expression. In other words: it is the anaphoric nature of Adrian's self-report that licenses the *de re* justification of (1) after all, independently of background assumptions about the doxastic state of the reportee. Furthermore, the type of anaphor that occurs in the *de dicto*-basis of a *de re* justification determines how a third person can charge the anaphora with

---

[5]It should be noted that I ignore here the fact that the *de dicto* justification of self-reports actually decomposes itself into different justification strategies, depending on e.g. whether or not the intrapersonal antecedent of a parasitic desire is in turn grounded in a causal relation of acquaintance.

[6]If "I want to buy it" would refer to an unpronounced belief of Adrian about a certain jacket but not to an object external to Adrian's state of mind, a reporter would not be able to report his desire with (1) just because there is no way in which "I want to buy it" and (1) could be externally about the same kind of jacket.

descriptive content. The basic case is where a singular anaphoric description such as "I want to buy it" licenses a *de re* justification of the descriptive content of (1) if the reporter is able to resolve *it* to the antecedent description *a jacket like Malte's jacket*. The complex case induces the reading that puzzled Fodor. The self-report of Adrian which underlies Fodor's puzzle involves a particular kind of complex anaphoric device of reference, e.g. as in *I want to buy one of those jackets* (cp. (5)). Compared to the interpersonal interpretation of the singular and simplex anaphoric expression *it*, what is special about *one of those* is that the expression involves not one but two anaphoric elements. Adrian's desire is directed towards an (unspecific) object that he (and only he himself) can identify with *one of*. But this intrapersonal anaphora cannot serve as the interpersonal anaphora that a reporter can pick up in a *de re* justification. Instead, what justifies (1) *de re* is the specific object that Adrian identifies with *those* (and similar conclusions hold for the other types of complex interpersonal anaphoric expressions discussed in section 2). The problem that emerges from complex uses of anaphora is that they require the interpreter to distinguish the role of anaphora in self-attributions of attitudes themselves (the intrapersonal specification of *one of*) from the role of anaphora in attitude reports (the interpersonal specification of *those*). Consequently, it is not surprising that justifications that involve complex interpersonal anaphoric constructions are puzzling from the perspective of conventional wisdom. It is because conventional wisdom conflates theorizing about the attitudes themselves with theorizing about their reports that it conceals the source of Fodor's puzzle: not only do anaphors structure the attitudes themselves intrapersonally and attitude reports interpersonally, but in *de re*-justification anaphors sit astride between attitudes and their reports in that they mediate between attitudes and their reports. I take this role of anaphora to be indicative of what justification strategies themselves are.

If there is a truth at all to the claim that attitude ascriptions obtain validity from the very fact that we accept *de re* and *de dicto* justifications as being suitable licensors for our talk about attitudes, then justification strategies are part of the pragmatic conditions for the successful *use* of attitude ascriptions. Considering *de re* and *de dicto* justification as regulating our use of attitude ascriptions makes them part of the system of social conventions according to which we ascribe attitudes: it is a social convention that a person has the descriptive authority over her own mental states and it is also a social convention that a third-person report cannot claim authority over the description of other persons' mental states. First-person attitude reports like (1) cannot be justified *de re* without giving up the social convention that a person has the authority over her own mental states. Thus, the function of *de re* and *de dicto* justification in regulating the practice of attitude ascriptions is part of a theory of the pragmatics of attitude reports. It is on the grounds of the acceptance of the implicit pragmatic success conditions of an attitude report generally that (1) is semantically informative to an interpreter. As mentioned earlier, one semantic function of attitude reports (what a reporter *does* when reporting an attitude) is the rationalization of an agent's behaviour. But the more important semantic function of attitude reports is in the way in which attitude ascriptions figure in practical reasoning. The semantics for desire ascriptions that I presented is motivated on exactly these grounds. Singh and Asher (1993) argue in detail how the action-theoretic analysis of belief, intentions and desires provides the basis for the definition of a logic of rational agency in which agents can reason about themselves and other agents. It is important to note at this point that also Stalnaker (1984) is concerned with the delineation of such a system of the logical relations that underly a rational agent's inquiry. A Stalnaker-type logic of desire and beliefs is thus not incompatible with the proposal for attitude reports made in this paper as long as we stick – as Stalnaker does anyway – to the first-person perspective when theorizing about the attitudes themselves.

6   Conclusion

This section concludes with some remarks on how the proposed account of desire reports can be generalized. First, the analysis of desire reports proposed can be modified so as to deal with other non-doxastic

attitudes such as *hope* or *wish*, the main difference between *hope, wish* and *want* manifesting not in their propositional content but in the strength with which they induce a tendency to act in the bearer of the attitude. While desires are understood as inducing a strong tendency to act in a certain way, a wish basically omits the forward-looking nature of desires and thus also can be used for goals which an agent cannot accomplish with his actions, like *I wish that I was never born.* vs. **I want that I was never born.* With respect to the relation of attitude and action, hopes seem to stand inbetween desires and wishes, as one the one hand, they share the forward-looking nature of desires but on the other also allow for outcomes of actions that an agent does not have under its own control like *I hope to pass the exams without preparation* vs. *I want to pass the exams without preparation*. Second, once the distinction between the first-person perspective on the attitudes themselves and their third-person reports has been established, the argument of this paper generalizes to doxastic attitudes like belief. On the one hand, like desires, an agent's beliefs can be observed from an external perspective only through the behaviour in which they manifest themselves under the hypothesis that "[t]o believe that P is to be disposed to act in ways that would tend to satisfy one's desires" (Stalnaker, 1984, p. 17). On the other, like desires, beliefs are parasitic attitudes: "[b]eliefs have determinate content because of their presumed causal connections with the world" (Stalnaker, 1984, p. 17). It is the combination of both aspects of beliefs that should be considered to be fundamental to their reports, given that the justification of belief reports necessarily involves an anaphoric specification of how the reported belief is parasitic on its causal relation of acquaintance, compare e.g. (Quine, 1956, p. 179)'s famous Ortcutt scenario in (17) (italics added).

(17)   "There is a certain man in a brown hat whom Ralph has *glimpsed* several times [. . .] Ralph suspects *he* is a spy. Also there is a gray-haired man, [. . .] whom Ralph is not aware of having *seen* except once at the beach. Now Ralph does not know it, but *the men* are one and the same."

References

Robert B. Brandom. *Making It Explicit. Reasoning, Representing, and Discursive Commitment.* Harvard University Press, Cambridge, MA, 1994.

Donald Davidson. Actions, reasons and causes. *Journal of Philosophy*, 60:695 – 700, 1963.

Janet Dean Fodor. *The linguistic description of opaque contents.* PhD thesis, MIT, 1970.

Peter T. Geach. Intentional identity. *Journal of Philosophy*, 64(20):627 – 632, 1967.

Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9:183–221, 1992.

Hans Kamp. Context, thought and communication. *Proceedings of the Aristotelian Society*, 85:239–261, 1985.

Hans Kamp, Josef van Genabith, and Uwe Reyle. Discourse Representation Theory. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic* 15:125 – 394. Springer, 2011.

Ezra Keshet and Florian Schwarz. De re/de dicto. 2014. URL http://florianschwarz.net/wp-content/uploads/papers/De_Re___De_Dicto.pdf.

Emar Maier. Parasitic attitudes. *Linguistics and Philosophy*, 38(3):205–236, 2015.

William Van Orman Quine. Quantifiers and propositional attitudes. *Journal of Philosophy*, 53(5):177 – 187, 1956.

Jacopo Romoli and Yasutada Sudo. De Re / De Dicto Ambiguity and Presupposition Projection. In Arndt Riester and Torgrim Solstad, editors, *Proceedings of Sinn und Bedeutung 13*, pages 425 – 438, 2009.

Tim Schroeder. Desire. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2015 edition, 2015.

Magdalena Schwager. Speaking of qualities. In Ed Cormany, Satoshi Ito, and David Lutz, editors, *Proceedings of SALT 19*, 2009.

Munindar P. Singh and Nicholas Asher. A logic of intentions and beliefs. *Journal of Philosophical Logic*, 22:513 – 544, 1993.

Robert C. Stalnaker. *Inquiry*. MIT Press, 1984.

Kai von Fintel and Irene Heim. Intensional semantics. MIT, Spring 2011 Edition, 2011.

Towards a Taxonomy of Textual Entailments[*]

Adam Przepiórkowski, Jakub Kozakoszczak, Jan Winkowski, Daniel Ziembicki,
and Tadeusz Teleżyński

[1] Institute of Computer Science, Polish Academy of Sciences, Warszawa, Poland
[2] University of Warsaw, Warszawa, Poland
adamp@ipipan.waw.pl, jkozakoszczak@gmail.com, jaswinko@gmail.com, seattlegrunge@wp.pl,
tadekte@gmail.com

Abstract

This paper reports on work in progress which aims at the creation of a comprehensive taxonomy of textual entailment rules and – as a proof of concept – its application to an RTE corpus. In particular, a new formalism for encoding entailment rules is proposed, with some emphasis on encoding the properties of such rules and their converses in different polarity contexts.

1 Introduction

Ever since Aristotle, entailment has been in the centre of interest in logic and semantics, and since mid-2000s the more loosely defined textual entailment has become increasingly important in Natural Language Processing (NLP), where semantic systems are evaluated by their performance on the Recognising Textual Entailment (RTE) task (Dagan *et al.* 2006, 2013, Sammons 2015; see also Bos 2008). To this end, RTE corpora have been created[1] consisting of ⟨T, H⟩ pairs tagged by humans with information whether the hypothesis H is textually entailed by the attested text T, or not; in case of some corpora, a third tag, signalling contradiction between T and H, is also employed. The quality of a system is measured by the degree of agreement between human judgements and tags assigned by the system. The additional difficulty of the task lies in the fact that textual entailment is understood very generally and a little vaguely; the following definition is typical: "[The] applied notion of *textual entailment* is defined as a directional relationship between pairs of text expressions, denoted by T – the entailing 'Text', and H – the entailed 'Hypothesis'. We say that T entails H if, typically, a human reading T would infer that H is most likely true" (Dagan *et al.* 2006: 178).

Evaluation against such RTE corpora is purely quantitative: the only information it provides is the percentage of correctly tagged pairs, and not the kinds of phenomena the system does not handle correctly or the kinds of knowledge it lacks. For this reason, since around 2010, the need has been voiced for corpora containing pairs tagged with the types of phenomena or knowledge involved in the process of reasoning (Sammons *et al.* 2010). Different types of textual inference had, of course, been discussed earlier (e.g. Cooper *et al.* 1996, Zaenen *et al.* 2005, Garoufi 2007, Clark *et al.* 2007, Bos 2008), but such discussions resulted in identifying some types of phenomena or kinds of knowledge, without an attempt at creating an exhaustive taxonomy of types of textual entailment. Other work discusses selected types or aspects of reasoning, e.g. monotonicity effects (MacCartney and Manning 2007), common-sense knowledge (LoBue and Yates 2011), linguistic modification (Toledo *et al.* 2012), etc.

---

[1]Some are listed at http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool#RTE_data_sets.

To the best of our knowledge, no comprehensive taxonomy of types of textual entailment has been proposed so far, and the most advanced attempt is still that of Bentivogli *et al.* 2010. There, 90 pairs extracted from the RTE-5 (Bentivogli *et al.* 2009) corpus were annotated with kinds of phenomena involved in the entailment. The 36 distinguished phenomena were grouped into 5 categories: lexical (e.g. *hypernymy* and *geographical knowledge*), lexical-syntactic (e.g. *causative* and *paraphrase*), syntactic (e.g. *negation*, *list* and *apposition*), discourse (e.g. *coreference* and *ellipsis*) and reasoning (e.g. *apposition* again, *meronymy*, *relative clause* and the most frequent "phenomenon" of *common background / general inferences*). As the authors note themselves, "the list is not exhaustive and reflects the phenomena [they] detected in the sample of RTE-5 pairs [they] analyzed". Unfortunately, the scope of particular phenomena and reasons for classifying, say, *geographical knowledge* as a lexical phenomenon and *meronymy* as reasoning, are not elucidated, as that short paper concentrates on general methodological issues.

The aim of the current paper, which reports on work in progress, is to start filling this gap, i.e. to report on early research leading to 1) the creation of a comprehensive taxonomy of textual entailment steps and – as a proof of concept – 2) its application to an RTE corpus. We assume that each ⟨T, H⟩ pair tagged as involving entailment may be associated with a sequence of atomic entailment steps, cf. §2, that each such a step may be formalised as a rule transforming one text into another text (or, more declaratively, specifying a relation between texts), cf. §3, and that such rules may be organised into a taxonomy of textual entailment types, cf. §4. In specifying the format of the rules and the resulting taxonomy, we combine and extend ideas from the literature which have not been put together before.

While all the examples provided in this paper involve English, the actual entailment corpus is a Polish translation of the development part of the RTE-3 corpus (Giampiccolo *et al.* 2007) consisting of 800 ⟨T, H⟩ pairs, including 411 pairs tagged as involving true entailment. At the point of submitting this paper to the proceedings, over 120 ⟨T, H⟩ pairs are fully tagged with almost 800 atomic entailment steps (over 6 steps per pair, on average), and some 50 of the rules used in these entailments – jointly covering about 120 atomic entailment steps – are fully formalised in the sense of §3. While, as discussed there, some of the rules are language-specific and others are language-independent, they are organised into a taxonomy outlined in §4, which is designed to be largely language-independent. The possible uses of such a corpus and taxonomy are mentioned in the concluding remarks in §5.

2   Entailment corpus

One novel aspect of the research presented here consists in the construction of a textual entailment corpus, where particular entailment pairs are not only tagged with kinds of reasoning involved, as already postulated (but not fully realised) in the literature, but where all transformation steps leading from the initial text to the final hypothesis are explicitly shown, as illustrated in Figure 1.[2] The five steps shown there conflate eight actual steps in the corpus. In the first step of the figure, appropriate rules for dropping a modifier are applied three times: for the two temporal modifiers *In 1927* and *until his death in 1962*, and for the subordinate *following on its success.* Similarly, the final step in the figure involves two applications of a rule substituting a term with its synonym: *penned* is replaced by *authored* and *numerous* – by many.

Obviously, this is only one of possible orderings of rule applications, but in many cases, including the example above, the ordering does not matter. An example of where the order of

---

[2]In the actual corpus, particular steps are labelled with the identifiers of specific rules. Moreover, the rule alluded to as *distributing shared dependent* is not strictly necessary here, but we leave it here as its formalisation will be discussed in the following section.

T: In 1927 Harnold Lamb wrote a biography of Genghis Khan, and following on its success turned more and more to the writing of non-fiction, penning numerous biographies and history books until his death in 1962.

    → Harnold Lamb wrote a biography of Genghis Khan, and turned more and more to the writing of non-fiction, penning numerous biographies and history books.   (dropping modifiers)

    → Harnold Lamb turned more and more to the writing of non-fiction, penning numerous biographies and history books.   (dropping conjunct)

    → Harnold Lamb penned numerous biographies and history books.   (extracting -*ing* modifier)

    → Harnold Lamb penned numerous biographies and numerous history books.   (distributing shared dependent)

    → Harnold Lamb penned numerous biographies.   (dropping conjunct)

H: → Harnold Lamb authored many biographies.   (substituting synonyms)

Figure 1: Entailment steps from text T to hypothesis H for the RTE-3 pair id=65

rule applications does matter is given in Figure 2, where the first sentence of the text T may be dropped only once the coreference rule fires, as otherwise *the team* could not be replaced by the (already dropped) *The Kinston Indians*.

T: The Kinston Indians are a minor league baseball team in Kinston, North Carolina. The team, a Class A affiliate of the Cleveland Indians, plays in the Carolina League.

    → The Kinston Indians are a minor league baseball team in Kinston, North Carolina. The Kinston Indians, a Class A affiliate of the Cleveland Indians, plays in the Carolina League.   (substituting coreferring term)

    → The Kinston Indians, a Class A affiliate of the Cleveland Indians, plays in the Carolina League.   (dropping sentence)

    → The Kinston Indians plays in the Carolina League.   (dropping modifier)

    → Kinston Indians play in the Carolina League.   (substituting synonym)

H: → Kinston Indians participate in the Carolina League.   (lexical implication)

Figure 2: Entailment steps from text T to hypothesis H for the RTE-3 pair id=92

Occasionally, it is possible that the same hypothesis may be reached from the text via an application of different sets of rules. Nevertheless, as the task of devising a single line of reasoning is already rather difficult, no attempt is made to provide all possible minimal sequences of rule applications, even once a set of rules is fixed. It should be noted that this procedure goes further than that of Bentivogli *et al.* 2010, where particular reasoning steps were only applied to the initial Text, without an attempt at constructing a complete transformation path from Text to Hypothesis.

3 Textual entailment rules

We propose an extension and further formalisation of the usual textual entailment rules (Szpektor *et al.* 2007) mapping one text into another (uni- or bidirectionally). We assume that such rules operate on dependency trees, i.e. their Left- and Right-Hand Sides are templates matching a dependency (sub)tree.[3] The rules constructed so far operate at the level of syntactic dependency trees, where word nodes are marked with lemmata and with morphosyntactic information (part of speech, case, gender, etc.), and with dependency labels such as *subject* and *object*,[4] but we envisage the possibility of some rules operating at the level of semantic dependencies, with labels such as *agent* or *instrument*.[5]

Formally, a rule is a quintuple whose first two elements are dependency trees (to match the input and the output of the rule), followed by a polarity signature,[6] strength specification and applicability conditions: ⟨LHS, RHS, P, S, C⟩. A preliminary version of a simple rule for dropping a premodifier expressed by a nominal or prepositional phrase is shown in (1):[7]

(1) ⟨ X.G Y.L, Y.L, ⟨+, ∘, ∘, +⟩, S, {} ⟩

The LHS and RHS elements are particularly simple here: LHS matches any dependency of the type `nmod` (nominal modifier) where the dependent (indicated by the variable X) is on the left of the head (indicated by Y), and the RHS is the result of dropping the dependent (i.e. leaving Y). An example of the operation of this rule is given in Figure 3. This example also illustrates
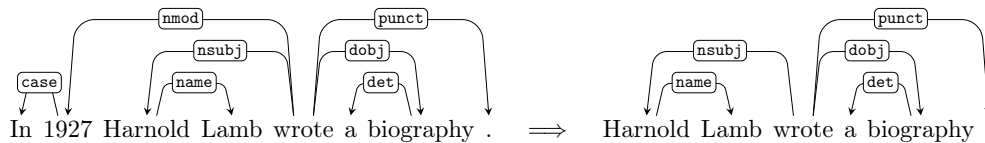
Figure 3: An example of the operation of rule (1)

the need for the markers L and G adorning the nodes in LHS and RHS: the former stands for *lazy match*, the latter – for *greedy match*. Lazy match occurs when a node is matched but those of its dependents which are not explicitly specified in the rule are not "consumed" during the match – such dependents will be rewritten to the output of the rule. On the other hand, greedy match indicates the need to "consume" – i.e. get rid of – any unspecified dependents. Hence, when Y.L matches *wrote*, all its dependents which are not explicitly mentioned by the rule – i.e. the subject *Harnold Lamb* and the direct object *a biography* – will be rewritten together with Y in the output. On the other hand, when X.G matches *1927*, its dependent[8] *In* will also be consumed rather than being adjoined to some higher node in the output.

---

[3]See Tesnière 1959, 2015 for the original work on dependency grammar and de Marneffe *et al.* 2014 for the approach assumed here.

[4]In all examples, part of speech tags and dependency labels follow the Universal Dependencies (`http://universaldependencies.github.io/docs/`).

[5]Two of the linguistic theories that posit multiple parallel dependency levels are Meaning-Text Theory (Mel'čuk 1981, 1988) and Functional Generative Description (Sgall *et al.* 1986).

[6]Generalisation to monotonicity is planned for future work.

[7]Analogous rules take care of nominal postmodifiers and clausal premodifiers, as also needed in the first step of Figure 1.

[8]Admittedly, treating nouns as heads and prepositions as their dependents in prepositional phrases is a controversial aspect of Universal Dependencies.

The third element of the rule quintuple is the polarity signature, which indicates the behaviour of the rule (the first two characters) and its converse (the next two characters) depending on the polarity of the context (to be discussed below). It is best to visualise this quadruple signature as a 2×2 table, where the first column indicates the behaviour of the rule in positive polarity contexts and the second column – in negative polarity contexts, and the first row describes the rule from LHS to RHS, while the second row – its converse, i.e. the rule from RHS to LHS:

(2) $\langle 1, 2, 3, 4 \rangle$ $\equiv$

	Pos	Neg
LHS $\rightarrow$ RHS	1	2
LHS $\leftarrow$ RHS	3	4

Following the discussion in Nairn *et al.* 2006, we assume that one of three possible values may occur in each of the cells:

(3)

+	the rule is valid in the specified direction and polarity context without any further conditions,
−	the rule is valid in the specified direction and polarity context, but the polarity of the consequent must be reversed,
∘	neither of the above, i.e. the rule is not applicable in the specified direction and polarity context.

Hence, the polarity specification $\langle +, \circ, \circ, + \rangle$ in (1) means that the rule may be applied in positive contexts (as in Figure 3), but not in negative contexts, and its converse may be applied in negative – but not in positive – contexts (e.g. to transform *Harnold Lamb didn't write a biography.* into *In 1927 Harnold Lamb didn't write a biography.*).

Following the suggestion of Zaenen *et al.* 2005, we also further classify rules as *strong* (S; including semantic entailments and presuppositions) and *defeasible* (DF; including conversational implicatures). Since an attempt is made to annotate with entailment steps all $\langle T, H \rangle$ pairs marked in the development section of RTE-3 as involving true entailment, even those where entailment steps are so weak or counterintuitive that they should not be even marked as DF, some rules are tagged as *wishful thinking fallacies* (WTF).

The final element is a set of additional constraints on the applicability of the rule. In case of (1), this set is empty, but it may involve references to word classes or relations between words, among other constraints, as in the following rule, which relates for example *a team of European astronomers* to *a team of astronomers from Europe* (Bentivogli *et al.* 2010: 3544). This rule contains constraints on parts of speech of all nodes and on the lexical relation of demonymy between nodes X and Z:

(4) $\langle$ X Y.L, Y.L from Z, $\langle +, +, +, + \rangle$, S, {ADJ(X), NOUN(Y), PROPN(Z), DEMONYM(X,Z)} $\rangle$

This rule defines a full equivalence, valid in both types of context and in either direction, hence the four pluses in the signature. Note also that some nodes in the LHS and RHS are not adorned with either L or G – they are assumed to have no further dependents (this effectively becomes another condition on the applicability of this rule).

We do not propose a subformalism for expressing constraints, but acknowledge that much of the expressiveness of the proposed formalism for textual rules lies in the expressiveness of the constraint subformalism. Consider the antepenultimate entailment step of Figure 1, where the relevant rule is alluded to as *distributing shared dependent.* The rule may be formalised as in (5), where D stands for any dependency relation, and an example of its use is given in Figure 4.

(5) $\langle$ S.L X.L C Y.L, S.L X.L C S.L Y.L, $\langle +, +, +, + \rangle$, *DF*, {} $\rangle$

Harnold Lamb penned numerous biographies and history books .

$\Downarrow$

Harnold Lamb penned numerous biographies and numerous history books .
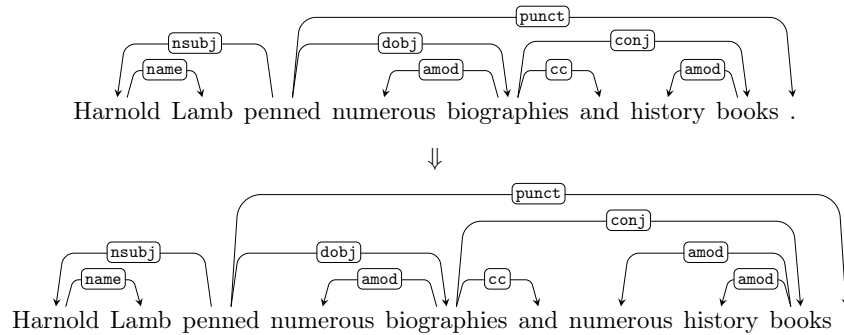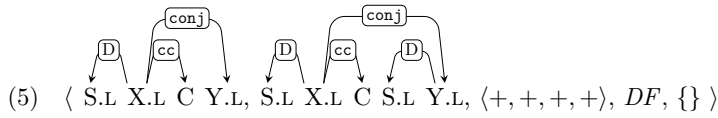
Figure 4: An example of the operation of rule (5)

This rule is marked as defeasible, as constructions such as *numerous biographies and history books* are infamously ambiguous (here *numerous* may refer to both conjuncts, *biographies* and *history books*, or just to the first conjunct, *biographies*), and the rule is justified only in one (perhaps the more natural one) of the two readings. However, if the texts are parsed and disambiguated semantically, a constraint could be added that refers to the disambiguated representation of LHS and makes sure that S indeed scopes over both X and Y – in such a case, the rule could be marked as strong.

So far we have only seen rules containing '+' and '∘' in their polarity signatures. Two examples of rules containing '−' are (6)–(7).[9]

(6) $\langle$ U.L X.G to Y.L, U.L Y.L, $\langle -, -, \circ, \circ \rangle$, *S*, {LEMMA(X) $\in$ {FORGET, FAIL, $\dots$}, $\dots$} $\rangle$

(7) $\langle$ U.L X.G to Y.L, U.L Y.L, $\langle \circ, -, \circ, \circ \rangle$, *S*, {LEMMA(X) $\in$ {HESITATE, $\dots$}, $\dots$} $\rangle$
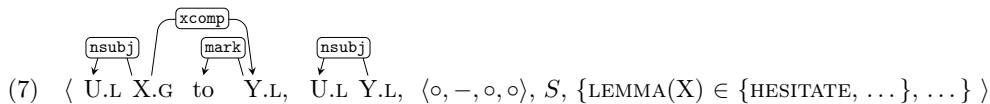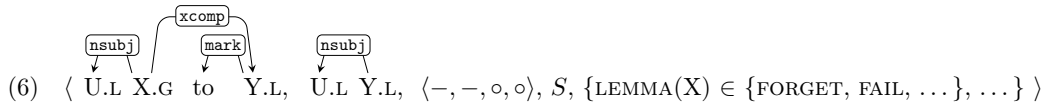
For simplicity, we assume that such rules are unidirectional, i.e., that the converse is not applicable (hence the final two '∘' characters in polarity signatures). The former of these rules targets constructions headed by negative two-way implicatives such as FORGET or FAIL: saying that John forgot or failed to come implies that John didn't come, and saying that Mary didn't forget or fail to come implies the Mary indeed came. On the other hand, the latter rule is applicable only in negative polarity contexts: saying John didn't hesitate to come implies that John came, but saying that Mary hesitated to come doesn't imply whether Mary came or not.

Such rules illustrate a certain subtlety when it comes to the definition of context for the purpose of judging its polarity and to reversing the polarity of the output. Consider the following sentences as the intended input and output of rule (7):

---

[9]These rules are much simplified. Ellipsis signals the omission of the lemmata of other relevant implicatives and of other constraints, including those taking care of appropriate forms of verbs X and Y.

(8)   Nobody hesitated to come.
(9)   Everybody came.

The polarity context of the input is clearly negative, but this negative force does not come from the context in which the rule is applied, as it would in case of *It's not the case that John hesitated to come*, where negation is expressed outside the match of LHS, and also in case of *John didn't hesitate to come*, where negation is expressed in the unspecified part of the subtree headed by X. Instead, in (8), negation is introduced by the negative quantifier in the subject position, i.e. matched by the variable U of the LHS of this rule. Hence, the polarity of the context should be understood as the relative polarity of the head of the match, and not as the relative polarity of the whole match: an expression reversing the polarity may be part of the match, as in (9).

What the input–output pair (8)–(9) also illustrates is the convenience of understanding the polarity reversing effect of such rules rather broadly. The narrowest interpretation of this effect would simply add negation to the output, as in *Nobody didn't come*, and perhaps another rule would transform this sentence into (9). Similarly, perhaps the intermediate *I didn't not get high* could be used to get from (10) to (11). But for the purpose of this paper we assume that polarity reversal expressed by '−' is powerful enough to get in one step from the input to the output in both (8)–(9) (via rule (7)) and (10)–(11) (via rule (6)).

(10) I forgot not to get high.
(11) I got high.

Let us finally note that the simple picture where all rules are rewriting rules, replacing one subtree with another (perhaps empty, effectively pruning the tree), cannot be maintained. As an example, take the following pair (id=44) from the RTE-3: dataset:

(12) T: A former employee of the company, David Vance of South Portland, said Hooper spent a lot of time on the road, often meeting with customers between Portland and Kittery.
(13) H: David Vance lives in South Portland.

Here, the noun phrase (NP) *David Vance of South Portland* must be transformed into the sentence *David Vance lives in South Portland*. But of course the resulting sentence cannot simply replace the original NP – this would result in ungrammaticality. Neither can it simply replace the minimal sentence containing this NP: *It's not the case that David Vance of South Portland arrived* should not be replaced by *It's not the case that David Vance lives in South Portland*. Finally, the output of the rule should not replace the whole matrix sentence, as it may contain other similar bits of information to be used in the reasoning (e.g. *David Vance of South Portland met Hooper of North Dakota*). Instead, a second type of rules is needed which transform the input text by adding a new sentence. We currently assume that such rules are strictly unidirectional and are only used to extract conventional implicatures, which are independent of the polarity of context, hence, they may be formalised as quadruples ⟨LHS, RHS, S, C⟩, where RHS is interpreted as being added to the text rather than replacing LHS.

A simplified example of such a rule for the above pair is given in (14), and the complete reasoning – in Figure 5.
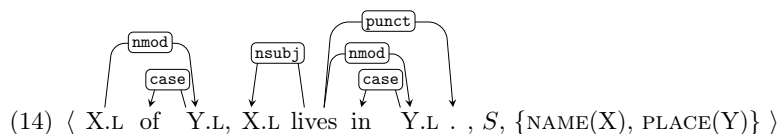
(14) ⟨ X.L of Y.L, X.L lives in Y.L . , S, {NAME(X), PLACE(Y)} ⟩

T: A former employee of the company, David Vance of South
       Portland, said Hooper spent a lot of time on the road, often
       meeting with customers between Portland and Kittery.

     → A former employee of the company, David Vance of South   (rule (14))
       Portland, said Hooper spent a lot of time on the road, of-
       ten meeting with customers between Portland and Kittery.
       David Vance lives in South Portland.

H: → David Vance lives in South Portland.   (dropping sentence)

Figure 5: Entailment steps from text T to hypothesis H for the RTE-3 pair id=44

4   Towards a taxonomy

Textual entailment rules are very diverse and involve various kinds of knowledge: lexical, syn-
tactic, semantic, pragmatic and world knowledge, including the awareness of common scripts or
frames (in the sense of Artificial Intelligence), so there is probably no single most natural way
to organise them in a taxonomical hierarchy. Any attempt at such a taxonomy should be made
with the view of its usefulness for the evaluation of semantic systems. For example, it makes
sense to group rules reflecting phenomena such as coreference, zero anaphora and ellipsis, as
a system which deals poorly on entailment pairs involving such rules probably lacks a reason-
able discourse component. Similarly, grouping rules referring to lexical synonymy, hyperonymy,
meronymy, etc., helps evaluate the use of wordnet-like knowledge sources by the system.

   At the highest level, the proposed taxonomy distinguishes:

1. rules which do not refer to co(n)text, e.g. those discussed in §3 and almost all of the rules
   alluded to in Figures 1–2,
2. rules referring to the *cotext* of a given constructions, i.e. to bits of text not structurally
   related to this construction; e.g. to an earlier noun phrase (NP) coreferent with a given
   NP, as in the first rule of Figure 2, or to a bit of text corresponding to the missing part of
   an elliptical construction,
3. rules referring to the non-textual *context*, e.g. to the date of publication of the text, as in
   the textual entailment from *John Coltrane died in 1967* to *John Coltrane died almost 50
   years ago* in case of texts published in 2015.

   Within each class, rules are grouped according to how much lexical knowledge they require:

a. rules reflecting lexical equivalence relations: not only the usual synonymy (as in wordnets
   or thesauri), but also (near-)synonymous diathesis,
b. rules reflecting various kinds of lexical implication: hyperonymy, meronymy, etc.,
c. rules reflecting the implication of predicate dependents of factive and implicational verbs
   (as in Nairn *et al.* 2006),
d. constructional rules involving some lexical knowledge, e.g. rules allowing for the dropping
   of intersective or subsective modifiers (requires the knowledge of which lexemes may act
   as such modifiers),
e. constructional equivalence correspondences: variations in word order, distribution of de-
   pendents of a coordination over conjuncts (e.g. from *nice boy and girl* to *nice boy and nice
   girl* in languages in which the adjective on the left-hand side is in the plural and the ad-
   jectives on the right-hand side – in the singular), etc.; also rules which add new sentences
   on the basis of existing ones, including rules which extract relative clauses into standalone
   sentences (e.g. from *. . . Coltrane, who died in 1967,. . .* to *Coltrane died in 1967.*),

    f. constructional implication relations: pruning of whole sentences, removing parentheticals, removing various kinds of modifiers, etc.

Additionally, a few groups of rules correspond to such subsystems of reasoning as temporal reasoning (e.g. from *Coltrane died on 17 July 1967* to *Coltrane died in July 1967*), numerical reasoning (further to *Coltrane died almost 50 years ago*; $2015 - 1967 = 48 = 50 - \epsilon$), etc.

At the time of submitting this paper to the proceedings, the complete taxonomy is not yet fully stable, so we do not provide it here. It should be emphasised, however, that while particular rules are often language-dependent, the taxonomy is constructed in a way that maximises its language-independence. This should facilitate the qualitative comparison of RTE systems for different languages.

5 Conclusion

The research reported here builds incrementally on a wide range of previous work in NLP (the RTE task) and linguistics, with the following novel features: 1) an extended and further formalised format of textual entailment rules, with multiple specific rules encoded, 2) a principled taxonomy of kinds of textual entailment (in place of previous more or less unordered collections of selected entailment types), 3) an entailment corpus, still at the initial stage of development, but already containing over 120 pairs fully tagged with almost 800 atomic entailment steps. We expect the corpus to be the most valuable result in practice, as it will not only lead to a qualitative evaluation of semantic NLP systems, but will also make it possible to construct training corpora – consisting of particular kinds of entailment steps – specialised for particular types of reasoning, as postulated already in Bentivogli *et al.* 2010.

References

Bentivogli, L., Dagan, I., Dang, H. T., Giampiccolo, D., and Magnini, B. (2009). The fifth PASCAL Recognizing Textual Entailment challenge. In *Proceedings of the Second Text Analysis Conference (TAC 2009)*.

Bentivogli, L., Cabrio, E., Dagan, I., Giampiccolo, D., Leggio, M. L., and Magnini, B. (2010). Building textual entailment specialized data sets: a methodology for isolating linguistic phenomena relevant to inference. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC 2010*, pp. 3542–3549, Valletta, Malta. ELRA.

Bos, J. (2008). Let's not argue about semantics. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC 2008*, pp. 2835–2840, Marrakech. ELRA.

Clark, P., Harrison, P., Thompson, J., Murray, W., Hobbs, J., and Fellbaum, C. (2007). On the role of lexical and world knowledge in RTE3. In *Proceedings of the ACL–PASCAL Workshop on Textual Entailment and Paraphrasing at ACL 2007*, pp. 54–59, Prague.

Cooper, R., Crouch, D., van Eijck, J., Fox, C., van Genabith, J., Jaspars, J., Kamp, H., Milward, D., Pinkal, M., Poesio, M., and Pulman, S. (1996). FraCaS: Using the framework. Deliverable D16, University of Edinburgh.

Dagan, I., Glickman, O., and Magnini, B. (2006). The PASCAL Recognising Textual Entailment challenge. In J. Quiñonero-Candela, I. Dagan, B. Magnini, and F. d'Alché Buc, eds., *Machine Learning Challenges: Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment. Proceedings of the First PASCAL Machine Learning Challenges Workshop (MLCW 2005)*, pp. 177–190, Berlin. Springer-Verlag.

Dagan, I., Roth, D., Sammons, M., and Zanzotto, F. M. (2013). *Recognizing Textual Entailment: Models and Applications*. Morgan & Claypool.

de Marneffe, M.-C., Dozat, T., adn Katri Haverinen, N. S., Ginter, F., Nivre, J., and Manning, C. (2014). Universal Stanford Dependencies: A cross-linguistic typology. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, eds., *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pp. 4585–4592, Reykjavík, Iceland. ELRA.

Garoufi, K. (2007). *Towards a Better Understanding of Applied Textual Entailment: Annotation and Evaluation of the RTE-2 Dataset*. Master's thesis, Universität des Saarlandes.

Giampiccolo, D., Magnini, B., Dagan, I., and Dolan, B. (2007). The third PASCAL Recognizing Textual Entailment challenge. In *Proceedings of the ACL–PASCAL Workshop on Textual Entailment and Paraphrasing at ACL 2007*, pp. 1–9, Prague.

LoBue, P. and Yates, A. (2011). Types of common-sense knowledge needed for recognizing textual entailment. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 329–334, Portland, OR.

MacCartney, B. and Manning, C. D. (2007). Natural logic for textual inference. In *Proceedings of the ACL–PASCAL Workshop on Textual Entailment and Paraphrasing at ACL 2007*, pp. 193–200, Prague.

Mel'čuk, I. (1981). Meaning-Text Models: A recent trend in Soviet linguistics. *Annual Review of Anthropology*, **10**, 27–62.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, NY.

Nairn, R., Condoravdi, C., and Karttunen, L. (2006). Computing relative polarity for textual inference. In *Proceedings of the Fifth International Workshop on Inference in Computational Semantics (ICoS-5)*, Sydney, Australia.

Przepiórkowski, A. (2015). Towards a linguistically-oriented textual entailment test-suite for Polish based on the semantic syntax approach. *Cognitive Studies / Études Cognitives*, **15**. To appear.

Sammons, M. (2015). Recognizing textual entailment. In S. Lappin and C. Fox, eds., *The Handbook of Contemporary Semantic Theory*, pp. 523–558. Wiley-Blackwell, 2nd edition.

Sammons, M., Vydiswaran, V., and Roth, D. (2010). "Ask not what textual entailment can do for you...". In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1199–1208, Uppsala, Sweden.

Sgall, P., Hajičová, E., and Panevová, J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.

Szpektor, I., Shnarch, E., and Dagan, I. (2007). Instance-based evaluation of entailment rule acquisition. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 456–463, Prague.

Tesnière, L. (1959). *Éléments de Syntaxe Structurale*. Klincksieck, Paris.

Tesnière, L. (2015). *Elements of Structural Syntax*. John Benjamins, Amsterdam.

Toledo, A., Katrenko, S., Alexandropoulou, S., Klockmann, H., Stern, A., Dagan, I., and Winter, Y. (2012). Semantic annotation for textual entailment recognition. In I. Batyrshin and M. G. Mendoza, eds., *Advances in Computational Intelligence: 11th Mexican International Conference on Artificial Intelligence, MICAI 2012San Luis Potosí, Mexico, October 27 – November 4, 2012*, pp. 12–25. Springer-Verlag.

Zaenen, A., Karttunen, L., and Crouch, R. (2005). Local textual inference: Can it be defined or circumscribed? In *Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, pp. 31–36, Ann Arbor, Michigan.

Three types of indefinites: evidence from Ga (Kwa)

Agata Renans

Universität Potsdam
Potsdam, Germany
renans@uni-potsdam.de

Abstract

There is an ongoing discussion whether wide-scope indefinites denote choice functions that are existentially bound [7, 11, 14] or remain free [6]. Data from Ga, an under-researched language spoken in Ghana, show that there are wide-scope indefinites denoting existentially bound skolemized choice functions whose parameter is bound by a higher quantificational NP, free skolemized choice functions with the speaker or a higher quantificational NP as a parameter, and narrow scope quantificational indefinites. Thus the data show that both existentially bound and free skolemized choice functions are attested in natural language shedding new light on the semantics of indefinites.

1 Introduction

In the history of formal semantics, indefinites, whose small subset is exemplified in (1), have obtained many different analyses, e.g., quantificational [10, 1, 4] or dynamic [3, 5].

(1)   Kofi read **a/some/one** book.

Moreover, in the recent literature, indefinites were analyzed as denoting choice functions. For example, [11] and [14] analyzed indefinites as existentially bound choice functions. By contrast, under the analysis of [6], choice functions denoted by indefinites remain free. This discrepancy lead to the still ongoing discussion whether the choice functions denoted by indefinites should be existentially bound or not.[1] In this paper, I argue that both free and existentially bound (skolemized) choice functions are attested in natural language.

The empirical focus of this paper is put on indefinites in Ga, i.e., two indefinite determiners *ko* and *kome* and bare NPs, which show non-homogeneous scopal properties with respect to negation and quantifiers. Based on these interactions, I argue that bare NPs can be properly analyzed as quantifiers, *ko* denotes an existentially bound skolemized choice function whose parameter is bound by a higher quantificational NP, if available, and *kome* denotes a free skolemized choice function, whose parameter can be bound either by the speaker or a higher quantificational NP. The structure of this paper is as follows. In section 2, I provide evidence that bare NPs, *ko*, and *kome* are interpreted as indefinites and I illustrate their scopal properties. Subsequently, section 3 presents the analysis of *ko* and *kome* as denoting different kinds of choice functions. Section 4 discusses some open issues and section 5 concludes.

2 Indefinites in Ga

Ga is an under-researched Ghanaian language spoken in The Greater Accra Region by ca. 600,000 speakers. It is an SVO language with two tones: high and low. All the data, unless written otherwise, stem from the author's original fieldwork in Accra with three Ga native

---

[1]For a discussion of problems of both types of analyses, see [2], [12] and [13].

speakers using the field research methodologies presented in [9]. All the language consultants were students at the time of conducting the fieldwork. None of them has a background in linguistics.

2.1 Bare NPs, *ko*, and *kome* are indefinites

Before I discuss the scopal properties of the bare NPs, *ko*, and *kome*, let me first provide empirical evidence that all of them are indeed indefinites. Below, I present the results of applying some diagnostics for detecting indefinites, whose design is based on the tests presented in [7].

The result of the diagnostic demonstrated in (2) shows that *bare NPs*, *NP ko*, and *NP kome*, as English indefinite determiners, can refer to two different discourse referents in a sentence. By contrast, if they were definite, they would refer to the same entities in both clauses leading to a pragmatically odd structure:[2]

(2)   context: Shikatoohe ∅/**ko**/**kome**   yɛ   Osu...
                bank   INDF/INDF/INDF be.at Osu
                'A bank is in Osu...'
        ...ni shikatoohe ∅/**ko**/**kome**   yɛ   Jamestown.
        and bank   INDF/INDF/INDF be.at Jamestown
        '...and a bank is in Jamestown.'
        #'The bank is in Osu and the bank is in Jamestown.'

Further, it turns out that unlike definite determiners, they can be used in contexts in which the discourse referent is not unique, as illustrated in (3):

(3)   context: There is a tree outside the window. There are three birds on the tree.
        Gbekɛ biihii lɛ   fɛɛ na looflo ∅/**ko**/**kome**.
        child boys DET ALL see bird   INDF/INDF/INDF
        'All the boys saw a bird.'
        #'All the boys saw the bird.'

Finally, the test illustrated in (4) is based on the observation that NPs associated with the *wh*-remnant in sluicing constructions cannot be definite (see [7] and references there). The fact that *bare NPs, ko*, and *kome* are acceptable in sluicing constructions suggest that they are indefinites.

(4)   John mii-tawo   wolo ∅/**ko**/**kome**,   shi mi-le   tenoniji.
        John PROG-look.for book INDF/INDF/INDF but 1SG-not.know which
        'John is looking for a book but I do not know which.'
        #'John is looking for the book but I do not know which.'

2.2 Scopal properties of *ko* and *kome*

Interestingly, bare NPs, *ko*, and *kome* exhibit non-homogeneous scopal properties with respect to various operators, e.g., negation and quantifiers. First, it turns out that whereas *ko* can take

---

[2]The glosses used in this paper are as follows: INDF = indefinite; DET = determiner; SG = singular; PL = plural; 3 = Third person; PRT = particle; NEG = negation; REL = relativizer; PFV = perfective; PROSP = prospective. An example marked with '#' means that the example was judged to be unacceptable in the given context and I hypothesize that it is for semantics or pragmatic reasons. Examples without any diacritics were judged as acceptable in the given context.

both wide and narrow scope with respect to negation, bare NPs can take only narrow scope, and *kome* only wide scope. Consider (5). Since the context specifies that Kofi bought a lot of fish, it clashes with the narrow-scope interpretation which would lead to the meaning that Kofi didn't buy any fish.

(5)   WIDE-SCOPE INTERPRETATION
        context: Kofi bought a lot of fish, but
        E-he-ko   loo #∅/**ko/kome**
        3SG-buy-PFV.NEG fish INDF/INDF/INDF
        'He didn't buy a certain fish.'

The context in (6), on the other hand, specifies a narrow-scope interpretation. Since it is known that Kofi didn't buy any fish, a wide scope interpretation which says that Kofi didn't buy a certain fish is ruled-out.

(6)   NARROW-SCOPE INTERPRETATION
        context: Kofi went to the market yesterday. He bought vegetables, shoes, and toys but he didn't buy any fish.
        Kofi he-ko   loo ∅/**ko/#kome**.
        Kofi buy-PFV.NEG fish INDF/INDF/INDF
        'Kofi didn't buy any fish.'

Second, whereas *kome* can obtain both a constant and a covarying interpretation with respect to quantifiers, as suggested by its acceptability in the context of (7) and (8), bare NPs and *ko* can get only a covarying interpretation, as suggested by their acceptability in the context of (8) but not in the context of (7):

(7)   CONSTANT INTERPRETATION
        context: There were four women in the library. It looked really funny because all of them were reading one book.
        Yei   lɛ   fɛɛ   kane wolo #∅/#**ko/kome**.
        women DET every read book INDF/INDF/INDF
        'Every women read some book.'

(8)   COVARYING INTERPRETATION
        context: When I came to the library yesterday, four women were reading a book. Each of them was reading a different book.
        Yei   lɛ   fɛɛ   kane wolo ∅/**ko/kome**.
        women DEF every read book INDF/INDF/INDF
        'Every women read some book.'

In addition, while both *ko* and *kome* can obtain an intermediate scope interpretation, bare NPs cannot. For example, the intermediate scope interpretation of (9) is the one in which most linguists chose one problem to work on but the choice of the problem varies with the linguist.

(9)   context: Four linguists chose one linguistic problem to work on. Linguist 1 chose the syntax of Ga, linguist 2 chose the syntax of Akan, linguist 3 chose the phonology of Ewe, linguist 4 chose the morphology of Avatime. Linguists 1, 2, and 3, but not 4, read all the analyses solving the respective problem.
        Otsiamii pii   ekwɛ   susumɔi saji   fɛɛ   ni   yeɔ boa sane
        linguist   most have.looked analysis analysis every that help solve problem

#∅/**kome/ko**   naaboamɔ.
INDF/INDF/INDF solve
'Most linguist have looked at every analysis that solves some problem.'

The discussed scopal properties of bare NPs, *ko*, and *kome* are summarized in Table 1.

Table 1: Scopal properties of *ko* and *kome*

	intermediate scope	negation	quantifiers
bare NPs (∅)	–	narrow	covarying
ko	✓	wide, narrow	covarying
kome	✓	wide	constant, covarying

3   Analysis

Based on the data presented in section 2, I argue that whereas bare NPs can obtain a quantificational analysis, both *ko* and *kome* denote skolemized choice functions. The former is empirically motivated by the observation that bare NPs cannot obtain an intermediate scope interpretation in (9), suggesting that they cannot move out of syntactic islands. The strongest evidence for the latter is the availability of intermediate scope readings [11, 14, 6, 7, 2]. For example, *ko* and *kome* indefinites in (9) can be neither quantificational (because quantifiers cannot move out of islands) nor referential (because the linguists in (9) choose different problems to solve). In the next subsection, I will explicate the semantics of *ko* and *kome*, leaving a more precise discussion of the semantics of bare NPs for a further occasion.

3.1   Skolemized choice functions

A choice function (CF) is a function of type $\langle\langle e,t\rangle, e\rangle$ which takes a set as its argument and returns one element from that set:

(10)   A **choice function** is a function from sets of individuals that picks a unique individual from any non-empty set in its domain. [6]

Choice functions can be existentially bound [11, 14, 7] or remain free [6, 8]. For example, under the analysis of the indefintie determiner *a* as denoting an existentially bound CF, the sentence in (11) obtains the interpretation in (11-a) and under the analysis of *a* as denoting a free CF in (11-b), where f is a variable ranging over choice functions. In the latter, the value of the CF is provided by the context, which usually is the one intended by the speaker:

(11)   Kofi read a book.
         a. $\exists f($Kofi read $f($book$))$
         ≈ There is a way of choosing a book such that Kofi read this book.
         b. Kofi read $f($book$)$
         ≈ Kofi read a book chosen in a way known to the speaker.

   Choice functions can also be skolemized, i.e., they can take an additional covert pronominal index (also called the parameter or the skolem index). The index, as overt pronouns, can be either interpreted with respect to the assignment function, usually relativized to the speaker, or

can be bound by a higher quantificational NP. The former leads to the wide scope interpretation in (12-a) and the latter triggers a narrow scope interpretation in (12-b):

(12)   Every student read a book.
         a. every student read f_1(book)
         $\approx$ I know a way of choosing a book such that every student read the book chosen that way
         b. every student$_z$ read f_z(book)
         $\approx$ every student read a book chosen in a way relative to every student

Analyzing indefinites as denoting skolemized choice functions can account for the intermediate scope interpretation. For example, the intermediate reading of (9) in the Kratzer's approach [6] to CF is as in (13):

(13)   for most x[linguist(x) $\to$ $\forall z$[analysis(z) $\wedge$ z solves f_x(problem) $\to$ x looked at z]]
         $\approx$ For most linguists, there is a way of choosing a problem such that they have looked at every analysis that solves that problem.

As it has been already written, I argue that both *ko* and *kome* denote skolemized choice functions.The question is whether they are bound or free, and what are the possible binders of their parameters. I argue that the answers for these questions are provided by different scopal properties of *ko* and *kome* presented in subsection 2.2.

Interaction with negation.   I argue that in a semantic fieldwork situation, interaction with negation is a good test method for determining whether a CF denoted by an indefinite is existentially bound or not. Crucially, a narrow scope interpretation with respect to negation is only possible if a CF denoted by an indefinite is existentially bound. Otherwise, only a wide scope interpretation is available. Consider (14):

(14)   Kofi didn't buy any fish.

With an existentially bound CF both a wide scope interpretation, as in (15-a), and a narrow scope interpretation, as in (15-b), is possible, because negation can scope above or below existential closure:[3]

(15)   a. $\exists f[\neg$buy[Kofi, f_1(fish)]] $\approx$ As for the speaker, there is a way of choosing a fish such that Kofi didn't buy it (there is a fish that Kofi didn't buy.)
         b. $\neg \exists f$[buy[Kofi, f_1(fish)]] $\approx$ As for the speaker, there is no way of choosing a fish such that Kofi bought it (Kofi didn't buy any fish.)

With a free CF, on the other hand, only a wide scope interpretation is possible, because there is no other operator that negation could scope over:

(16) $\neg$buy[Kofi, f_1(fish)] $\approx$ The speaker knows a way of choosing a fish such that Kofi didn't buy it (there is a fish that Kofi didn't buy.)

Since *ko* can get both a wide and a narrow scope interpretation with respect to negation, I argue that it denotes an existentially bound skolemized CF. By contrast, since *kome* can only

---

[3]I argue that both *ko* and *kome* denote skolemized CFs. Since there is no binder that could bind the index in (15), its value is provided by the context, i.e., by the assignment function. Even though in the case of (b) the CF is bound by the context, due to the presence of existential closure it does not obtain a wide scope interpretation.

get a wide scope interpretation with respect to negation, it denotes a free skolemized choice function.

Interaction with quantifiers. The interaction with quantifiers, on the other hand, can detect possible parameter (skolem index) binders. If an indefinite has the speaker as its parameter, it gets a constant interpretation. Conversely, if the parameter is bound by a higher quantificational NP, an indefinite gets a covarying interpretation. Consider (17):

(17)   Every woman read some book.

The representations of (17) given in (18) illustrate what happens when the parameter is bound by the speaker. In both cases, i.e., when the CF is existentially closed or remains free, the indefinite obtains a constant interpretation:

(18)   a. $\exists f \forall z[\mathrm{woman}(z) \rightarrow \mathrm{read}(z, f_1(\mathrm{book}))]$
        $\approx$ As for the speaker, there is a way of choosing a book such that every woman read a book chosen that way.
        b. $\forall z[\mathrm{woman}(z) \rightarrow \mathrm{read}(z, f_1(\mathrm{book}))]$
        $\approx$ The speaker knows a way of choosing a book such that every woman read a book chosen that way

By contrast, if the parameter is bound by a higher quantificational NP, then the indefinite invariably obtains a covarying interpretation. Note that in the case of indefinites denoting an existentially bound CF, it does not matter whether $\exists$ scopes over $\forall$ or vice versa: since in both cases the parameter is bound by the higher quantificational NP, both give rise to the same truth conditions.

(19)   a. $\exists f \forall z[\mathrm{woman}(z) \rightarrow \mathrm{read}(z, f_z(\mathrm{book}))]$
        $\approx$ There is a way of choosing a book relative to every woman such that she read a book chosen that way.
        b. $\forall z[\mathrm{woman}(z) \rightarrow \mathrm{read}(z, f_z(\mathrm{book}))]$
        $\approx$ For every woman there is a way of choosing a book such that she read a book chosen that way.

In subsection 2.2, it was shown that whereas *kome* can get both a constant and a covarying interpretation, *ko* can only get a covarying interpretation. Building on these data and the observations presented in (18) and (19), I argue that while the parameter of *kome* can be bound either by the speaker or a higher quantificational NP, the parameter of *ko* can only be bound by a higher quantificational NP. Putting all the elements together, *kome* denotes a free skolemized CF whose pronominal parameter can be bound either by the context or by a wider scope quantificational NP. *Ko*, on the other hand, denotes an existentially bound skolemized CF, whose parameter is bound by a higher quantificational NP (if available).

3.2   *Ko* and *kome* in downward entailing contexts

The way I have set things up makes predictions for the behavior of *ko* and *kome* in downward entailing contexts. In particular, it predicts different scopal behavior of *ko* and *kome* in the contexts of (20) and (25).[4] First, consider (20):

---

[4]The acceptability of the following two example were checked with one Ga native speaker in Berlin.

(20)   context: There were three woman in the literature course: Mrs Smith, Mrs Müller,
         and Mrs Laryea. They were supposed to read three books of their choice. Mrs Smith
         chose 'Anna Karenina,' 'Gone with the wind,' and 'Madame Bovary.' She read 'Anna
         Karenina' and 'Gone with the Wind' but she didn't read Madame Bovary. Mrs Müller
         chose 'The Hobbit,' 'Pride and Prejudice,' and 'Madame Bovary.' She read 'The Hobbit'
         and 'Pride and Prejudice' but she didn't read 'Madame Bovary.' Mrs Laryea chose 'The
         Lord of the Rings,' 'Anna Karenina,' and 'Madame Bovary.' She read 'The Lord of the
         Rings' and 'Anna Karenina' but she didn't read 'Madame Bovary.'
         Jeee yei   lɛ   fɛɛ   kane wolo **#ko/kome**.
         NEG woman.pl DET every read book INDF/INDF
         'Not every woman read a book.'

The context of (20) is presented schematically in (21), where the capital letters stand for books'
titles and the underlined capital letters for the books read by the respective woman:

(21)   Mrs Smith: {AK, GW, MB}
         Mrs Müller: {TH, PP, MB}
         Mrs Laryea: {LR, AK, MB}

The analysis of *ko* as an existentially bound CF with a higher quantificational NP as the
parameter can account for the unacceptability of *ko* in the context of (21). The semi-formal
representation of the target sentence with *ko* is given in (22).[5]

(22) $\neg\exists f$[every woman$_z$ read f_z(book)]

It says that there is no way of choosing a book such that every woman read a book chosen by
'her CF.' Crucially, it is false in the context of (20), because there is a way of choosing a book
such that every woman read a book chosen that way, as illustrated in (23).

(23)   Mrs Smith: {AK, GW, MB} → AK
         Mrs Müller: {TH, PP, MB} → PP
         Mrs Laryea: {LR, AK, MB} → LR

Conversely, the target sentence with *kome* is judged to be acceptable in the context of (20).
Kome denotes a free skolemized CF. When the value of the skolem index of *kome* is relativized
to the speaker, it obtains the following representation:

(24) $\neg$[every woman read f_1(book)]

(24) says that the speaker knows a way of choosing a book such that it is not the case that
every woman read this book. This is true in the context of (20), because it is possible to choose
a book in the relevant way. For example, 'Madame Bovary' is the book that was not read by
every woman (in fact nobody read it). Now, consider (25):

(25)   context: The same as before, but this time Mrs Smith didn't read 'Madamy Bovary',
         Mrs Müller didn't read 'Pride and Prejudice', and Mrs Smith didn't read 'Anna Karen-
         ina.' They read all other books they chose.

---

[5]An open issue is whether Ga native speakers can also obtain an interpretation of (22) with the existential
closure scoping above the negation, which is not excluded by the proposed analysis of *ko*, and why my Ga native
speaker preferred the reading with the existential closure below the negation. I plan to explore this issue in
future research.

Jeee yei   lɛ   fɛɛ   kane wolo **#ko/kome**.
NEG woman.pl DET every read book INDF/INDF
'Not every woman read a book.'

Again, the context is presented schematically in (26), where the capital letters stand for books' titles and the underlined capital letters for the books read by the respective woman:

(26)   Mrs Smith: {AK, GW, MB}
         Mrs Müller: {TH, PP, MB}
         Mrs Laryea: {LR, AK, MB}

The same as in the example discussed before, the target sentence with *ko* is unacceptable in the context of (26). Again, it is judged to be wrong because there is a way of choosing a book, relativized to every woman, such that the respective woman read a book chosen by 'her CF,' contrary to what is suggested by the formal representation of the target sentence with *ko*:

(27) $\neg\exists f$[every woman$_z$ read f_z(book)]

Conversely, the target sentence with *kome* is judged to be acceptable, because *kome* takes invariably a wide scope with respect to negation. Consider (28) with the speaker as the parameter:

(28) $\neg$[every woman read f_1(book)]

It can be paraphrased as: the speaker knows a way of choosing a book such that it is not the case that every woman read this book. This is true in the context of (25), because for example 'Anna Karenina' was not read by every woman.

4   Some open issues

Throughout the paper, I simplified the semantics of *kome* a bit. Namely, it derives from *ekome* 'one' and the cardinality one forms part of its meaning. This claim is based on the data presented in (29) – (31), which show that *kome* can only combine with singular count nouns:

(29)   SINGULAR COUNT NOUN:
         Q: What did Kofi buy yesterday?
         A: Kofi he   adafitswawolo kome nyɛ.
            Kofi buy newspaper   INDF yesterday
            'Kofi bought (one) newspaper yesterday.'

(30)   PLURAL COUNT NOUN:
         Q: What did Kofi buy yesterday?
         A: #Kofi he   adafitswawo-ji kome nyɛ.
            Kofi   buy newspaper-PL INDF yesterday
            'Kofi bought one newspapers yesterday.'

(31)   MASS NOUN:
         Q: What did Lisa buy yesterday?
         A: #Lisa he   fɔ kome nyɛ.
            Lisa   buy oil INDF yesterday
            'Lisa bought (one) oil yesterday.'

I propose modeling the meaning of *kome* as in (32), where *g* is an assignment function determining which CF will be used in a particular context:

(32) $[\![\text{kome}_i]\!]^g = \lambda P_{\langle e,p \rangle} : ([g(i)](P))\,\text{is atomic.}[g(i)](P)$

Unfortunately, I do not have data which would illustrate an interaction of *ko* with different types of common nouns. This definitely should be examined in order to determine whether it is or it is not another dimension in which the semantics of both indefinites differ.

In addition, there are some open issues and data that still need to be accounted for and/or double-checked. In particular, the proposed analysis cannot account for the following data:

(33)   There were four women in the library. Three of them were reading a book, i.e., the first one was reading 'Pride and Prejudice,' the second 'Gone with the Wind,' and the third 'Anna Karenina.' The fourth woman was writing an article, she was not reading any book.
Jeee yei   lɛ   fɛɛ   kane wolo **#ko/#kome**.
NEG woman.pl DET every read book INDF/INDF
'Not every woman read a book.'

The analysis presented so far predicts *kome* to be acceptable in this context, contrary to fact. Note, however, that this sentence was checked only with one Ga native speaker and its unacceptability should be double-checked.

Moreover, I did not discuss in this paper the scopal properties of *ko* and *kome* with respect to if-clauses. The relevant data are presented below:

(34)   E-baa-ŋɔɔ   Mary naa   kɛji onukpa **ko/kome**   ba.
3SG-PROSP-happy1 Mary happy2 if   elder   INDF/INDF come
'Mary will be happy if an elder comes.'
<u>context 1</u>: Mary doesn't know if there are any elders, but...
→ both *ko* and *kome* are acceptable in this context
<u>context 2</u>: There are many elders in the community.
→ both *ko* and *kome* are acceptable in this context
<u>context 3</u>: There are bunch of elders in this community. Mary dislikes most of these elders and doesn't want them to come, but there is a particular elder who she likes and wants her to come.
→ both *ko* and *kome* are unacceptable in this context

Since in order to account for these data, the systematic field research on the semantics of conditionals needs to be conducted, it is left for future research.

5   Summary

Based on novel data from Ga, I argued that both existentially bound skolemized choice functions and free skolemized choice functions are attested in natural language. In particular, it was shown that there are three types of indefinites in Ga: bare NPs, *NP ko*, and *NP kome*. Bare NPs denote quantifiers, *ko* denotes an existentially bound CF that takes a higher quantificational NP as its parameter (if available) and *kome* denotes a free skolemized CF that always can take either the speaker or a higher quantificational NP as its parameter.

References

[1] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.

[2] Gennaro Chierchia. A puzzle about indefinites. In C. Cecchetto, G. Chierchia, and M.T. Guasti, editors, *Semantic Interfaces*, pages 51–89. CSLI, Stanford, 2001.

[3] Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst, 1982.

[4] Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Oxford: Blackwell, 1998.

[5] Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Dordrecht: Kluwer, 1993.

[6] Angelika Kratzer. Scope or pseudoscope? Are there wide-scope indefinites? In Susan Rothstein, editor, *Events and Grammar*, pages 163–196. Kluwer Academic Publishers, 1998.

[7] Lisa Matthewson. On the interpretation of wide-scope indefinites. *Natural Language Semantics*, 7:79–134, 1999.

[8] Lisa Matthewson. Quantification and the nature of crosslinguistic variation. *Natural Language Semantics*, 9:145–189, 2001.

[9] Lisa Matthewson. On the methodology of semantic fieldwork. *International Journal of American Linguistics*, 70:369–415, 2004.

[10] Richard Montague. The proper treatment of quantification in ordinary English. In J. Hintikka, J. Moravcsik, and P. Suppes, editors, *Approaches to Natural Language*, pages 221–242. Dordrecht: Reidel, 1973.

[11] Tanya Reinhart. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy*, 20:335–397, 1997.

[12] Bernhard Schwarz. Two kinds of long-distance indefinites. In Robert van Rooy and Martin Stokhof, editors, *Proceedings of the 13th Amsterdam Colloquium*, pages 192–197, Amsterdam: ICCL, 2001.

[13] Bernhard Schwarz. Long distance indefinites and choice functions. *Language and Linguistic Compass*, 5/12:880–897, 2011.

[14] Yoad Winter. Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy*, 20:399–467, 1997.

Conditional plans and imperatives: A semantics and pragmatics for imperative mood

Craige Roberts

The Ohio State University, Columbus, OH, USA

`roberts.21@osu.edu`

1 Characteristics of imperatives[*]

Of the three grammatical moods which appear to be universally attested in human language, there is now strong consensus about the basic semantics and pragmatics of two—the declarative and the interrogative. But despite the fact that a great deal of progress has been made in the study of the imperative over the past fifteen years, its basic semantics, and even the semantic type of imperative clauses is still a matter for debate; and accordingly, the pragmatic effect on a context of utterance of proffering an imperative clause still requires clarification. Drawing upon that recent work, I propose a semantics and pragmatics for the imperative. I focus on English; but this basic account can readily be extended to cover languages whose imperatives are somewhat more flexible, like the Korean jussive.

That literature makes evident a number of important properties of imperative clauses. They:

a) typically have no subject (a strong cross-linguistic tendency), though they may:
    (1) Eat your soup!
    (2) Johnny, eat your soup!
    (3) Somebody help me up!
    I'll call the entity, typically an agent (but see (6)), to whom an imperative is directed the **target** of the imperative. Note that (3) shows that the target needn't be specific.

b) display evidence of tense and aspect, but always pertain to a present or future time:
    (4) Please have this done by the time I get back.
    (5) [In the short story *The lady or the tiger*, a captive must choose one of two doors, knowing that behind one is a beautiful lady, behind the other a vicious tiger. Silently to himself before opening one of the doors:] Be the lady! [Carl Pollard, p.c.]
    (6) [speaker is unexpectedly taking a friend home for coffee, can't remember what shape the house was in when she left. Silently to herself:] Please don't be a mess!
    (7) Vote tomorrow!
    (8) #Please had this done by last night.

c) may occur embedded. In English this is only as the complement of a verb of saying, and only as directed to the actual addressee:
    (9) John$_i$ said eat his$_i$ share of the chicken. He won't get home til late.
    In (9) the third person *his*, coreferential with the subject *John*, precludes a direct quotation interpretation. In some languages, complement imperatives may have a shifted target, not the actual addressee (Zanutinni et al. 2012).

d) may be explicitly or implicitly conditional:
    (10) If you're hungry, have some cheese and crackers.
    (11) [Army combat instructor to students:] Before you walk into an area where there are lots of high trees, if there might be snipers hiding in the branches, use your flamethrowers to clear away the foliage. [after von Fintel & Iatridou 2003]
    (12) [two crooks planning a robbery:]
         A: What should I do if the cops arrive?

B: Start shooting.
modal subordination interpretation: 'if the cops arrive, start shooting'

e) display a range of flavors, with two main types:

Practical: something the target can do. Only felicitous if so far as the speaker believes it's possible for the target to realize the property denoted by the VP.

commands and prohibitions
(13) [Boss to tardy employee:] Tomorrow get to work on time!
(14) And don't dawdle!

permission
(15) Take your time!
(16) Have a cookie.

suggestion
(17) [To a friend who's been ill:] See if you can take a day off to recuperate.

pleas: see (3) above

advice: speaker may be disinterested
(18)   [Two friends chatting:]
          A: I'm worried that this contractor will put a lien on my property. But the guy's completely unreasonable. I can't talk to him.
          B: Hire an attorney.

instructions/directions
(19)   A: How do I get to Harlem?
          B: Take the A-train.
(20)   To prepare an artichoke, pull out the central leaves and the fuzzy part down to the heart.

warnings
(21) Be careful! There are sharks in the water!

concessives
(22) OK, go to the silly party! See if I give a damn.

Expressive: nothing can be done; either the matter is already settled, or the target isn't in a position to do anything about it. Grounded in the wishes, desires, etc. of the speaker.

wishes: see (5), (6) above.
(23) Enjoy the movie! (Kaufmann 2012)

f) are closely related to deontic modal statements, in that they:

- permit valid inference of their deontic modal counterparts, as in the following pairs:
(24) [father to son:]   Finish your homework before you surf the web.
                          You must finish your homework before you surf the web.
(25) [to a friend in trouble:]   Hire an attorney.
                                   You should hire an attorney.

- display constraints on interpretation of sequences of imperatives parallel to those on sequences of modal statements (Portner 2007, his example in (26), modified (27)):
(26) a. Be there at least two hours early.
     b. Then, have a bite to eat.   [odd as permission after an order interpretation of (a)]
(27) a. You must be there at least two hours early.
     b. Then have a bite to eat at that cute little place on the corner. [odd as suggestion after the moral injunction in (a)]

- display similar performative constraints on follow-up to those displayed by *must* but not *should* (Ninan 2005, his examples):
(28) You should go to confession, but you're not going to.
(29) #You must go to confession, but you're not going to.
(30) #Go to confession! You're not going to go to confession.

- display a **Deontic Moore's Paradox** (Kaufmann 2015, her examples): Even if the speaker has no interest in realization of the prejacent, as with concessions or disinterested advice, they commit her to endorsing it in some fashion:
(31) #You should go to Paris, but in fact, I think it is not advisable.

(32) A: How do I get to Harlem?
   B: Take the A-train. #But I don't want you to do this.
- display non-Boolean behavior with disjunction ("Free Choice disjunction"), in some sense entailing both disjuncts:
   (33) Pay the bill online or take it to the gas company.
   (34) You can pay online or at the gas company.
g) presuppose an **Epistemic Uncertainty Condition** (Kaufmann 2012): So far as the speaker knows, there are some future courses of events where the imperative is realized by the target, and others where it is not.
h) cannot be used to make assertions.
i) unlike assertions, are not felicitously subject to judgments of truth or falsity.
   (35) A: How do I get to Harlem?
       B: Take the A-train.
       C: #That's false!
       C′: No, take the number 37 bus.
   (30C′) is not a truth value judgment, but a rejection of B's directions, i.e. a correction of B's proposed answer to A's question.
j) cannot occur with sentential adverbials (36) (Gärtner 2015), unlike deontic modal statements (37) or performatives (38):
   (36) #Unfortunately, go to bed!
   (37) Unfortunately, you must go to bed!
   (38) Unfortunately, I now pronounce you man and wife.
k) display non-Boolean behavior: In addition to the Free Choice phenomena noted above in (f), embedded imperatives cannot occur under the scope (syntactic or semantic) of negation or in the antecedent of a conditional.
l) strongly tend, across languages, to be used with directive illocutionary force, just as indicative mood tends to be used to make assertions, interrogative to pose questions.

2 Previous proposals

Recent work has contributed enormously to our understanding of the semantics and pragmatics of imperative clauses. But problems remain, as we see in the following brief overview of some of these accounts. Space precludes review of important work by Condoravdi & Lauer (2012) and Starr (2013).

2.1 Kaufmann (2012, 2015)

Kaufmann's imperative root clauses denote Kratzerian modal propositions, with an implicit necessity operator relativized to a modal base (given by a contextually understood function f that takes the world of evaluation to yield a set of propositions) and ordering source (set of ideals— e.g. rules or laws, wishes or wants, best outcomes, etc.—captured by a function g). The functions f and g may have many different "flavors", so that choice of f and g can account for why one and the same imperative, e.g. *Take a taxi!*, can be an order (from the boss), a suggestion (from a helpful friend), or a plea (from one's worried husband), etc., with relevant, contextually given variations on each of those types. This permits Kaufmann to beautifully capture characteristics (c) – (f) above. For example, in Kratzer's modal semantics, conditionals are just modal statements with an extra, explicit premise, expressed by the *if*-clause, which is added to the modal base; since imperative modals use a modal base, we would expect such explicit modification to be possible here as well (d). Kaufmann makes many excellent observations about the presuppositions associated with use of an imperative, and predicts the full range of imperative flavors. But the modal semantics also means that imperatives denote propositions and have truth conditions, failing to satisfactorily explain why they cannot be asserted (h), or why we cannot respond to them directly with *that's true/right*, unlike to the corresponding modal statements (i), or why they

are incompatible with sentential adverbials (j). She does offer an explanation of (h), (i), arguing that imperatives are "performatives", and that indicative performatives are supposedly not asserted or assessed for their truth conditions, either. But first, it isn't clear that performative declaratives are not asserted; there is a long tradition arguing that they *are* assertions, but for pragmatic reasons are simply self-verifying (see Condoravdi & Lauer 2012; Roberts 2015). Second, it seems that only practical directions are performative, not expressives, but the latter also do not license response as to an assertion. This tack also fails to address the infelicity with sentential adverbials (j), especially since those may be acceptable with indicative performatives (38); nor does it explain their non-Boolean behavior (k). Finally, Kaufmann doesn't yet satisfactorily tackle the pragmatics of imperatives (l), despite the useful discussion in Kaufmann (2015). One consequence of this is that many of the presuppositions she attributes to imperatives should follow from general pragmatic principles, given the proper pragmatics (below).

2.2 Portner (2004, 2007)

Portner (2004) takes imperative clauses to denote directed properties (type $\langle s, \langle e, t \rangle \rangle$)— properties which can only be true of the target. He assumes that in the context of utterance there is a record of each interlocutor's To-Do list, the set of evident actions which that interlocutor is publicly committed to doing. The type of speech act canonically associated with use of an (unembedded) imperative is issuance of a direction; and directions, if accepted, are added to the addressee's To-Do list. This account straightforwardly explains why imperatives cannot be used as assertions (h) or take truth judgments (i), why they don't occur with sentential adverbials (which arguably modify propositions), their non-Boolean behavior (k), and their default correlation with Directives (l). Portner (2007) then focuses on explaining a direction's deontic implications (f). But Portner doesn't relativize the interpretation of imperatives to flexible modal parameters, so cannot readily address the wide range of imperative flavors accounted for by Kaufmann (e). He does attempt to capture this flexibility, arguing that there are different types of To-Do lists (e.g., deontic/moral, buletic, and teleological, with an indefinite class of sub-types of each), so that different main-clause imperatives lead to enrichment of different lists. But though this captures some aspects of how different flavors of imperatives are correlated with different modal flavors, it only does so via the pragmatic function of main clause imperatives to update the To-Do list, and then only with respect to the priorities reflected in a relevant "selection function" (related to Kratzerian conversational backgrounds), which both selects the relevant type of To-Do list to which the property is to be added and leads to a corresponding modal update in the Common Ground. This is not entirely satisfactory. For example, he cannot naturally explain why imperatives tend to be (overtly or implicitly) conditional (d), since his account doesn't make the modal base f play a role in the update of the To-Do list itself; the latter involves the simple addition of the property denoted by the main imperative clause. This problem is compounded in the interpretation of embedded imperatives, where both the conditional sense (*f*) and relativization to other priorities (*g*) of the imperative may be conveyed; for example, we could modify (9) to yield *John_i said eat his_i share of the chicken if you're hungry*. Portner (2007:380) assumes that a monster shifts the context of issuance for embedded imperatives in Korean to the one reported in the matrix clause; but since an embedded imperative is not used to issue a direction, it's not clear why or how the pragmatic condition involving the selection function would be supposed to apply in such embeddings; and in any case that doesn't explain the conditional force. This is a strong suit for Kaufmann, who uses Kratzer's f and g in the *semantics* of imperative modals. Also, Portner cannot naturally capture the Expressive imperative uses, since in these uses there is no practical action to undertake to do.

2.3 Charlow (2011)

Charlow (2011) illuminates how imperatives propose modification of a body of preferences associated with the target interlocutor's complex plans, as well as how those plans and associated goals bear on the imperative's interpretation. But to do this, (a) he makes the semantic type of imperatives be that of a function from a body of preferences (roughly, an ordering source) to a

proposition including a necessity modal, and (b) he builds illocutionary force into the semantics of the imperative. E.g., the semantics for a conditional imperative like his (39):

(39) If you're cold, shut the window! (conditional imperative)

proposes the introduction of "a complex planning state in [the target] agent—one represented very roughly, by sequentially pairing facts (relevant contingencies, like the target being cold) with planned outcomes (that the target shut the window)." Because the imperative contains a modal, in principle this type of account can satisfactorily capture most of the same characteristics that Kaufmann does. But the built-in illocutionary force is an important barrier to explaining embedded imperatives. And since the semantics yields a modal proposition, given its preference-set argument, it isn't clear why imperatives cannot occur with sentential adverbials (j).

3 A New Proposal

The present proposal adopts the best features of each of the accounts just reviewed.

3.1 Background: Context of utterance

A context of utterance is a body of information captured on a scoreboard in the sense of Lewis (1979), as developed in Roberts (1996/2012, 2015), given here with new detail about G:

The **scoreboard for a language game** is a tuple, $<I, G, M, <, CG, QUD>$, where:

I, the set of interlocutors at t

$G = \{ G_i \mid i \in I\}$, a set of sets of goals, plans, ideals and priorities in effect at t, where:

for all $i \in I$, there is a (possibly empty) G_i which is the set of i's publicly evident prioritized desiderata, including those goals which i is publicly committed at t to trying to achieve

for all $i \in I$, for all $g \in G_i$, g is a conditional goal $<c,\gamma>$, representing the intention to achieve the target goal γ should the possibly trivial conditions c be realized in the actual world.

for all $G_i \in G$, there are several relations over G_i, including:

Subs$_i$: a pre-order (reflexive, transitive) s.t. Subs$_i$(g,g′) iff g subserves g′ in G_i

Plan$_i$: Plan$_i$($<g,\{g_m,...,g_n\}>$) iff i has a plan to accomplish g via realizing $g_m,...,g_n$, and $\forall g' \in \{g_m,...,g_n\}$: Sub$_i$(g′,g)

$\leq_i$: a partial order (reflexive, antisymmetric, and transitive) s.t. g $\leq_i$ g′ iff g is a higher ranked priority (more ideal) for i than g

and in addition we define:

$G_{com} = \{g \mid \forall i \in I: g \in G_i\}$, the set of the interlocutors' common desiderata at t.

$G_Q = \{g \in G_{com} \mid$ there is some Q$\in$QUD and g is the goal of answering Q$\}$

M, the set of moves made by interlocutors up to t, with distinguished sub-sets:

A $\subseteq$ M, the set of assertions

Q $\subseteq$ M, the set of questions

S $\subseteq$ M, the set of suggestions

Acc $\subseteq$ M, the set of accepted moves

< is a total order on M, the order of utterance

CG, the common ground, the set of propositions treated as if true by all $i \in I$ at t.

The CG reflects all the information on the scoreboard. I.e., if in G_i for some interlocutor i there is the goal of addressing some question or realizing some plan for action, then the fact that i is so committed—that i should realize that goal—is reflected in the CG.

QUD $\subseteq$ Q$\cap$Acc, the ordered set of questions under discussion at t, s.t. for all m$\in$M at t:

a. for all q ∈ Q∩Acc, q ∈ QUD(m) iff CG fails to entail an answer to q and q has not been determined to be practically unanswerable.

b. QUD is (totally) ordered by <.

c. for all q, q' ∈ QUD, if q < q', then the complete answer to q' contextually entails a partial answer to q.

and in addition:

d. for all Q∈QUD there is a g∈G_{com} such that g is the goal of answering Q, and

e. for all Q∈QUD, it is not the case that CG entails an answer to Q

RELEVANCE: Since the QUD reflects the interlocutors' goals at any point in a discourse, in order for an utterance to be rationally cooperative it must address the QUD. Given QUD *q*, a move *m* is **relevant** iff *m* addresses *q*.

An utterance *m* **addresses a question** *q* iff *m* either contextually entails a partial answer to *q* (*m* is an assertion) or is part of a strategy to answer *q* (*m* is a question) or suggests an action to the target which, if carried out, might help to resolve *q* (*m* is a direction).

Since the CG includes all that the interlocutors take to be true, it includes information about the discourse scoreboard as well. The point of the more articulated scoreboard is not so much to *replace* the CG as to clarify the different types of information that interlocutors crucially track in discourse, and the different roles these types of information play in the evolution of felicitous discourse.

In the absence of an evident QUD, we can understand relevance to require that *m* address the interlocutors' immediate, evident goals in a task at hand or other practical problem; in such a case the goal can be understood as addressing a decision problem. Kaufmann (2012,2015) and Kaufmann & Kaufmann (2015) model a decision problem as a kind of question: a partition of the Context Set which represents the answers to the question *What should x_i do?* in given circumstances.

3.2   Semantics for the English imperative

The **semantics** for an imperative yields a conditional, directed property (type <s,<e,t>>). As in Portner, such a property is indexically **directed** to a target agent. In English, the target is always the addressee, in both root and embedded imperative clauses, and the function corresponding to the property is only defined when its target argument is the addressee. In other languages the target of an embedded imperative may be shifted, reminiscent of shifted indexicals (Portner 2004); and in the closely related Korean jussive (Pak et al. 2004) even matrix clauses may be directed to the speaker, yielding a promise. As in Kaufmann (2012), the denotation of an imperative is **conditional** in that it depends upon a Kratzerian modal base *f* and ordering source *g*. But here there is no modal *per se*; instead, *f* and *g* determine the applicable circumstances in which the property should be realized (accessible world/times in a branching future). Thus, instead of truth conditions, imperative clauses have **realization conditions**, spelling out what the world would have to come to be like for the property to count as realized by the addressee to which it's directed, in the applicable circumstances.

Take ‼$_i$[$_S$VP] to be the logical form of an English imperative clause, uttered in context K (the scoreboard, as above), with modal base *f* and ordering source *g*. As in Kratzer, *f* takes a world *w* and time *t* as argument to yield a set of propositions, each a set of worlds. The ordering source *g* then facilitates an ordering of the worlds in which all those propositions are true, ∩f(<w,t>): *g*(<w,t>) also yields a set of propositions—reflecting some relevant ideals—and the worlds in ∩f(<w,t>) are ordered according to how close they come to realizing all those ideal propositions.

We define the **applicable circumstances** for a directed property, relative to *f, g,* and the world and time of issuance <w,t>:

Applic$_{f,g}$(<w,t>) = {<w′,t′> | w′ ∈ f(<w,t>) & ∀w″∈ ∩f(<w,t>) : w′ ≤$_{g(<w,t>)}$ w″ & t′ ≥ t}

The applicable circumstances are those <w′,t′> which are the most ideal present or future circumstances among those in which all the propositions in *f*(<w,t>) are true.

Then an imperative's proffered content is its **realization conditions**, presupposing the target addressee and a modal base and ordering source for the applicable circumstances:

<u>CONVENTIONAL CONTENT of English</u> $!_{K,f,g}[_SVP_i]$: (proffered type $<s,<e,t>>$)
Given context K, with x_i the addressee, t the UT, $<w,t>$ the circumstance of evaluation:
> **Presupposed content**:
>> The addressee x_i is the target to which the proffered content is directed.
>> f is a circumstantial modal base, consistent with the interlocutors' common ground.
>> g is an ordering source that ranks actions relative to the QUD and the interlocutors' goals and plans, and for consistency with overarching goals, priorities and ideals.
>
> **Proffered content**:
>> $\lambda<w',t'> \lambda x: x \in \{x_i\} \cdot <w',t'> \in Applic_{f,g}(<w,t>) \rightarrow x \in [[VP]]^{<w',t'>}$

The circumstance of evaluation $<w,t>$ will be the circumstance of issuance. In matrix clauses, this will be the speech time/world $<w^*,t^*>$; in embedded clauses, the reported eventuality. The domain of λx is restricted to the singleton set containing the target. The imperative is realized in case the target has the property denoted by VP in all the applicable circumstances.

As in Kratzer (1981), an *if*-clause is a modifier of the modal base f, adding the proposition expressed by its clause to the set of propositions $f(<w,t>)$, which are ordered relative to $g(<w,t>)$.

Many of Kaufmann's (2012) presuppositions of imperatives follow on this account, as we will see, from general principles relating information in discourse, or principles pertaining to what it is to rationally plan some action. The relationship to deontics is also pragmatic and is independently motivated in the framework in §3.1 above. In the present account, none of this need be stipulated.

3.3 Pragmatics of imperatives

The canonical use of a root imperative clause is to issue a direction, a natural use given its semantic type (as in Portner 2004). The direction might be intended to address a contextually relevant decision problem ('what should I do?'), satisfy some buletic goal (make the speaker or the target happy), and/or answer a question (A: *Where are my socks?*, B: *Look in the closet.*). The pragmatics of directions is parallel to that of Stalnaker's (1979) for assertions, Roberts' (1996) for questions. These are the three principal kinds of moves in a discourse game, given the scoreboard above, characterized formally as follows, where for constituent κ $|\kappa|^D$ is the interpretation of κ in discourse context D, and the diacritics **.**, **?**, and **!** stand for declarative, interrogative, and imperative mood, respectively:

Assertion: (following Stalnaker 1979)
> If an assertion of $.\alpha$ is accepted by the interlocutors in a discourse D, $|.\alpha|^D$ is added to CG.

Interrogation: (Roberts 1996)
> If a question posed by $?\alpha$ is accepted by the interlocutors in a discourse D, then $|?\alpha|^D$, a set of propositions, is added to the QUD.
> A question is removed from QUD iff its answer is entailed by CG, or it is determined to be practically unanswerable, or it is no longer relevant to some question or goal it subserves in the strategy of inquiry (the super-question or goal has been answered or abandoned).

Direction:
> If a proffered direction $!_iP$ is accepted by target x in context K, containing information G about the evident goals and plans of the interlocutors, then
> (a) **PRACTICAL DIRECTIONS**: if so far as the interlocutors know x can reasonably intend to realize $!_iP$, update x's goals and associated plans in G_x to include the realization, under the applicable circumstances, of $!_iP$.
> (b) **EXPRESSIVE DIRECTIONS**: if so far as the interlocutors know x can*not* reasonably intend to realize $!_iP$, update the speaker y's ideals in G_y to include the realization, under the applicable circumstances, of $!_iP$ by x.
> The realization of $!_iP$ is removed from the interlocutors' ideals in G once it is no longer potentially applicable (it has been realized, or it is determined that it cannot be practically realized) or any over-arching goals and plans it subserves have been realized or abandoned.

If a proffered imperative is accepted, this leads to modification of the publicly evident goals, plans and intentions of the interlocutors, and more generally of their overarching ideals and

desiderata. If the realization conditions of the imperative are in principle actionable, the speaker proposes that, under the understood conditions, it would be ideal if the target found a way to realize the proffered content. When such practical directions are accepted, the target is committed to planning to realize the corresponding conditional goal *should the applicable circumstances obtain*, insofar as it's within her power. Practical directions can also modify the speaker's ideals unless the speaker is understood to be disinterested—cf. commands vs. advice (Kaufmann 2015). Expressive directions are not actionable (Condoravdi & Lauer 2012, Kaufmann 2012): either the matter is already settled (as in (6) above), or there's little or nothing the target can do about it (as in (5), and (23)), and the target may not even be an agent (as in (6)). In such a case the imperative is understood as the expression of the speaker's desires or wishes, an ideal to which she is committed. Thus, how the ideals and intentions of the interlocutors are modified, and whose are modified, is a function of the practicality of the imperative, as well as of other evident intentions.

Semantically, the proposed ideal is typically conditional, as are goals generally (§3.1): This may be explicit, as in (10) – (11), the latter constituting generic advice; or clear from context, as in the modal subordination example in (12), where Relevance to A's question makes the condition evident. And just like deontic modals (Thomason 1984), we always adopt ideals for action relative to a certain kind of *ceteris paribus* assumption: If conditions change, or the realization of the imperative would conflict with other, higher goals or ideals, one may drop the commitment to realizing it. For example, though the Army instructor doesn't say so in (11), if one of the soldiers comes to an area where snipers might be hiding in the branches *but there are children collecting wood directly under the trees*, the regard for innocent lives may override the goal of destroying potential snipers' hiding places. Thus, a practical direction is more than the proposal that the target agent adopt a goal, because (a) that goal is conditional on the applicable circumstances obtaining, and (b) the adoption is not proposed as an isolated matter, but as a revision of the target agent's overall complex structure of plans and intentions, with f and g reflecting the evidently relevant circumstances and priorities (Charlow 2011). Moreover, not all directions are practical, and the above pragmatics, unlike Portner's To-Do list account, admits of expressive directions. This also correctly predicts that there should be embedded expressive directions:

(40) John said to tell you be well while he's gone.

Other felicity conditions imposed on imperatives by Kaufmann (2012, 2015) follow from the pragmatics in the framework described in §3.1, the requirement of relevance of the utterance to the QUD, the nature of a decision problem, and from what it is to adopt an intention to act *in view of such a problem* (Charlow 2011). These include (with rough paraphrases):

- **Authority Condition**: the speaker is an expert on f and g

This is rather too strong for non-commands, a fact acknowledged in Kaufmann (2012, §4.2.2), where it is clear that Authority is intended to predict performativity. But a weaker, more plausible sense in which the cooperative speaker is presupposed to believe that her advice is sound follows from the assumption that the utterance must be relevant to the decision problem addressed, offering successful resolution. Then this is just the imperative counterpart of Gricean Quality.

- **Epistemic Uncertainty Condition**: the speaker holds as possible some future courses of events where the imperative prejacent p comes about and some where $\neg p$ does.

An uninformative response to a question is irrelevant. Kaufmann (2015:fn.27, p.23) notes: ". . .at least in practical contexts, this follows independently from the requirement that the prejacent answer an open decision problem for the addressee."

- **Ordering Source Restriction**: the prejacent either answers a salient decision problem for the hearer (practical), in which case the ordering source g provides the relevant criteria for resolving that problem; or there is no such decision problem (expressive) and g is speaker-bouletic.

Given the pragmatics of directions, adding a goal or priority to G, it follows from acceptance of a direction that one will consider other relevant priorities in G in grasping how the prejacent property is to be integrated. Since adopted actionable goals will be something the sincere, rational agent attempts to achieve *by virtue of what it is to be committed to a goal* (Bratman 1987), if the goal is to serve the resolution of a problem, she must consider the other relevant criteria.

Kaufmann takes practical imperatives to be those that address salient decision problems for the target agent x_i. Kaufmann & Kaufmann assume that it is "a defining characteristic of decision problems that they contain only propositions the agent is able to bring about" (cf. Kaufmann's 2012 *Ability to Act*). In such a case Kaufmann (2012) imposes two other presuppositions:

- **Curious George**: A rational hearer facing a decision problem Δ_c will try to find out whether $\Box^{f,g}p$ for all $p \in \Delta_c$. ['. . .will aim to answer the question of 'what do to'…']
- **Rational Choice**: A rational hearer who believes of some $p \in \Delta_c$ that it is the solution under the relevant criteria f, g will aim to bring about p. ['...and knowing the solution, will aim to realize it']

But again, these just follow from what it is to sincerely adopt a goal; cooperative interlocutors address the QUD (discourse goals) and more generally attempt to achieve their (domain) goals.

If the question is a decision problem, this guarantees that so far as the speaker is concerned, the realization of the proffered content would constitute the optimal solution to the problem. And this plus cooperativity entails that so far as the issuer knows the property can be realized by x_i in a future branch of the actual world of issuance, in order to actually address the problem.

What these conditions capture is that the combination of the circumstances both at the time of utterance (relative power and wishes of the interlocutors, etc.) and in the applicable conditions, plus the interlocutors' understood priorities influence the interpretation of both modal statements and imperatives. And in turn, both deontic modal statements and imperatives influence the agents' understood goals, plans and intentions in G. So CG and G both constrain and are constrained by these prioritizing speech acts, as first argued by Portner (2007).

4 Reviewing the Characteristics of Imperatives

The theory in §3 addresses the problems noted for other accounts while retaining their virtues:

(a) Overt subjects can be addressed by (a) making them optional at LF, and (b) presupposing that any overt subject denote or have as domain a subset of the set of addressees.
(b) Tense restrictions are understood as a function of the definition of applicable conditions.
(c) The illocutionary neutrality of the semantics for imperative clauses permits an account of embedded imperatives, given an appropriate semantics for embedding predicates, and can easily be modified for other, shiftable languages.
(d) As in Kaufmann, the proffered content is relativized to a contextually salient modal base f and ordering source g, explaining the conditional flavor of imperatives, just as in overt modal statements. In modal subordination, as in (12), the modal base is enriched with contextually relevant hypothetical assumptions, just as in indicatives (Roberts 1989).
(e) Different contextually relevant f, g yield a wide range of imperative flavors, as in deontic modals, partly suggested by information in the context of utterance about the interlocutors' relative power relationships, their overarching goals and ideals, the QUD, etc. We can adopt Kaufmann's (2012) detailed working out of the various kinds of modal base and ordering source involved.
(f) Content in G, for all interlocutors, is automatically reflected in CG (§3.1). Hence, as in Portner (2007), all adopted goals (practical directives) and updated ideals (expressives) are reflected in CG as deontic modal propositions, restricted via the same modal base and ordering source used to give the applicable conditions on the adopted goal or ideal. Nothing need be stipulated. Ninan (2005) models the performative aspect of *must* by making it contribute to the addressee's To-Do list, like the imperative; but that is just to say that *must* tells us something about what the addressee's priorities have to be like, leading to obligatory modification of G, unlike *should*, with leaves open options. We can adopt Kaufmann's (2015) account of the Deontic Moore's Paradox. Similarly, whatever account of Free Choice disjunction is suitable in the case of modals can be extended to the imperatives insofar as it depends on the relationship of the modal base to the common ground and of the ordering source to what is known about the speaker's priorities.

(g) Epistemic Uncertainty follows from pragmatic considerations, as discussed in §3.3.

(h) Imperatives do not denote propositions so they cannot be used to make assertions.

(i) Imperatives have no truth conditions, so are not felicitously subject to judgments of truth or falsity. .

(j) Sentential adverbials are inapplicable because they require a propositional argument

(k) Given that imperative clauses don't denote propositions, we don't expect them to behave like propositions in the standard calculus, so that their non-Boolean behavior isn't surprising.

(l) The account lets us capture universals about grammatical mood and the universal default correlations between mood and speech act type; see Portner (2004) and Roberts (2015).

References

Bratman, Michael E. (1987) *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA.

Charlow, Nathan (2011) *Practical language: Its meaning and use*. Ph.D.Thesis, University of Michigan.

Condoravdi, Cleo & Sven Lauer (2012) Imperatives: Meaning and illocutionary function. In C. Piñon (ed.) *Empirical Issues in Syntax and Semantics*, Volume 9, pp. 1–21.

von Fintel, Kai & Sabine Iatridou (2003) Epistemic containment. *Linguistic Inquiry* 34(2):173-98.

Gärtner, Hans-Martin (2015) Root infinitivals and modal particles. An interim report. Ms., HAS-RIL, Budapest.

Kaufmann, Magdalena (2012) *Interpreting Imperatives*. Springer, Studies in Linguistics and Philosophy 88. Revised version of M. Schwager (2006) Ph.D. thesis, University of Frankfurt.

Kaufmann, Magnalena (2015) Fine-tuning natural language imperatives. Ms., UConn.

Kaufmann, Magdalena & Stefan Kaufmann (2012) Epistemic particles and performativity. *Proceedings of SALT 22*:208-225.

Kratzer, Angelika (1981) The notional category of modality. In H. J. Eikmeyer & H. Rieser (eds.), *Words, Worlds and Contexts* (pp. 38--74). Berlin: de Gruyter.

Lewis, David (1979) Scorekeeping in a language game. In R. Baüerle, U. Egli, and A. von Stechow (eds.) *Semantics from a Different Point of View*. Springer, Berlin.

Pak, Miok, Paul Portner & Rafaella Zanuttini (2004) Deriving Clause Types: Focusing on Korean. In *Proceedings of the Linguistic Society of Korea 2004*, Yonsei Institute of Language and Information Studies (pp. 359–368). Seoul: Hanshin Publishing Company.

Portner, Paul (2004) The semantics of imperatives within a theory of clause types. In K.Watanabe & R. Young (eds.) *Proceedings of SALT 14*. CLC Publications.

Portner, Paul (2007) Imperatives and modals. *Natural Language Semantics* 15:351-383.

Roberts, Craige (1989) Modal Subordination and Pronominal Anaphora in Discourse. *Linguistics and Philosophy* 12.6:683-721.Roberts, Craige (1996/2012) Information Structure: Toward an integrated theory of formal pragmatics. Published, with a new afterword, in 2012 in *Semantics and Pragmatics* 5.7:1-19.

Roberts, Craige (2015) Speech acts in discourse context. To appear in a volume on speech acts for Oxford University Press, edited by Daniel Fogal, Daniel Harris & Matt Moss.

Starr, William B. (2013). A Preference Semantics for Imperatives. Ms. Cornell University, URL http://williamstarr.net/research/a_preference_semantics_for_imperatives.pdf.

Thomason, R. H. (1984) Combinations of tense and modality. In D. Gabbay & F. Guenthner (eds.) *Extensions of Classical Logic*, Volume 2 of Handbook of Philosophical Logic, pp. 135–165. D. Reidel.

Zanuttini, Rafaella, Miok Pak & Paul Portner (2012) A syntactic analysis of interpretive restrictions on imperative, promissive and exhortative subjects. *Natural Language and Linguistic Theory* 30.4:1231-1274.

Splitting hairs[*]

Chung-chieh Shan[1]

Indiana University
`ccshan@indiana.edu`

1 Introduction

Besides conjoining an assertion into the common ground (Stalnaker 1978), another kind of move that we often propose and accept in discourse is to change our notion of what is a thing and what is the same thing. For example, depending on how a conversation goes, in the middle of it we might start treating 'a' and 'A' as different letters, while preserving our common knowledge that it is a vowel. (After all, the English alphabet has 26 letters—or is it 52?) Or we might start to distinguish 'a' from 'a', or even 'a' from 'a'. (After all, the word 'lava' contains two vowels—or is it just one?) Similarly in the case of copredication: after agreeing that today 'lunch was delicious but took forever' (Asher 2011), we might start to distinguish the food from the event, in order to clarify how we plan to have 'the same lunch' tomorrow.

This paper begins a study of these moves. They are made by interlocutors such as linguists, meteorologists, and radiologists as they work together to conjure things like grammatical constructions, cold fronts, and tumors from the stuff that is our shared reality. These moves are useful because two interlocutors can agree completely on facts of the matter—be both omniscient, even—yet appear to disagree because they individuate the world into things differently. For example, two linguists may appear to disagree about whether Sita speaks Rama's language, not because they disagree about how Sita and Rama speak, but because they group idiolects into languages differently. And two meteorologists may appear to disagree about how many cold fronts formed in this country today, only because they distinguish cold fronts, countries, or days differently.

I don't know about you, but this kind of apparent disagreement happens all the time around me. When it does, instead of accusing each other of inaccurate perception, we interlocutors can use these moves to change and align our domains of individuals. These moves are essential for productive conversation, because what individuals we should distinguish and how (Aloni 2000) depends on the current context of the conversation, including what we are trying to do together. Indeed, I can refer to a thing without even myself having an ultimate refinement of it in mind ('this letter is beautiful'). And as the examples in the previous paragraph suggest, the moves are not specific to words like English 'same' and 'different' (Beck 2000; Barker 2007).

2 Hair-splitting in action

To give a more formal example, suppose you and I introduce a discourse marker i in conversation and agree to ascribe to it various properties—i is P, and moreover i is Q—by successive updates to the discourse state:

$$\cdots \longrightarrow \boxed{\begin{array}{c} i \\ \hline P(i) \end{array}} \xrightarrow{\;Q(i)\;} \boxed{\begin{array}{c} i \\ \hline P(i) \\ Q(i) \end{array}} \tag{1}$$

Then we turn to the topic of whether i is R and find upon investigation that we need to distinguish between two refinements of i, namely j, which is R, and k, which is not. So we make that distinction—we split the hair i into j and k—while preserving our common knowledge that it is P and Q. In other words, j and k are both P and Q:

$$\cdots \longrightarrow \boxed{\begin{array}{c} i \\ \hline P(i) \end{array}} \xrightarrow{Q(i)} \boxed{\begin{array}{c} i \\ \hline P(i) \\ Q(i) \end{array}} \xrightarrow{\;?\;} \boxed{\begin{array}{cc} j & k \\ \hline P(j) & P(k) \\ Q(j) & Q(k) \end{array}} \xrightarrow{\substack{R(j) \\ \neg R(k)}} \boxed{\begin{array}{cc} j & k \\ \hline P(j) & P(k) \\ Q(j) & Q(k) \\ R(j) & \neg R(k) \end{array}} \qquad (2)$$

It is impossible to model the update marked '?' above as a relation on standard possible worlds. (By 'standard possible world', I mean what amounts to a first-order model with a domain of individuals.) The reason is that the valuation in a standard possible world is a *total* function—it always returns either true or false. Such a valuation conflates propositions whose truth values persist after a split (like $P(i)$ and $Q(i)$) with propositions whose truth values may vary after a split (like $R(i)$). Now to see the impossibility, suppose w is a world possible in the common ground before the update '?' above. In each such w, we have not only both $P(i)$ and $Q(i)$, but also one of $R(i)$ and $\neg R(i)$. Consider one such w, where we have $R(i)$, say. If '?' above were a relation on worlds, then we would want it to relate w to only worlds where both $P(j) \wedge P(k)$ and $Q(j) \wedge Q(k)$ hold, but not only worlds where $R(j) \wedge R(k)$ holds. But w contains no information to help the relation treat R so differently from P and Q.

The upshot is that a discourse marker cannot naively refer to an individual in a standard possible-world semantics. On that naive view, there is a fact of the matter as to how many letters are in the word 'lava', so if you count 4 letters and I count 3, then we can't both be right. And there is a fact of the matter as to how many things are on the dining table, so if you include the coasters under the glasses whereas I pair up each coaster and glass as one thing, then we can't both be right. But we can. So when we agreed in (1) that there is a P that is Q, what did we existentially quantify over?

We could start to answer this question by blaming underspecification. We would then have to explain what kind of underspecification would let us equivocate between ascribing the properties P and Q to one individual i on one hand, and to two individuals j, k on the other hand. What *are* these underspecified individuals that an indefinite could existentially quantify over? (Perhaps plurals of some sort, with parts?)

As a first step, the rest of this paper defines a modal logic in which a point models not a possible world but a discourse state.

3   A modal logic with individual accessibility

To model how discourse markers split and merge in the course of a conversation, we introduce a modal logic in which accessibility relates not possible worlds but discourse states, and crucially, the individuals at each state as well.

3.1   Frames

Formally, a *frame* for us is a quadruple $F = \langle S, R_S, D, R_D \rangle$ consisting of

1. a set S of *states* (depicted in Figure 1 as $\{s, t, u\}$);
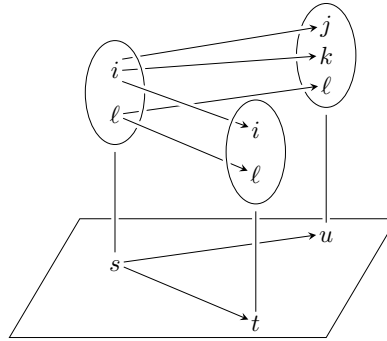
Figure 1: A frame

2. a relation $R_S \subseteq S \times S$ on states, called the *(state) accessibility relation* (the arrows at the bottom of Figure 1);

3. a function D mapping each state $s \in S$ to a set $D(s)$, called the *domain of individuals* at s (the balloons in Figure 1);

4. this is what's new: a function R_D mapping each pair $\langle s, t \rangle \in R_S$ to a relation $R_D(s, t) \subseteq D(s) \times D(t)$, called the *(individual) accessibility relation* (the arrows at the top of Figure 1).

If $\langle s, t \rangle \in R_S$, then we say that the state t is *accessible* from the state s, and notate the relationship infix as $s\,R\,t$. If moreover $\langle x, y \rangle \in R_D(s, t)$, then we say that the individual y at t is *accessible* from the individual x at s, and notate the relationship infix as $x\,_sR_t\,y$. We visualize individual accessibility as lifting state accessibility to a *covering space* (Munkres 2000). It can be implemented by a sort of counterpart relation (Lewis 1968), with postulates modified to suit the difference that our states are discourse states, not possible worlds. (For example, whereas according to Lewis 'it would not have been plausible to postulate that the counterpart relation was transitive', it *is* plausible to postulate that our accessibility relation be transitive, as discussed below.)

3.2   Truth

We define formulas ϕ, terms t, and models M as usual in first-order modal logic. But to define truth, we need not only the notion of *valuations at a state* but also the notion of *valuation accessibility*. A valuation v at a state s is just a function that maps each variable name x to an individual at s. If the state t is accessible from the state s, then we say that a valuation w at t is *accessible* from a valuation v at s just in case v and w (have the same set of variable names for their domains and) map each variable name to an accessible pair of individuals. We notate this relationship infix as $v\,_sR_t\,w$. In short,

$$v\,_sR_t\,w \quad \text{just in case} \quad \forall x.\, v(x)\,_sR_t\,w(x). \tag{3}$$

Finally, we define the *truth* of a formula ϕ in a model M under a valuation v at a state s, notated $M, s, v \Vdash \phi$. The definition is standard except for the modal operators $\Box$ and $\Diamond$, which are dual. We define

$$M, s, v \Vdash \Box\phi \quad \text{just in case} \quad \forall t.\, (s\,R\,t \rightarrow \forall w.\, (v\,_sR_t\,w \rightarrow M, t, w \Vdash \phi)). \tag{4}$$

3.3 Consequences

Two hallmark consequences distinguish our logic from standard modal logic. On one hand, $(x = y)$ does not entail $\square(x = y)$, because one individual at the current state may be related to multiple individuals at future states. On the other hand, if ϕ and ϕ' are the same formula except with free variables x and y exchanged throughout, then $(x = y) \wedge \square\phi$ does entail $\square\phi'$. In particular, ϕ could be $P(x) \wedge Q(x)$, so to continue the example (2), the formula $(x = y) \wedge \square(P(x) \wedge Q(x))$ does entail $\square(P(y) \wedge Q(y))$. Both of these consequences are desirable for modeling our discourse moves of interest, where discourse markers split and merge.

These consequences also offer one way to relate our logic to the *modal logics of relations* that Marx and Venema (1997) study under the rubric of *multi-dimensional modal logic*. Starting at the 'local cube' condition on multi-dimensional frames (LC_n in their Definition 5.3.4), our logic amounts to extending their modal similarity type with a unary modality $\square$. It would be enlightening if their metatheoretical results could be generalized—either to specify $\square$ so that every dimension moves along the same individual accessibility relation, or to axiomatize our logic so that the two hallmark consequences fall out.

3.4 Frame correspondence

Like standard modal logic, our logic enjoys natural frame conditions corresponding to modal axioms. The reflexivity ('T') axiom $\square\phi \to \phi$ is valid for a frame if and only if

$$\forall s.\ (s\,R\,s \wedge \forall x.\ x\,_sR_s\,x). \tag{5}$$

And the transitivity ('4') axiom $\square\phi \to \square\square\phi$ is valid for a frame if and only if

$$\forall s, t, u.\ (s\,R\,t \wedge t\,R\,u) \to (s\,R\,u \wedge \forall x, y, z.\ (x\,_sR_t\,y \wedge y\,_tR_u\,z) \to x\,_sR_u\,z). \tag{6}$$

The difference from standard modal logic in the two conditions are the subformulas $\forall x \ldots$ constraining individual accessibility.

4 Outlook

The axioms (5) and (6) are both intuitive, but they only begin to chip away at a myriad of discourse possibilities made available by individual accessibility: as conversation proceeds, discourse markers may split or merge as well as go in and out of existence. Investigating how interlocutors navigate this sea of states leads us back to the question: what does a discourse marker refer to in our shared reality, if not an individual in a standard possible-world semantics? In short, what is a thing? Our logic suggests that a possible world is a path along the state accessibility relation and a thing is a path along the individual accessibility relation.

References

Maria Aloni. *Quantification under Conceptual Covers*. PhD thesis, Universiteit van Amsterdam, 2000.

Nicholas Asher. *Lexical Meaning in Context: A Web of Words*. Cambridge University Press, Cambridge, 2011.

Chris Barker. Parasitic scope. *Linguistics and Philosophy*, 30(4):407–444, 2007.

Sigrid Beck. The semantics of *different*: Comparison operator and relational adjective. *Linguistics and Philosophy*, 23(2):101–139, 2000.

David K. Lewis. Counterpart theory and quantified modal logic. *Journal of Philosophy*, 65(5): 113–126, 7 March 1968.

Maarten Marx and Yde Venema. *Multi-Dimensional Modal Logic*. Kluwer, Dordrecht, 1997.

James R. Munkres. *Topology*. Prentice-Hall, Englewood Cliffs, NJ, second edition, 2000.

Robert C. Stalnaker. Assertion. In Peter Cole, editor, *Pragmatics*, number 9 in Syntax and Semantics, pages 315–332. Academic Press, San Diego, CA, 1978.

Alternative Representations in Formal Semantics:
A case study of quantifiers

Shane Steinert-Threlkeld[1], Gert-Jan Munneke[2], and Jakub Szymanik[2]

[1] Department of Philosophy, Stanford University
`shanest@stanford.edu`
[2] Institute for Logic, Language, and Computation, University of Amsterdam
`G.J.Munneke@uva.nl, J.K.Szymanik@uva.nl`

Abstract

How do semantic theories fit into the psychology of language more generally? A number of recent theoretical and experimental findings suggest that specifications of truth-conditions generate biases for different verification procedures. In this paper, we show how considerations of different representations of a visual scene in the semantic automata framework can generate predictions for differential working memory activation in proportional quantifier sentence verification. We present experimental results showing that different representations do impact working memory in sentence verification and that 'more than half' and 'most' behave differently in this regard.

1 Introduction

Semantic theories for (fragments of) natural languages typically provide a compositional assignment of truth-conditions to sentences in the repsective language. Such assignments are used to predict competent speakers' judgments of entailment and the truth of sentences in context. This level of generality, however, leaves open exactly how a specification of truth-conditions integrates with cognition more broadly to generate these judgments. There are at least two natural views to take on this issue. The relationship may be *permissive*: once a specification of truth-conditions is "exported" to general cognition, anything goes. There is no systematic connection between the ways that truth-conditions are specified and judgments of truth in context are made. On the other hand, the relationship may be *constrained*: the ways in which truth-conditions are specified correlates with and constrains the methods of verification of sentences in context.

A number of philosophers of language, developing ideas rooted in Frege, have argued that knowing the meaning of a sentence consists in having 'internalized' an algorithm for computing the truth-value of that sentence in a context [Dummett, 1973, Suppes, 1980, Moschovakis, 2006, Horty, 2007]. This line of thought has been put to experimental test recently by psychologists and linguists [Hackl, 2009, Pietroski et al., 2009, Lidz et al., 2011]. These experiments have led the theorists to argue that the relationship between specifications of truth-conditions and verification procedures is in fact constrained. The evidence comes in two forms. In one case, subjects are asked to verify the truth of a single sentence against visual scenes which differ in how amenable they are to different verification procedures. [Pietroski et al., 2009] and [Lidz et al., 2011] find that in the case of sentences involving 'most', such manipulations do not effect verification accuracy. This suggests that the specification of truth-conditions for 'most' constrains in some way the verification procedures available.[1] In another case, subjects

---

[1] In particular, [Lidz et al., 2011] propose the *Interface Transparency Thesis*: "speakers exhibit a bias towards the verification procedures provided by canonical specifications of truth conditions" (p. 229). While we find this thesis imminently plausible, we keep the discussion at a more general level for reasons that will become clear in the later discussion of results.

are asked to verify two truth-conditionally equivalent sentences against visual scenes in the same experimental paradigm. [Hackl, 2009] finds that there are differences in the way subjects perform self-paced counting tasks when verifying sentences containing 'most' and 'more than half'. Since the sentences are truth-conditionally equivalent, this suggests that the two quantifiers possess different specifications of the truth-conditions, which constrain the methods of verification.[2]

In this paper, we contribute to the growing body of evidence for a constrained relationship by exploring the impact of different presentations of a visual scene on working memory load in proportional quantifier sentence verification. First, we present a computational model for quantifier meanings which has made empirically verified predictions about working memory in sentence verification. Then, we show how to extend that framework to handle different representations of the visual scene. This extension motivates a prediction that how a visual scene is presented will effect working memory involvement. We then present experimental results that make good on this prediction. The results also indicate that 'most' and 'more than half' are effected differently by the visual scene manipulation, providing further evidence for a constrained relationship between truth-condition specifications and verification procedures. The paper concludes with a discussion of future directions.

2   Semantic Automata and Working Memory

The semantic automata approach to generalized quantifiers associates with each quantifier Q a formal language L_Q and a machine M_Q accepting L_Q. Definability of quantifiers in various logics corresponds to levels of the Chomsky hierarchy of languages/machines. All first-order definable quantifiers have a finite-state automaton (FSA) / regular language, while others – notably proportional quantifiers like *most* and *more than half* – require a pushdown automaton (PDA) / context-free language [van Benthem, 1986, Mostowski, 1998]. A pushdown automaton essentially augments a finite-state automaton with a stack, a form of *memory*. If one views the automata as verification procedures somehow internalized in the minds of competent speakers of the language, this leads one to expect that verifying sentences with quantifiers which require a PDA will use more working memory than verifying sentences with quantifiers that have FSAs.

This hypothesis has been demonstrated true in a number of experiments over the past decade. In a pioneering study, [McMillan et al., 2005] had participants verify sentences of the form 'Q of the balls are blue' while in an fMRI machine. They varied Q between quantifiers that have FSAs ('some', 'all', 'at least three') and those that only have PDAs ('most'). In addition to behaviorally finding that the latter are more difficult to verify, they found differential activation in dorsolateral prefrontal and inferior frontal cortices bilaterally. These brain regions have previously been found to be highly involved in working memory by, among others, [Braver et al., 1997].[3]

---

[2] An apparently related line of work concerns the relationship between superlative quantifiers, such as 'at least n', and the corresponding comparative quantifier 'more than $n-1$'. Traditional semantic theories, such as generalized quantifiers theor, consider these to be truth-conditionally equivalent. [Geurts et al., 2010] show experimentally that the two quantifiers are processed differently. They, however, build on [Geurts and Nouwen, 2007], in which it is argued that the two quantifiers do in fact differ in subtle ways in their truth-conditions. Nevertheless, [Cummins and Katsos, 2010] provide further experiments in support of the view that superlatives and comparatives are truth-conditionally equivalent but have different psychological profiles due to the presence of strict versus weak inequality in the specification of the truth-conditions. Under this interpretation, these quantifiers exhibit a similar phenomenon to the one being discussed here.

[3] This line of work has continued in [McMillan et al., 2006, Troiani et al., 2009a, Troiani et al., 2009b] and has been nicely summarized by [Clark and Grossman, 2007, Clark, 2011].

This differential activation of working memory in quantifier verification has also been demonstrated behaviorally. [Szymanik and Zajenkowski, 2010] show that reaction times in a sentence verification task are much higher for proportional quantifiers than for first-order ones. [Szymanik and Zajenkowski, 2011] combined quantifier sentence verification with a memory span task. In particular, participants were asked to memorize a sequence of either 4 or 6 digits before performing the verification task and were then asked to recall it afterwards. They found that reactions times were much longer and accuracy much worse in verification for proportional quantifiers than for FSA quantifiers.[4] Moreover, in the 4-digit case, digit recall performance was significantly lower after proportional quantifier verification. Finally, [Zajenkowski et al., 2011] have compared the verification of natural language quantifier sentences in a group of patients with schizophrenia (and associated working memory deficits) and a healthy control group. In both groups, the difficulty of the quantifiers was consistent with the computational predictions. Patients with schizophrenia took more time to solve the problems with every quantifier. They were significantly less accurate, however, only with proportional quantifiers.[5] These results, together with the neuroimaging ones, show that the semantic automata model makes good predictions about working memory involvement in quantified sentence verification.

2.1 Representation in Semantic Automata

All of the aforementioned work, however, neglects a crucial component of the psychological task: *how* the visual scene is represented. To show how the semantic automata model can also be used to make predictions about the effect of visual presentation, we must explain the model in slightly more detail. A generalized quantifier Q is a class of finite models of the form $\langle M, A, B \rangle$ [Barwise and Cooper, 1981]. To generate L_Q, a mapping $\tau(\cdot)$ from such models into strings of 0s and 1s is defined by sending elements of $A \cap B$ to 1 and elements of $A \setminus B$ to 0.[6] L_Q, then, just is the set of strings generated by τ from the models in Q.

As an example, consider verifying the sentence 'Most of the dots are blue' against a visual scene with 3 blue dots and 2 yellow (i.e. non-blue) dots. We can represent this scene by a model $\mathcal{M}$ with $M = A = \{d_1, \ldots, d_5\}$ and $B = \{d_1, d_4, d_5\}$. Here, A is the set of dots d_i and B is the set of blue dots. Thus, $A \cap B = B = \{d_1, d_4, d_5\}$ while $A \setminus B = \{d_2, d_3\}$. The encoding described above will generate the string $\tau(\mathcal{M}) = 10011$. The corresponding machine $M_{>1/2}$ has two states: one representing a 'yes' response and one a 'no' response. Inuitively, it pairs off 1s and 0s and returns a 'yes' if and only if there are more 1s than 0s. It processes the string $\tau(\mathcal{M})$ as follows: it pushes the first 1 on to the stack, but pops it off of the stack when encountering the first 0. Then, it pushes the second 0 onto the stack, but pops it off of the stack when encountering the second 1. Then, it pushes the final 1 onto the stack. Because, at the end of processing the string, only 1s are on the stack, the machine accepts the string. This reflects the fact that in $\mathcal{M}$, most of of the dots are in fact blue: $A \cap B$ – the set of blue dots – has a bigger cardinality than $A \setminus B$ – the set of non-blue dots.

In the semantic automata literature thus far, only a single τ has ever been considered.[7] From the psychological perspective, this is surprising: this perspective takes the visual scene to

---

[4]They used *parity quantifiers* like 'even number of' for reasons that do not concern us here.

[5]See also [Zajenkowski and Szymanik, 2013] for the relationship between intelligence, working memory, executive functions and complexity of quantifiers. [Szymanik, 2016] provides a summary of these experimental results.

[6]That $A \cap B$ and $A \setminus B$ are the only necessary sets follows from the properties of Conservativity and Extensionality. See [van Benthem, 1986].

[7]The only exceptions are when extending the framework to handle more complicated kinds of quantified sentences, e.g. with quantifiers in subject and object position [Steinert-Threlkeld and Icard III., 2013, Steinert-Threlkeld, 201x, McWhirter, 2014].

provide a model $\langle M, A, B \rangle$ with τ embodying how that scene is represented in the mind. Given that different representations of the same situation can effect performance in cognitive tasks, it is natural to suppose that different encodings of finite models might make a relevant difference in quantifier sentence verification. Consider, for instance, the task of verifying 'More than half of the dots yellow' against a visual scene of yellow and blue dots. The τ above will work in every case. If, however, yellow and blue dots are paired together (with all the remaining being of a single color), the agent could make sure that each pair has one dot of each color and that all of the remaining are yellow or blue (corresponding to yes/no answers).

We can model this situation by supposing that the scene makes salient a certain pairing of dots so that the models which are input to the encoding function actually have the form $\langle M, A, B, R \rangle$ where $R \subseteq M \times M$ (here, A is the set of dots and B the set of blue dots). A new mapping τ' can be defined which maps such models into an alphabet containing *pairs* of symbols in addition to 0s and 1s. τ' will map all pairs in R to pairs of symbols in the natural way: if, for instance, $\langle b, a \rangle \in R$ where $b \in A \cap B$ and $a \in A \setminus B$, then $\langle b, a \rangle$ will get mapped to $\langle 1, 0 \rangle$. Then, any elements of the model that are not paired will get mapped to 0 or 1 as before. Using this encoding τ', if the only pairs in the encoding are $\langle 1, 0 \rangle$ or $\langle 0, 1 \rangle$ and all individual symbols are 1s, then the sentence is true. Similarly, if all individual symbols are 0s, it is false. These facts, however, can paradigmatically be checked by a finite-state automaton.

Given that the difference between PDA and FSA-acceptable quantifiers has been shown to correlate very strongly with working memory involvement in verification, the above discussion of encoding suggests that verifying a proportional quantifier sentence against a visual scene in which elements are paired will require less working memory than doing so against a visual scene with randomly scattered elements. Moreover, if distinct but truth-conditionally equivalent quantifiers exhibit different levels of sensitivity to the type of visual scene, this would provide evidence for a constrained relationship between specifications of truth-conditions and verification procedures.

3   Experiment

3.1   Methods

To test this hypotheses, we ran an experiment in which participants had to answer a question while being shown a visual scene containing two types of objects. There were three main conditions, corresponding to the following questions. In parentheses are listed the types of objects in the visual scene.

1. Are more than half of the dots yellow?

   (Yellow and blue dots)

2. Are more than half of the letters 'E'?

   (Characters 'E' and 'F')

3. Are most of the dots yellow?

   (Yellow and blue dots)

The visual scenes came in two types: random, with objects distributed randomly across the image, and paired, with objects presented in pairs consisting of one object of each type. We manipulated the proportions of the two types of objects – 8/7, 9/8 and 10/9 – as well as the correct answer to the question (yes/no). Examples of stimuli demonstrating these variables are given in Figure 1. Participants were randomly assigned to one of the three main conditions and one of the three proportions; the other variables were within-subjects.
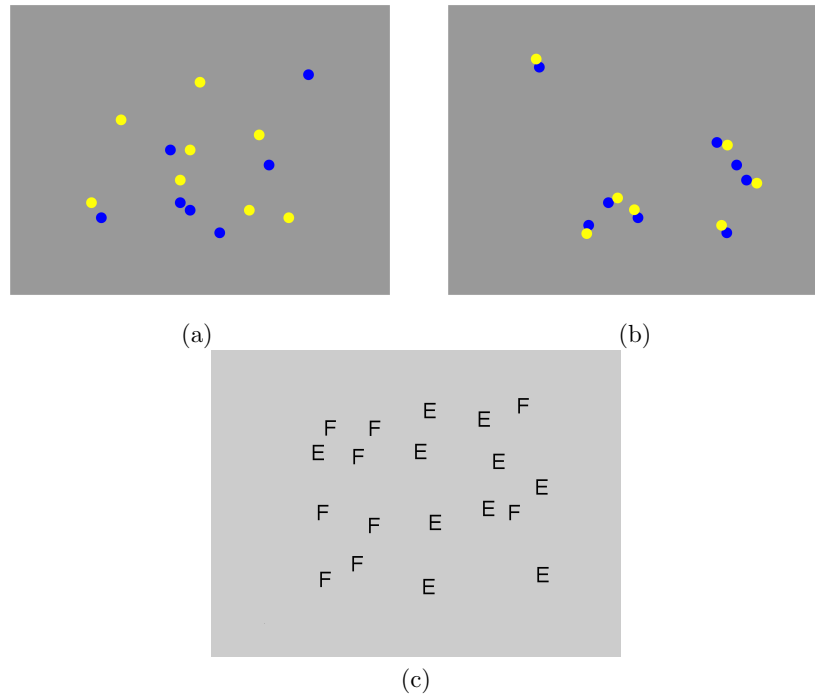
Figure 1: Example stimuli. (a) Condition: (1) and (3); Proportion: 8/7; Type: random; Answer: yes. (b) Condition: (1) and (3); Proportion: 8/7; Type: paired; Answer: no. (c) Condition: (2); Proportion: 10/9; Type: random; Answer: yes.

To manipulate working memory load, we included a digit recall task. Before the onset of a picture, a participant saw a string of 5 digits for 1500ms. After answering the question against the picture (sentence verification), the participant was presented with one of the digits as a probe and asked to give the digit in the sequence following the probe. In the low memory condition, the same sequence of digits – (0,1,2,3,4) – was used in every trial; in the high memory condition, the digits were randomized [de Fockert et al., 2001]. Participants performed one block in the low memory condition and one in the high condition, with a forced 30 second break in between. Each block consisted of 40 trials (10 in each combination of yes/no and random/paired). The order of the blocks was random as was the order of trials within the block. These main blocks were preceded by a 4-trial training block, after which the participants received feedback on their performance.

3.2   Participants

We recruited participants from Mechanical Turk, all from the United States with HIT approval rate of at least 99%. They were compensated with $1.50. In condition (1), we had 79 participants. We excluded 20 participants who made more than 10 errors on the sentence verification task. The remaining ($N = 59$) were aged between 20 and 59 ($M = 33$, $SD = 9.9$) with 28 male and 31 female. For conditions (2) and (3), we had 60 participants and excluded those

who committed more than 30 errors in the verification task. For condition (2), this left us with $N = 54$, 27 male, 27 female, aged between 20 and 69 ($M = 35, SD = 12$). For condition (3), this left us with $N = 57$, 28 male, 27 female, aged between 20 and 68 ($M = 35, SD = 9.6$).

3.3   Results

In both conditions (1) and (2), there is a significant main effect of stimulus type on both the accuracy and reaction time of the sentence verification task. Participants answer faster and more accurately when pictures show paired rather than random dots. Similary, we found main effects of WM condition on both accuracy and RT of the digit recall task in all proportions in both conditions.

To test whether the stimulus type – random or paired – makes a difference in working memory demands, we ran tests to see whether reaction time and accuracy in the *digit recall* task were thereby effected. In particular, we ran a multiple regression of digit recall RT on stimulus type and WM condition as well as a log-likelihood difference test of digit recall accuracy on the same two variables.

Crucially, we found a significant interaction effect of stimulus type and WM condition in digit recall RT of condition (1) in proportion 8/7 ($\chi^2(4) = 4, p = 0.043$). The difference in the RT of the digit task between low and high WM conditions is greater for random pictures ($M = 1049$ ms, $SD = 2193$ ms) than for paired pictures ($M = 671$ ms, $SD = 1191$ ms). We also find an interaction effect on digit recall accuracy in proportion 8/7 for condition (2) ($\chi^2(1) = 4.19, p < .0407$). The increase in error rate due to the hard WM condition was higher for random pictures (11.25%) than for paired pictures (9%). In condition (2), we also found a trend towards a significant interaction effect on accuracy in the 9/8 proportion ($\chi^2(1) = 3.61, p < .057$). The increase in error rate due to the hard WM condition was higher for random pictures (20%) than for paired pictures (11.18%).

The results for condition (3) – 'most' – are a bit different. While there was a main effect of WM condition on accuracy and RT of digit recall in all proportions, there was no main effect of stimulus type on RT in proportion 9/8 nor on accuracy in proportion 10/9. More importantly, we find no significant interaction effects of stimulus type and WM condition on digit recall RT or accuracy in any proportions. Figure 2 shows the observed main effects and Figure 3 shows the observed interaction effects for all of the conditions.

4   Discussion

Consideration of the role of representation in the semantic automata framework led us to hypothesize that working memory activation in proportional quantifier sentence verification should depend on how a visual scene is presented. In a limited context, we do indeed find such a dependence: in the 8/7 proportion for conditions (1) and (2), the effect of working memory load depended on whether the stimulus type was random or paired. We hypothesize that the effect occurs only at this proportion because participants are more likely to approximate in the larger proportion conditions [Halberda and Feigenson, 2008]. Moreover, that the strongest interaction effects were found in condition (2) – with E/F images – supports this approximation interpretation. Because the two types of objects in the visual scene are so similar to each other, it becomes nearly impossible to approximate. Thus, in the case when approximation is most difficult, we get the strongest interaction effects. This also suggests future manipulations: if we made the 'E's and 'F's colored, perhaps the effect would weaken as approximation becomes available.
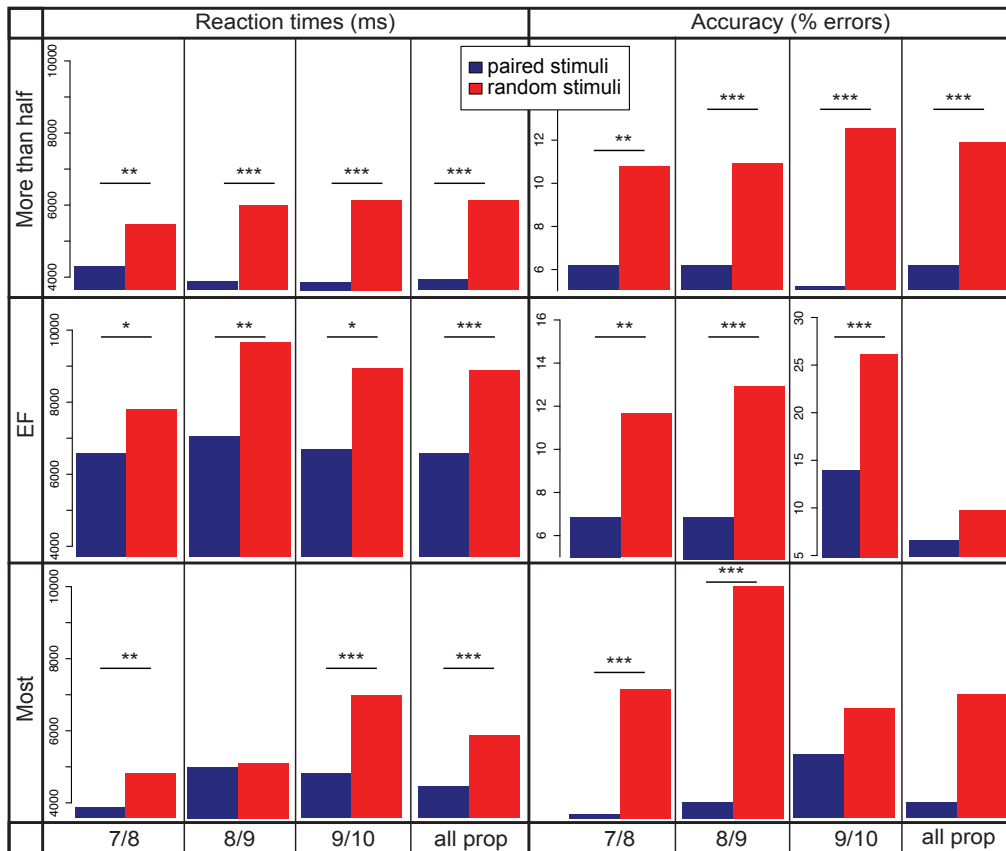
Figure 2: Main effects of stimulus type on the quantifier verification task. Above a column, * means $p < .05$, ** means $p < .01$, and *** means $p < .001$.

4.1 'Most' and 'More than half'

A very striking feature of the experimental results concerns the very different results between conditions (1) and (3), which differ only in the use of 'more than half' in the former and 'most' in the latter. In particular, the lack of any interaction effects in the latter suggests that the manipulation of stimulus type does not effect working memory demands for 'most' in the way that it does for 'more than half'. Minimally, this provides evidence for a constrained relationship between specifications of truth-conditions and verification procedures because two truth-conditionally equivalent sentences have very different verification profiles.

Can more about the nature of this constraint be said? [Hackl, 2009] argues that the difference amounts to the following: 'most' is a superlative while 'more than half' makes explicit size comparisons in terms of proportion. He argues that this difference manifests itself in terms of verification procedures in that 'most' will lend itself to a "vote-counting" strategy which resembles pairing colored dots and seeing what colors remain. Interestingly, such an interpretation does not fit well with our data: if 'most' lent itself to pairing strategies, one would expect to

Increase in digit task RT and errors due to high versus low WM load
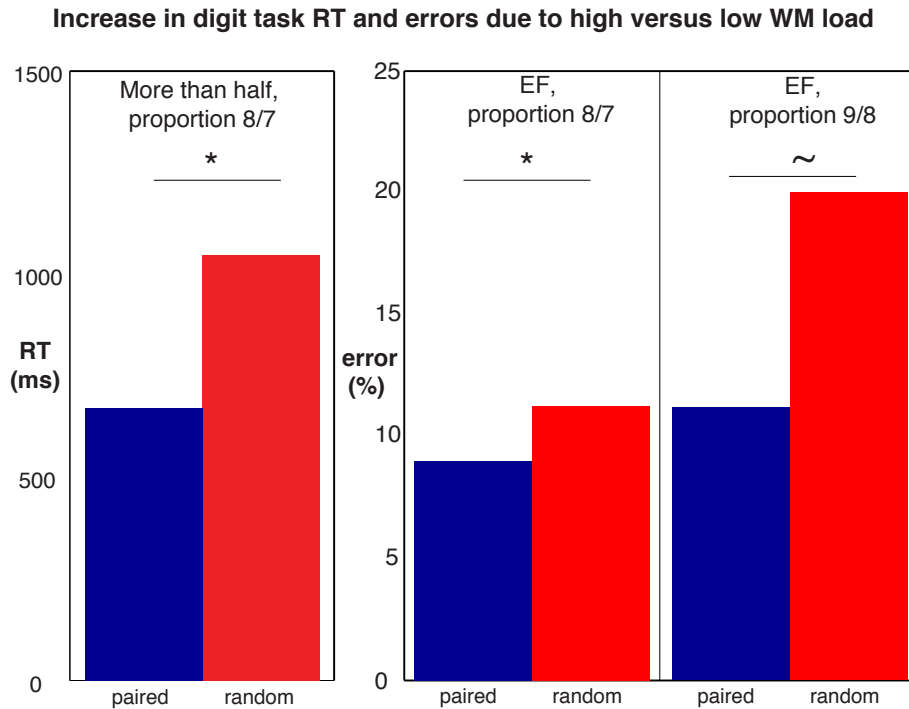
Figure 3: Effects of the interaction of stimulus type and working memory on the digit recall task. Here, $*$ indicates $p < .05$ and $\sim$ indicates a trend toward significance $(0.05 < p < 0.057)$.

see a marked difference in WM demand between random and paired stimuli.[8]

Moreover, [Pietroski et al., 2009] present evidence that 'most' does not favor a vote-counting procedure. They conducted an experiment where participants verified a 'most' sentence against a visual scene of yellow and blue dots which is flashed for 200ms. They find that manipulating the scene between random and paired placement of dots does not effect the accuracy of judgments. Were 'most' to exhibit a bias towards vote-counting verification, one would expect accuracy to improve in paired scenes. At the present, then, we can conclude that 'most' and 'more than half' do indeed verify in their verification behavior; exactly how remains undetermined.

5   Conclusion and Future Directions

In this paper, we have presented novel experimental support for the idea that specifications of truth-conditions for sentences constrain methods of verification. In particular, we used the semantic automata framework to show how different presentations of a visual scene can effect working memory demand. Verifying sentences containing 'more than half', but not those

---

[8][Solt, 201x] argues for a very similar difference in the logical forms of 'most' and 'more than half' which places constraints on the types of measurement scale required. She uses this to predict interesting distributional properties and argues that the scale requirements can also account for Hackl's results.

containing 'most', exhibit effects on working memory demand when verified against paired versus random visual scenes. These results present further evidence for a constraint between representations of truth-conditions and verification and a proof-of-concept that working memory can be used to probe such constraints.

Much work, however, remains to be done. On the experimental side, a next step consists in running these experiments using EEG. The memory protocol used in this paper has been shown to be effective in such a context [de Fockert et al., 2001]. As alluded to earlier, more manipulations of our present set-up – color(s) of the items in the stimuli, types of items, the number of proportions considered – could be included. These would provide a fuller picture of the range of data for which a theory of the relationship between truth-conditions and verification needs to account.

On the theoretical side, significant modeling work needs to be done to predict the results here. In particular, how do different quantifiers exhibit a bias for different encodings when visual scenes make them salient? This would help answer why 'more than half', but not 'most', appears to have this bias. At the present moment, the semantic automata framework treats 'most' and 'more than half' equivalently. The experimental results here indicate that this assumption should be relaxed. Moreover, a fully developed model would need to incorporate the proportion of items in a visual scene as a predictor. The frameworks of resource-rational [Griffiths et al., 2015] or boundedly-rational [Icard, 2014] analysis are likely to be useful in this endeavor.

References

[Barwise and Cooper, 1981] Barwise, J. and Cooper, R. (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4(2):159–219.

[Braver et al., 1997] Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory. *NeuroImage*, 5(1):49–62.

[Clark, 2011] Clark, R. (2011). Generalized Quantifiers and Number Sense. *Philosophy Compass*, 6(9):611–621.

[Clark and Grossman, 2007] Clark, R. and Grossman, M. (2007). Number sense and quantifier interpretation. *Topoi*, 26(1):51–62.

[Cummins and Katsos, 2010] Cummins, C. and Katsos, N. (2010). Comparative and Superlative Quantifiers: Pragmatic Effects of Comparison Type. *Journal of Semantics*, 27:271–305.

[de Fockert et al., 2001] de Fockert, J. W., Rees, G., Frith, C. D., and Lavie, N. (2001). The Role of Working Memory in Visual Selective Attention. *Science*, 291(5509):1803–1806.

[Dummett, 1973] Dummett, M. (1973). *Frege: Philosophy of Language*. Harper & Row, New York.

[Geurts et al., 2010] Geurts, B., Katsos, N., Cummins, C., Moons, J., and Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25(1):130–148.

[Geurts and Nouwen, 2007] Geurts, B. and Nouwen, R. (2007). At least et al.: The semantics of scalar modifiers. *Language*, 83(3):533–559.

[Griffiths et al., 2015] Griffiths, T. L., Lieder, F., and Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7(2):217–229.

[Hackl, 2009] Hackl, M. (2009). On the grammar and processing of proportional quantifiers: *most* versus *more than half*. *Natural Language Semantics*, 17(1):63–98.

[Halberda and Feigenson, 2008] Halberda, J. and Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5):1457–1465.

[Horty, 2007] Horty, J. (2007). *Frege on Definitions*. Oxford University Press, Oxford.

[Icard, 2014] Icard, T. (2014). Toward boundedly rational analysis. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

[Lidz et al., 2011] Lidz, J., Pietroski, P., Halberda, J., and Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics*, 19(3):227–256.

[McMillan et al., 2005] McMillan, C. T., Clark, R., Moore, P., Devita, C., and Grossman, M. (2005). Neural basis for generalized quantifier comprehension. *Neuropsychologia*, 43(12):1729–1737.

[McMillan et al., 2006] McMillan, C. T., Clark, R., Moore, P., and Grossman, M. (2006). Quantifier comprehension in corticobasal degeneration. *Brain and Cognition*, 62(3):250–260.

[McWhirter, 2014] McWhirter, S. (2014). *An Automata-Theoretic Perspective on Polyadic Quantification in Natural Language*. Masters of logic thesis, Universiteit van Amsterdam.

[Moschovakis, 2006] Moschovakis, Y. N. (2006). A logical calculus of meaning and synonymy. *Linguistics and Philosophy*, 29(1):27–89.

[Mostowski, 1998] Mostowski, M. (1998). Computational semantics for monadic quantifiers. *Journal of Applied Non-Classical Logics*, 8(1-2):107–121.

[Pietroski et al., 2009] Pietroski, P., Lidz, J., Hunter, T., and Halberda, J. (2009). The meaning of 'most': Semantics, numerosity and psychology. *Mind and Language*, 24(5):554–585.

[Solt, 201x] Solt, S. (201x). On measurement and quantification: The case of most and more than half. *Language*.

[Steinert-Threlkeld, 201x] Steinert-Threlkeld, S. (201x). Some Properties of Iterated Languages. *Journal of Logic, Language and Information*.

[Steinert-Threlkeld and Icard III., 2013] Steinert-Threlkeld, S. and Icard III., T. F. (2013). Iterating semantic automata. *Linguistics and Philosophy*, 36(2):151–173.

[Suppes, 1980] Suppes, P. (1980). Procedural semantics. In Haller, R. and Grassl, W., editors, *Language, Logic, and Philosophy: Proceedings of the 4th International Wittgenstein Symposium*, pages 27–35. Hölder-Pichler-Tempsy, Vienna.

[Szymanik, 2016] Szymanik, J. (2016). *Quantifiers and Cognition: Logical and Computational Perspectives*. Studies in Linguistics and Philosophy. Springer.

[Szymanik and Zajenkowski, 2010] Szymanik, J. and Zajenkowski, M. (2010). Comprehension of simple quantifiers: empirical evaluation of a computational model. *Cognitive Science*, 34(3):521–532.

[Szymanik and Zajenkowski, 2011] Szymanik, J. and Zajenkowski, M. (2011). Contribution of working memory in parity and proportional judgments. *Belgian Journal of Linguistics*, 25(1):176–194.

[Troiani et al., 2009a] Troiani, V., Peelle, J. E., Clark, R., and Grossman, M. (2009a). Is it logical to count on quantifiers? Dissociable neural networks underlying numerical and logical quantifiers. *Neuropsychologia*, 47(1):104–11.

[Troiani et al., 2009b] Troiani, V., Peelle, J. E., McMillan, C., Clark, R., and Grossman, M. (2009b). Magnitude and parity as complementary attributes of quantifier statements. *Neuropsychologia*, 47(12):2684–5.

[van Benthem, 1986] van Benthem, J. (1986). *Essays in Logical Semantics*. D. Reidel Publishing Company, Dordrecht.

[Zajenkowski et al., 2011] Zajenkowski, M., Styła, R., and Szymanik, J. (2011). A computational approach to quantifiers as an explanation for some language impairments in schizophrenia. *Journal of Communication Disorders*, 44(6):595–600.

[Zajenkowski and Szymanik, 2013] Zajenkowski, M. and Szymanik, J. (2013). MOST intelligent people are accurate and SOME fast people are intelligent. Intelligence, working memory, and semantic processing of quantifiers from a computational perspective. *Intelligence*, 41(5):456–466.

The role of preferred outcomes in determining implicit questioning strategies

Jon Scott Stevens*

Center for General Linguistics, Berlin, Germany

Abstract

We look at a particular class of indirect answers to yes/no questions, where the speaker supposes that the hearer's question was motivated by some underlying goal and aims to provide the hearer with an alternative way of accomplishing it. We posit that the questions in these cases are sub-questions in an implicit questioning strategy which is anchored to the hearer's goal, and which must be inferred probabilistically by the speaker. The use of these implicit questioning strategies is licensed when the hearer has a preferred outcome, i.e., better and worse ways to accomplish her goal. We formally model the generation and interpretation of questions and answers in this context as a Bayesian game.

1 Indirectness & implicit questioning strategies

We consider a particular class of phenomenon whereby a yes/no question ('YN-question') is answered indirectly with a particular intonation pattern, lacking the falling pitch contour that typically characterizes the end of a declarative answer.

(1)   Q: Does Bob's Market sell turnips?
        A: Freshtown sells turnips...
            H*   L-H%

The placement of the H* pitch accent in the answer is determined by focus. The salience of $\lambda x.sell(x, turnips)$ introduced by the question requires focus on the subject in (1) (Rooth, 1992; Schwarzschild, 1999; Wagner, 2012b), and in a different context, e.g. in (2), we find the pitch accent occurring elsewhere to reflect the different focus structure.

(2)   Q: Does Bob's Market sell sour cream and onion chips?
        A: Bob's Market sells cheese and onion chips...
                    H*   L-H%

But independently of focus placement, we find a rising boundary tone which signals discourse non-finality, typically suggesting that the question under discussion (Roberts, 1996) is still open (see Lai, 2012). The answer does have a flavor of non-finality in that it is interpreted by default as a mention-some answer, indicating in (1) that perhaps other markets sell turnips as well (see also Wagner, 2012a). But rather than signaling that the question is unresolved, the answers in (1) and (2) both clearly implicate a 'no' answer to the YN-question (assuming a context in which the answerer is assumed to know who sells what), allowing the discourse to potentially conclude without any follow-ups.

The license for these answers is not so straightforward under previous approaches to indirect answerhood which hold that indirect answers must allow for a clear inference of a 'yes' or 'no' answer based on the denotation of the response given (Asher and Lascarides, 2003; de Marneffe et al., 2009), as in (3).

Who ate what?

What did Alice eat?   What did Bob eat?
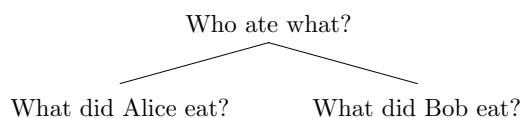
Figure 1: A questioning strategy.

(3)   Q: Does Bob's Market sell turnips?
        A: Bob's Market has a terrible selection of produce.

Here, there is a clear probabilistic inference from the answer given to the proposition 'Bob's Market does not sell turnips', a valid literal answer to the YN-question. But in (1) and (2), this does not hold. There is no clear inferential link, for example, between Bob's Market selling turnips and Freshtown selling turnips.

Such indirect answers are called 'alternative answers' by Stevens et al. (2015), who model the generation of responses to YN-questions by representing and reasoning about the questioner's *goals*. Intuitively, questions are a means to some end, and often a question serves to gather the information required to accomplish some goal in the real world (van Rooij, 2003; Benz and van Rooij, 2007). Indirect answers can be licensed by anticipating these goals and offering alternative solutions. In (1), for example, one assumes that the questioner likely has the goal to go shopping for turnips, and needs to know which store(s) she can go to in order to do this. It is only in virtue of this assumption that the answer in (1) is licensed. Consider some other logically possible alternative answers.

(4)   Q: Does Bob's Market sell turnips?
        A: ?Bob's Market sells carrots. . .

(5)   Q: Does Bob's Market sell turnips?
        A: #Bob's Market sells hand soap. . .

Whereas the exchange in (1) is acceptable more or less out of the blue, (4) would require some explicit evidence that carrots were an acceptable alternative to turnips (perhaps the questioner simply wants some root vegetables, and is not picky), and (5) would require a much stranger context, and it is infelicitous out of the blue. This difference can be seen as a difference in probability of goals.

Formally, the difference between (1), (4) and (5) can be analyzed as a constraint on *implicit questioning strategies*, a type of questioning strategy in the sense of Roberts (1996) and Büring (2003). Originally devised to account for contrastive topics in answers to multiple wh-questions, a questioning strategy is a way of breaking questions down into sub-questions. Sub-questions are defined, after Groenendijk and Stokhof (1984), such that Q1 is a sub-question of Q0 iff exhaustively answering Q0 logically provides an answer to Q1. Strategies may be structured as trees as in Fig.1, which represents a possible strategy for the multiple wh-question 'Who ate what?'. Imagine a discourse in which the only two relevant individuals are Alice and Bob. The question 'Who ate what?' can be exhaustively answered by first answering 'What did Alice eat?' and then answering 'What did Bob eat?', resulting in an answer like, "Alice ate turnips, and Bob ate carrots."

We can construct a similar tree for (1), as in Fig.2. But crucially, although the root node question in structures like Fig.1 which motivate the placement of contrastive topics, e.g. a multiple wh-question like 'Who ate what?', is typically taken to be a proper question under discussion—that is, it is known to both interlocutors—the questioning strategy in Fig.2 in the context of (1) is one where the root node question is private to the questioner, and thus must be inferred by the answerer. In other words, only one of the subordinate

Where can I get some turnips?

Does Bob's Market sell turnips?   Does Freshtown sell turnips?
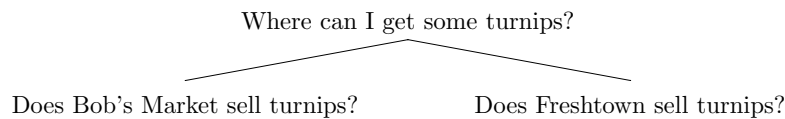
Figure 2: An implicit questioning strategy.

YN-questions is explicitly put on the table in the dialogue. The overarching strategy is implicit.[1]

Intuitively, the answerer in (1) is making two inferences. First, the observed YN-question is likely to have been motivated by an underlying super-question, 'Where can I get some turnips?'. Second, the hearer's goal in asking the question is to find a single place to buy turnips, and therefore a mention-some answer is appropriate. Taken together, these two inferences provide a clear path to an indirect answer: If Bob's Market sells turnips, then say so, and if not, then anticipate any sister questions by offering a single alternative turnip seller. Because the indirect answer is only ever licensed in the event that Bob's Market does not sell turnips, assuming that the speaker knows who sells what, an implicit 'no' can always be inferred.

The infelicity of (4) and (5), then, is reduced to the unavailability in a neutral context of the inference that the observed YN-question is part of an implicit questioning strategy headed by 'Does Bob's Market sell turnips and/or carrots?' or 'Does Bob's Market sell turnips and/or hand soap?', respectively. The availability of these inferences is highly context-dependent—with enough context, almost any implicit questioning strategy can be inferred.

(6)   Context: The questioner has a rather strange Secret Santa assignment this year, who wants either turnips or hand soap for Christmas. The questioner is going to Bob's Market later and wants to make a shopping list.
        Q: Does Bob's Market sell turnips?
        A: Bob's Market sells hand soap...

It is possible to account for these inferences within a game-theoretic framework (Benz et al., 2006; Parikh, 2010; Clark, 2011; Franke, 2011; Benz, 2012) by modeling such QA-exchanges as Bayesian games, where the behavior of both the questioner and answerer is determined by probabilistic reasoning about unknown variables (for the answerer, the questioner's underlying goal, and for the questioner, the true state of the world). Under this approach, it is the common-sense knowledge that finding a store at which to buy a particular item is a likely enough goal to have, combined with a drive toward efficiency in dialogue (anticipating follow-up questions, etc.), which licenses the indirect answer.

Stevens et al. (2015) offer a computational model along these lines, where the problem of producing alternative answers is reduced to the problem of learning a prior probability distribution over possible hearer goals. But that model is oriented toward generating answers to questions, and doesn't offer a full picture of how the questions themselves are selected. Therefore, one important puzzle remains: Why would the questioner ever employ an implicit questioning strategy in the first place? Instead of Fig.2, why not simply ask, 'Where can I get some turnips?' And why ask about Bob's Market first and not Freshtown? In this paper we argue that implicit questioning strategies of this type are to be expected whenever the questioner has a *preferred outcome*. For example, it might be the case that Bob's Market is cheaper or closer to the questioner's house than Freshtown, and thus the

---

[1]That the question is not shared, but rather merely inferrable, explains why we do not find narrow focus on *Bob's Market* in the question.

best outcome is one where the questioner buys turnips from Bob's. To ask the more general question would risk a mention-some answer that guides the questioner to a sub-optimal market when she could have gone to the preferred market instead. Acknowledging that there are better and worse ways to accomplish a goal gives us a fuller picture of how questions and answers are generated and interpreted in these contexts.

We begin by outlining our game-theoretic model, which combines insights from Franke's (2011) 'iterated best response' approach and Benz's (2012) 'error models'. We then work through a derivation of examples (1), (5) and (6).

2 Model

We model QA-exchanges as signaling games, a class of Bayesian games (Fudenberg and Tirole, 1991). The players in the game are the questioner and the answerer, who both want the questioner's goal to be accomplished. Based on the question that was asked, the answerer strategically selects a *message*, m, which carries some conventional propositional content. The questioner must then reason, based on m, about the current state of the world, ω (e.g., which stores sell which items). Using these inferences about the world state, the questioner can choose an *action*, a (e.g., going to Bob's or Freshtown), that will lead to an optimal outcome. The optimality of an outcome—its *utility*—is determined by whether and how well it accomplishes the questioner's *goal*, γ. The questioner can anticipate what messages might be generated by the answerer in order to strategically choose questions that most efficiently optimize her outcome. The *selection procedures* whereby the players select questions, answers and actions can be represented as functions which yield sets of strategically optimal questions, answers and actions from which the player randomly selects (allowing for a tie between multiple equally beneficial moves). In the end, any set of procedures for selecting questions, answers and actions for which neither player can improve her expected outcome given the behavior of the other player is an *equilibrium*. Our goal is to account for implicit questioning strategies and their associated indirect answers by showing that these are felicitous only when they are part of an equilibrium in a signaling game.

2.1 Goals & expected utility of actions

We begin by representing a space of possible questioner goals, where a goal is a set of instructions for accomplishing some concrete task. More precisely, for our purposes, a goal is an ordered tuple of conditional imperatives, where executing any one of the imperatives under the right condition will lead to the goal being accomplished. For example, instructions for buying turnips could be represented as an ordered tuple ⟨'Go to Bob's if Bob's sells turnips', 'Go to Freshtown if Freshtown sells turnips'⟩, where the second instruction is only executed if the first fails. If both fail, then the questioner should simply do nothing. Consider three possible goals, representations of which are given in Table 1, where each goal involves shopping and is accomplished by going to one of two area stores and/or adding relevant items to a shopping list. The ordering of the instructions encodes the questioner's preferences. γ_t instructs the questioner to go to Bob's if Bob's sells turnips, regardless of what Freshtown sells. Only if that condition fails to hold should the questioner consider going to Freshtown. This encodes a preference to go to Bob's. Similarly, γ_B, which corresponds to a context like in (6), encodes a preference to buy turnips over hand soap (perhaps because turnips are cheaper). Finally, the most specific goal to be inferred from the question 'Does Bob's Market sell turnips?' is the singleton goal γ_0: The questioner wants to buy turnips from Bob's, and no other store-item pair will do.

The first step in building up our model will be to use these goal representations to create a generalized schema for generating *utility functions* which we will call *action utility*

Label	Representation	Description
γ_t	$\langle go.to(B) \text{ if } sell(B,t),\ go.to(F) \text{ if } sell(F,t)\rangle$	Questioner wants to buy turnips from either Bob's or Freshtown, and prefers to go to Bob's.
γ_B	$\langle add(t) \text{ if } sell(B,t),\ add(h) \text{ if } sell(B,h)\rangle$	Questioner wants to go shopping at Bob's for either turnips or hand soap, and prefers to buy turnips.
γ_0	$\langle add(t) \ \&\ go.to(B) \text{ if } sell(B,t)\rangle$	Questioner wants to buy turnips from Bob's, or else do nothing.

Table 1: Three possible questioner goals. B stands for 'Bob's Market', F for 'Freshtown', t for 'turnips' and h for 'hand soap'. Possible imperatives are *go.to*—go to a given store—and *add*—add a given item to your shopping list.

functions. An action utility function returns a numerical value for each combination of real-world action a (going to Bob's, buying turnips, etc.), goal γ, and current *world state* ω, where the current world state encodes which relevant information is true in the world, i.e. which stores sell which products. Action utility is higher when the questioner's action leads to a preferred outcome. Therefore, every questioner's ultimate goal, in the most general terms, is to maximize action utility. One simple schema for one- and two-member goal tuples, which suffices for current purposes, is given in (7) below, where action $\emptyset$ corresponds to doing nothing, and where the questioner has the option of asking a follow-up question (Benz, 2012).

(7)   For any one- or two-member goal γ of the form $\langle P \text{ if } \Phi,\ (R \text{ if } \Psi)\rangle$, action utility
        $U(\gamma, \omega, a)$ is equal to:
        a.   2 iff $a = P$ and Φ is true in ω
        b.   1 iff $a = R$ and Ψ is true in ω
        c.   1 iff $a = \emptyset$ and $\neg(\Phi \lor \Psi)$ is true in ω
        d. $1 - c$ iff the questioner asks a follow-up question
        e.   0 elsewhere

The highest action utility value, 2, is awarded when the action leads to the questioner's preferred outcome. "Runner-up" outcomes are awarded utility 1. Moreover, doing nothing yields utility 1 in worlds where the questioner's goal is impossible to accomplish; this encodes an avoidance of fool's errands. Finally, the questioner always has the option to ask a follow-up question. To unpack the implications of follow-ups, consider a questioner with goal γ_t. She could ask, "Does Bob's Market sell turnips?", as in (1), and the answerer could simply reply, "no". If the questioner remains uncertain as to whether Freshtown sells turnips, then she can follow up with, "Does Freshtown sell them?" (the sister question in Fig.2). If the answer is 'yes', the questioner will know to go to Freshtown, and if the answer is 'no', the questioner will know to do nothing rather than looking for turnips at either store to no avail. Under the assumption that the answerer will supply a 'yes' or 'no' answer to a follow-up, the outcome is always one that will lead to action utility 1. However, the questioner incurs a small *effort cost* along the way: An additional QA-exchange was required to accomplish the task at hand. We represent this as a small constant c.   Action utilities generated for γ_t, γ_B and γ_0 from Table 1, broken down by possible world state, are given in Table 2.

The questioner wants to maximize action utility, but it is impossible to do this directly, because the questioner does not know the current world state ω. Information about ω must be drawn from the message m which was selected by the answerer. Indeed this is

ω	γ_t			γ_B			γ_0	
$\langle B, F \rangle$	$go.to(B)$	$go.to(F)$	$\emptyset$	$add(t)$	$add(h)$	$\emptyset$	$add(t)$ & $go.to(B)$	$\emptyset$
$\langle t+h, t+h \rangle$	2	1	0	2	1	0	2	0
$\langle t+h, t \rangle$	2	1	0	2	1	0	2	0
$\langle t+h, h \rangle$	2	0	0	2	1	0	2	0
$\langle t+h, \emptyset \rangle$	2	0	0	2	1	0	2	0
$\langle t, t+h \rangle$	2	1	0	2	0	0	2	0
$\langle t, t \rangle$	2	1	0	2	0	0	2	0
$\langle t, h \rangle$	2	0	0	2	0	0	2	0
$\langle t, \emptyset \rangle$	2	0	0	2	0	0	2	0
$\langle h, t+h \rangle$	0	1	0	0	1	0	0	1
$\langle h, t \rangle$	0	1	0	0	1	0	0	1
$\langle h, h \rangle$	0	0	1	0	1	0	0	1
$\langle h, \emptyset \rangle$	0	0	1	0	1	0	0	1
$\langle \emptyset, t+h \rangle$	0	1	0	0	0	1	0	1
$\langle \emptyset, t \rangle$	0	1	0	0	0	1	0	1
$\langle \emptyset, h \rangle$	0	0	1	0	0	1	0	1
$\langle \emptyset, \emptyset \rangle$	0	0	1	0	0	1	0	1

Table 2: Values of action utility function $U(\gamma, \omega, a)$, grouped by questioner's goal, where each world state ω is characterized by which relevant items Bob's and Freshtown sells. An additional action "follow-up question" is always available with utility $1 - c$. World states specify which relevant items Bob's sells, then which relevant items Freshtown sells.

a key function of questions (van Rooij, 2003): to obtain information about the world in order to better maximize utility. Because the link between m and ω is conventional, and not absolute, the hearer encodes the effect of m in a conditional probability distribution over world states. This probability, which encodes a *belief* on the part of the questioner, is conditioned by m as well as by the question that was asked, q, and the questioner's assumption about the answerer's procedure, $\mathcal{S}_M$, for generating messages based on q and ω. The best the questioner can do, then, is to choose actions which maximize utility in the aggregate by maximizing the weighted average of action utilities over possible worlds, weighted by $P(\omega | m, q, \mathcal{S}_M)$. This average is the *expected action utility*.

$$EU(a | \gamma, m, q, \mathcal{S}_M) = \sum_{\omega} P(\omega | m, q, \mathcal{S}_M) \cdot U(\gamma, \omega, a) \tag{1}$$

2.2 Expected utility of answers

We now work backwards from actions to answers, to determine how to strategically optimize answers to YN-questions. Assuming a cooperative Gricean scenario, the answerer wishes above all for the questioner to achieve her immediate goal. At the same time, we can reasonably assume a preference for direct answers, all things being equal. To encode this, we posit a *message utility* function U_M which, for any given outcome of the QA-exchange, is the action utility awarded to questioner for that outcome minus the *cost* of the message, $C(m)$, which we take here to be zero for direct answers, and some small constant k for indirect answers.

$$U_M(\gamma, \omega, m, a) = U(\gamma, \omega, a) - C(m) \tag{2}$$

There are two unknown variables: The answerer does not know the identity of γ or the action a which the questioner will take after receiving m. The expected utility of a message must therefore be calculated by first making inferences about a based on values of γ, and then making inferences about γ based on the question that has been asked, q. The latter inference requires a representation of the questioner's procedure $\mathcal{S}$ for using γ, m and q

to select an action, and of the questioner's procedure $\mathcal{S}_Q$ for using γ to select questions. The resulting formulation is more complex than expected action utility, but it can be boiled down as follows: The expected utility of a message is the weighted average of the message utility of all possible outcomes involving that message, where the probability of an outcome depends on assumptions about how the questioner will react to each message—$P(a|\gamma, m, q, \mathcal{S})$—and about how the question was selected to further the questioner's goal—$P(\gamma|q, \mathcal{S}_Q)$.

$$EU_M(m|\omega, q, \mathcal{S}, \mathcal{S}_Q) = \sum_{\gamma, a} P(\gamma|q, \mathcal{S}_Q) \cdot P(a|\gamma, m, q, \mathcal{S}) \cdot U_M(\gamma, \omega, m, a) \qquad (3)$$

2.3   Expected utility of questions

Again stepping backward through the QA-exchange, we move from the answerer's expected utility of sending a message given a certain question to the questioner's expected utility of posing a question given a certain goal. This also requires an assumed message selection procedure, $\mathcal{S}_M$.

$$EU_Q(q|\gamma, \mathcal{S}_M) = \sum_{\omega, m} P(\omega) \cdot P(m|\omega, q, \mathcal{S}_M) \cdot U(\gamma, \omega, \arg\max_a EU(a|\gamma, m, q, \mathcal{S}_M)) \qquad (4)$$

For any world-message pair, we can calculate the utility the questioner will receive in that world if she takes the action which maximizes EU given that message and $\mathcal{S}_M$. The weighted average for all such world-message pairs is the expected utility of the question. More simply, questions are selected by considering what answers they might elicit, and what best-case outcomes will result from receiving those answers.

We now have three quantities, EU, EU_M and EU_Q, corresponding respectively to the optimality of an action given a question/answer pair and goal, the optimality of an answer given a question and world state, and the optimality of a question given a goal. An equilibrium in this sequential game is one where the questioner chooses a question that maximizes EU_Q, the answerer provides a response that maximizes EU_M and the questioner responds to the answer with an action that maximizes EU, where the players' beliefs, $\mathcal{S}_Q$, $\mathcal{S}_M$ and $\mathcal{S}$, on which the expected utility values depend, reflect the assumption that all players maximize their own expected utility.[2]

2.4   Finding equilibrium

Where the model of Stevens et al. (2015) generates answers by optimizing EU and EU_M, taking the question to be background information, the current model must balance these with EU_Q in order to select questions as well as answers. This complicates the procedure somewhat, but it is nonetheless possible to derive an equilibrium via back-and-forth reasoning (see e.g. Franke, 2011). The idea is to find a sensible default set of beliefs to start with, find optimal selection procedures for those beliefs, then incorporate those procedures into a new set of beliefs, iteratively refining procedures until the iteration converges on a fixed point. We propose the following back-and-forth reasoning schema.

1. Set default procedures $\mathcal{S}^0$ and $\mathcal{S}_Q^0$ to stand in for $\mathcal{S}$ and $\mathcal{S}_Q$, respectively

2. Let $\mathcal{S}_M^1(\omega, q)$ be $\arg\max_m EU_M(m|\omega, q, \mathcal{S}^0, \mathcal{S}_Q^0)$

3. Let $\mathcal{S}_Q^2(\gamma)$ be $\arg\max_q EU_Q(q|\gamma, \mathcal{S}_M^1)$

4. Let $\mathcal{S}^2(\gamma, m, q)$ be $\arg\max_a EU(a|\gamma, m, q, \mathcal{S}_M^1)$

5. Let $\mathcal{S}_M^3(\omega, q)$ be $\arg\max_m EU_M(m|\omega, q, \mathcal{S}^2, \mathcal{S}_Q^2)$

---

[2]The solution concept relevant to these games is 'perfect Bayesian equilibrium' (Fudenberg and Tirole, 1991).

6. Iterate until $\langle S_M^n, S_Q^{n+1}, S^{n+1} \rangle = \langle S_M^{n+2}, S_Q^{n+3}, S^{n+3} \rangle$

The tuple $\langle S_M^n, S_Q^{n+1}, S^{n+1} \rangle$ constitutes an equilibrium. The final piece of the puzzle is to specify $\mathcal{S}^0$ and $\mathcal{S}_Q^0$. Following the idea from Franke (2011) and others of a *naïve hearer*, we propose default procedures for determining the questioner's behavior which do not rely on any reasoning about the answerer's utility. For our model, this means that questions and actions are chosen only in accordance with the questioner's goal. The questioner can ask about her goal directly via a wh-question, or, if the goal is a non-singleton, employ a YN-question as part of an implicit questioning strategy; this selection is made at random. Then, after receiving an answer to the question, the naïve questioner simply acts on the first relevant condition that is met, e.g., if her goal is γ_t, γ_B or γ_0 from Table 2, and the answer is "Bob's sells turnips", she goes to Bob's. These naïve strategies can be formulated as follows, assuming a one- or two-member goal of form $\langle P \text{ if } \Phi, (R \text{ if } \Psi) \rangle$.

$$\mathcal{S}^0(\gamma, m) = \{P\} \text{ if } [\![m]\!] \rightarrow \Phi, \{R\} \text{ if } [\![m]\!] \rightarrow \Psi, \{\emptyset\} \text{ otherwise}$$
$$\mathcal{S}_Q^0(\gamma) = \{q \mid [\![q]\!] \in \{\{\Phi, \neg\Phi\}, (\{\Phi, \Psi, \Phi \wedge \Psi, \neg(\Phi \vee \Psi)\})\}\} \tag{5}$$

We now apply this model to (1), (5) and (6) in Section 1.

3   Explaining the examples

If we work through the iterated reasoning schema given above to simulate contexts for (1), (5) and (6), we find the following:

- The optimal question for a questioner with goal γ_t is always the YN-question, "Does Bob's Market sell turnips?" rather than the broader wh-question, as long as γ_t has a non-zero probability.

- In a world state where Bob's does not sell turnips, but where Freshtown sells turnips and Bob's sells hand soap, the optimal response to the question "Does Bob's Market sell turnips?" is to give the alternative "Freshtown sells turnips..." when $P(\gamma_t)$ is sufficiently high and higher than $P(\gamma_B)$, to give the alternative "Bob's Market sells hand soap" when $P(\gamma_B)$ is sufficiently high and higher than $P(\gamma_t)$, and to give a direct 'no' answer when the two goals are equiprobable.

The easiest way to show this is via computational simulation. We implemented the model in a simplified context where the questioner can have goals γ_t, γ_t' (same as γ_t but with a preference for Freshtown rather than Bob's), γ_B, γ_B' (where hand soap is preferred over turnips), or one of three singleton goals, γ_0^{Bt}, γ_0^{Ft} or γ_0^{Bh}, depending on what item was sought at which store. The questioner can ask about her goal directly, or, for the non-singletons, ask YN-questions which pertain to her goal. The answerer may respond with direct yes/no answers or else convey any potentially relevant proposition of the form $sell(X, y)$. The questioner may respond to the answer with actions $go.to(B)$, $go.to(F)$, $add(t)$, $add(h)$, $add(t)\&go.to(B)$, $add(t)\&go.to(F)$, $add(h)\&go.to(B)$ or $\emptyset$. In order to calculate expected utility, conditional probabilities $P(\omega|m, q, \mathcal{S}_M)$, $P(a|\gamma, m, q, \mathcal{S})$ and $P(m|\omega, q, \mathcal{S}_M)$ were calculated by considering possible outputs of selection procedures, as follows, where the function BOOL takes a proposition and returns 1 if true and 0 if false.

$$P(\omega|m, q, \mathcal{S}_M) = \frac{\text{BOOL}([m \in \mathcal{S}_M(\omega, q)])}{|\{\omega' \mid m \in \mathcal{S}_M(\omega', q)\}|} \tag{6}$$

$$P(a|\gamma, m, q, \mathcal{S}) = \frac{\text{BOOL}([a \in \mathcal{S}(\gamma, m, q)])}{|\mathcal{S}(\gamma, m, q)|} \tag{7}$$

$$P(m|\omega, q, \mathcal{S}_M) = \frac{\text{BOOL}([m \in \mathcal{S}_M(\omega, q)])}{|\mathcal{S}_M(\omega, q)|} \tag{8}$$
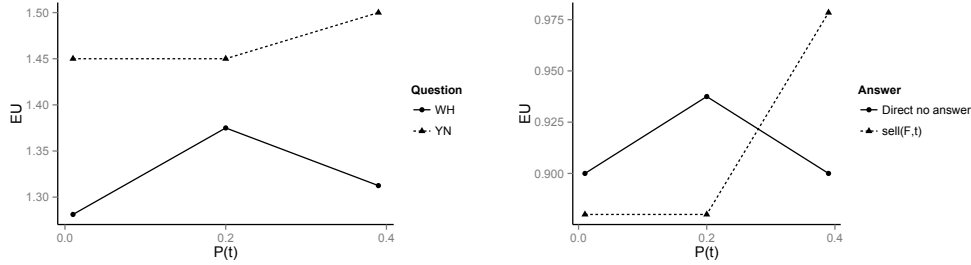
Figure 3: Values of $EU_Q(q|\gamma, \mathcal{S}_M^1)$ for a questioner with goal γ_t deciding between wh-question "Where can I get some turnips?" and YN-question "Does Bob's Market sell turnips?" (left) and values of $EU_M(m|\omega, q, \mathcal{S}^2, \mathcal{S}_Q^2)$ for an answerer having been asked the YN-question in a world state where $sell(B, t)$ is false but $sell(F, t)$ and $sell(B, h)$ are true (right), both plotted by prior probability of the questioner wanting to buy turnips ($P(t) = P(\gamma_t) + P(\gamma_t')$). For illustration, the space of possible goals was limited to those relevant to (1), (5) and (6), with fixed probability 0.2 given to each singleton goal, and with cost terms $c = 0.1$ and $k = 0.02$.

Finally, $P(\gamma|q, \mathcal{S}_Q)$ is calculated via Bayes' rule, where the likelihood term $P(q|\gamma, \mathcal{S}_Q)$ is calculated by considering possible outputs of the assumed question selection procedure.

$$P(\gamma|q, \mathcal{S}_Q) \propto P(q|\gamma, \mathcal{S}_Q) \cdot P(\gamma) \tag{9}$$

$$P(q|\gamma, \mathcal{S}_Q) = \frac{\text{BOOL}([q \in \mathcal{S}_Q(\gamma)])}{|\mathcal{S}_Q(\gamma)|} \tag{10}$$

The graphs in Fig.3 show the expected utility of some possible questions and answers as a function of the prior probability of wanting to buy turnips (γ_t or γ_t'). We assumed that $P(\gamma_t) = P(\gamma_t')$ and that $P(\gamma_B) = P(\gamma_B')$, and reserved a fixed probability for the singletons, such that raising the value of γ_t/γ_t' lowers the value of γ_B/γ_B', and vice versa. Already on the second iteration ($\mathcal{S}^2$, $\mathcal{S}_Q^2$ and $\mathcal{S}_M^3$) we observe the optimal behavior. A questioner with any preferred outcome (e.g., buying turnips from Bob's rather than Freshtown) does best to ask a YN-question about whether that outcome is possible ("Does Bob's Market sell turnips?"). If one particular goal is sufficiently probable, the answerer should provide either a 'yes' answer, if true, or an indirect answer addressing that goal where appropriate ("Freshtown sells turnips..."). Given an indirect answer, the questioner should infer that the preferred outcome is not possible and go for the runner-up (go to Freshtown).

In the end, this reduces the felicity of the answer "Bob's sells hand soap" for (5) and (6) to a threshold on the prior probability of γ_B. The expected utility of that answer will be the mirror image of the expected utility of "Freshtown sells turnips..." in Fig.3: When the prior probability of γ_t is higher than the prior probability of γ_B, "Bob's sells hand soap" is ruled out, but when the reverse is true, it is in fact the optimal answer. General world knowledge tells us that γ_B is a rather odd goal to have—hand soap does not often serve as a practical substitute for turnips—whereas γ_t seems uncontroversial. This is reflected in the prior probability distribution over goals. Thus, in an out-of-the-blue context, (1) is good while (5) is bad. But we can manipulate the prior by explicitly creating a context where γ_B is more likely, as in (6), in that case, it is acceptable to supply hand soap as an alternative to turnips.

References

Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press.

Benz, A. (2012). Errors in pragmatics. *Journal of Logic, Language and Information*, 21(1):97–116.

Benz, A., Jaeger, G., and van Rooij, R., editors (2006). *Game Theory and Pragmatics*. Palgrave MacMillan, New York.

Benz, A. and van Rooij, R. (2007). Optimal assertions, and what they implicate. A uniform game theoretic approach. *Topoi*, 26(1):63–78.

Büring, D. (2003). On D-trees, Beans, and B-accents. *Linguistics and Philosophy*, 26(5):511–545.

Clark, R. (2011). *Meaningful Games*. MIT Press, Cambridge, MA.

de Marneffe, M.-C., Grimm, S., and Potts, C. (2009). Not a simple yes or no. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 136–143.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantic and Pragmatics*, 4:1–82.

Fudenberg, D. and Tirole, J. (1991). Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2):236–260.

Groenendijk, J. and Stokhof, M. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam.

Lai, C. (2012). *Rises all the way up: The interpretation of prosody, discourse attitudes and dialogue structure*. PhD thesis, University of Pennsylvania.

Parikh, P. (2010). *Language and Equilibrium*. MIT Press, Cambridge, MA.

Roberts, C. (1996). Information Structure: Towards an integrated formal theory of pragmatics. In Yoon, J. and Kathol, A., editors, *OSUWPL Volume 49: Papers in Semantics*. The Ohio State University.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1:75–116.

Schwarzschild, R. (1999). Givenness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics*, 7:141–177.

Stevens, J., Benz, A., Reuße, S., and Klabunde, R. (2015). A strategic reasoning model for generating alternative answers. *Proceedings of the 53rd Annual Meeting of the ACL*, pages 534–542.

van Rooij, R. (2003). Questioning to resolve decision problems. *Linguistics and Philosophy*, 26(6):727–763.

Wagner, M. (2012a). Contrastive topics decomposed. *Semantics and Pragmatics*, 5(8):1–54.

Wagner, M. (2012b). Focus and givenness: A unified approach. In Kučerová, I. and Neeleman, A., editors, *Information Structure: Contrasts and Positions*. Cambridge University Press, Cambridge.

On the focus-sensitivity of counterfactuality

Jos L. Tellings*

University of California, Los Angeles
Los Angeles, CA
j.tellings@gmail.com

Abstract

I show that the presence of the additive particle 'also' can have the effect of cancelling the counterfactuality of the consequent in counterfactual conditionals, but only when 'also' focus-associates with material inside the antecedent. Adopting Karttunen's (1971) idea that *conditional perfection* (CP; the strengthening of conditionals to biconditionals) is a necessary ingredient of the counterfactual inference of the consequent (which I write 'CF_q'), it follows that when CP is blocked, CF_q is also blocked. I show that when 'also' associates with material in the antecedent of the conditional, it reflects a context in which multiple causes for the same consequent are salient (a *multiple-cause context*). I then argue that multiple-cause contexts block CP. This gives rise to a more general characterization of the set of contexts that block CF_q: both other realizations of multiple-cause contexts (that do not contain 'also'), and contexts that block CP for other reasons are included in this set.

1 Introduction

Counterfactual conditionals typically come with inferences that their antecedent is false and that their consequent is false. Writing the antecedent of a conditional as p, and the consequent as q throughout, I will refer to the corresponding counterfactual inferences as CF_p and CF_q, respectively:

(1)   If John had taken the bus, he would have been on time.
    $\mapsto CF_p$: John didn't take the bus
    $\mapsto CF_q$: John wasn't on time

In most of the literature on counterfactuality, attention is focused exclusively on CF_p, perhaps under the implicit assumption that what has been said about CF_p extends, mutatis mutandis, to CF_q as well. In this paper I will show that this implicit assumption is not warranted, and the empirical phenomena regarding CF_q are distinct from those CF_p, and therefore call for a different theoretical explanation.

Two major questions regarding counterfactual inferences concern their *source* (English has no dedicated counterfactual marker, so how is counterfactuality realized?) and their *cancellability*. This last property relates to the observation that the counterfactuality of p is not an entailment or a presupposition, because it can be cancelled in certain contexts. The most well-known example is due to Anderson (1951):

(2)   If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show. [So, it is likely that he took arsenic].

Here CF_p gets cancelled, but CF_q can also be cancelled, and crucially it can be cancelled independently from CF_p. This has has been noted at a few places in the literature, but has not been analyzed formally; I cite here an example from Declerck and Reed (2001:266):

(3)　　A:　　We are in time because we have taken the road I said we should take.
　　　　B:　　If we'd taken the other road, we would **also** have been here in time.
　　　　B′:　　#If we'd taken the other road, we would have been here in time.

Here CF_p is triggered in the regular fashion (we did *not* take the other road), but CF_q is cancelled (we *were* in fact in time). Reply B′ shows that the lexical item 'also' is in some way responsible for the cancellation of CF_q: leaving it out leads to a conditional that triggers CF_q in the normal fashion, and the contradiction between A's utterance (we are on time) and CF_q (we are not on time) makes it infelicitous.[1]

The presence of 'also' is not the only way to cancel CF_q (we will see more examples further on), nor do all instances of 'also' in the consequent of a subjunctive conditional have the effect of cancelling CF_q. This latter observation is an important empirical point, and in section 2 I will show that whether or not CF_q is cancelled depends on how 'also' associates with focus. This illustrates how the generation of CF_q is a focus-sensitive phenomenon, although I will reach a more general conclusion: the examples with 'also' will turn out to be just one instance of a more general characterization of contexts in which CF_q is cancelled, which depends on the information-structural context in which the conditional is uttered (section 3). In section 4 I will adopt a view, originally due to Karttunen (1971), in which *conditional perfection* (the pragmatic strengthening of conditionals to biconditionals; abbreviated CP) is a necessary ingredient for CF_q to be triggered. This means that when CP is absent for some reason or other, CF_q does not arise either. This is precisely the explanation I will provide in section 5: what the various CF_q-cancelling contexts have in common is that they are contexts that block CP. Some of these contexts have already been described as unperfectable contexts in the literature on CP, but I also propose a novel characterization of a set of contexts that blocks CP. I argue that CP is blocked in contexts that make several distinct causes (antecedents) for the same consequent salient, which I will refer to as *multiple-cause contexts*. The focus-association of 'also' relates to the information-structure of the context surrounding the conditional (e.g. which question under discussion it answers), and thus whether that context is multiple-cause or not. This solves the puzzle from section 2 of how blocking of CF_q can be sensitive to the focus-association of 'also'.

The focus-sensitivity of CF_q is thus different in nature from other cases in which focus interacts with counterfactual conditionals, such as the so-called Dretske-counterfactuals (Dretske 1972) and the cases discussed in Ogihara (2000). In those cases, focus plays a semantic (truth-conditional) role in restricting the set of worlds over which the conditional quantifies, while the type of focus-sensitivity I discuss in this paper is more pragmatic in nature: it reflects a certain information-structural environment that blocks conditional perfection, and hence also cancels the generation of CF_q.

2　Local and non-local 'also'

An important empirical observation is that not all instances of 'also' in the consequent of a subjunctive conditional have the effect of cancelling CF_q. In order to see this I introduce a

---

[1]As Declerck and Reed note, for many speakers the word 'still' can be used instead of 'also' in these contexts. It is surprising that 'still' and 'also' in this context appear to have very similar meanings, while in ordinary sentences the additive particle and (aspectual) 'still' diverge in meaning. I take this puzzle up in Tellings (2016).

context in which causal relations are made very explicit. This will help facilitate the somewhat subtle judgments, but examples of this kind commonly occur in natural speech.[2]

(4) **The Quiz Scenario**
Mary participates in a quiz show, in which she blindly opens one of the following four boxes, after which she wins its contents.

Box A	Box B	Box C	Box D
$100	$100 + laptop	empty	laptop

Suppose that Mary participated in this quiz, and she opened Box A, thus winning $100. After all the boxes are opened and their contents have been revealed, the following two statements are felicitous.

(5) [context: Mary opened Box A, so she won $100.]
    a.   If Mary had opened Box B, she would ALSO have won $100.
    b.   If Mary had opened Box B, she would also have won a LAPtop.

Focal stress, here indicated by capital letters is crucial in the assessment of these examples. The key point is that there is a difference with regard to CF_q in the pair in (5). In (5a), CF_q is cancelled: Mary did in fact win $100, so here 'also' has the same effect as in my first example in (3). In (5b), on the other hand, CF_q is not cancelled but triggered in the normal fashion (as in (1)): Mary did not win a laptop. In addition, (5b) becomes infelicitous when the context makes the consequent true (so that Mary did win a laptop):

(6) [context: Mary opened **Box D**, so she won a laptop.]
    #If Mary had opened Box B, she would also have won a LAPtop.   (= (5b))
    (possible response: . . . Wait a minute! She DID win a laptop!)

I claim that the difference between (5a) and (5b) lies in how 'also' associates with focus. As is well known, additive particles such as 'also' and 'too' have a focus-marked **associate** as well as a **presupposed alternative**, which is often thought to be anaphoric.

(7) I invited $Bill_i$ to the party, and I $also_i$ invited MARY.
    associate of 'also' = Mary, presupposed alternative of 'also' = Bill

One way to identify the associates of 'also' in (5a) and (5b) is by considering the different meanings the sentences express. Sentences (5a) and (5b) make different counterfactual statements: in (5a) an additional hypothetical way of winning $100 is provided (opening Box B in addition to opening Box A), while (5b) provides an additional hypothetical prize (a laptop in addition to $100). Since the focus-marked associate is the item for which alternatives are taken into consideration, the associate of 'also' in (5b) is 'a laptop'. This intuition is further confirmed by intonation: the associate of 'also' receives focal stress.

In (5a) on the other hand, the relevant alternatives are opening different boxes (Box A, Box B, . . .), so the associate of 'also' is the phrase 'Box B' which is syntactically located outside the consequent clause. This corresponds with a different intonation: focal stress in (5a) is on the focus particle itself, a pattern that is typical of postposed additive particles (Krifka 1999, to be discussed in section 3 below).

---

[2]Example (3) and the one in Iatridou (2000:232n) are examples from the literature; see Tellings (2016) for several examples from corpora.

Another way to state the difference is that in (5b), the particle 'also' is part of the proposition q (q = 'Mary also won a LAPtop'), but in (5a) it is not (q = 'Mary won \$100'). Therefore I will refer to 'also' in (5b) as **local** 'also' (since 'also' and its associate are clause-mates), and in (5a) as **non-local** 'also' (as it associates with material in a different clause).[3] The generalization is as follows:

	example	CF_q behavior	associate of 'also'
local 'also'	(5b)	does not block CF_q	in the consequent
non-local 'also'	(5a)	blocks CF_q	in the antecedent

This sets up the main puzzle: how can a difference in focus-association affect the generation of counterfactual inferences?

Since the conditionals with local 'also' behave like regular counterfactuals w.r.t. CF_q, the puzzle only concerns the non-local instances of 'also' (of course we will make sure that the explanation for cancellation with non-local 'also' I will present does not overgenerate to local 'also'). Indeed, the semantic behavior of local 'also' is correctly captured by existing theories. Since 'also' only contributes presuppositional content, we only need to worry about the behavior of that presupposition, and can remain neutral about the semantics of conditionals. The presupposition of 'also' triggered in the consequent becomes an unconditional presupposition, requiring that in the actual world Mary won something else than a laptop (in this case anaphorically satisfied by '\$100').[4]

(8)   *Local 'also'*
       ⟦if Mary had opened Box B, she would also have won a LAPtop⟧^{w} is defined if Mary won something else than a laptop in w;
       when defined, it is true if ⟦if Mary had opened Box B, she would have won a laptop⟧^{w} is true (defined in the reader's favorite semantics for counterfactuals)

Now turning our attention to non-local instances of 'also', first there is a syntactic issue to be taken care of.

Syntax   In the surface structure of (5a), non-local 'also' does not c-command its associate because the latter is in a different (higher) clause, violating a generally assumed requirement for focus particles. I argue that this requirement is still met for (5a), but at an earlier level in its syntactic derivation. The problem with c-command does not arise in sentences with a sentence-final *if*-clause. To see this, I adopt some basic assumptions regarding the syntactic structure of conditionals (Bhatt and Pancheva 2006) and focus particles. In particular, a sentence-final *if*-clause is a VP-adjunct, and the particle 'also' adjoins to VP as well (see e.g. Rullmann 2003). There are then three VPs: the smallest VP is that of the main clause (the consequent; VP_1), to VP_1 the *if*-clause right-adjoins (forming VP_2), and to VP_2 'also' left-adjoins forming VP_3.

---

[3]For reasons of exposition in (5a) the associate of 'also' is a proper part of the antecedent, but this is not a requirement. Suppose the quiz is such that you can win \$100 either by opening Box A or by answering a special bonus question. Then (i) would be another example of non-local 'also', with a larger associate:

(i)   If Mary had [answered the BOnus question], she would ALSO have won \$100

Thanks to an anonymous reviewer for bringing this point up.

[4]The 'proviso problem' (the question whether or not an embedded presupposition becomes unconditional) for counterfactual conditionals remains a difficult issue (see e.g. Lassiter 2012). There are additional data with additive particles in conditionals that present a complicated picture of presupposition projection behavior, that I cannot discuss here for reasons of space (but see Tellings 2016 for further discussion). As far as I am aware, existing theories (based for example on probabilistic independence) cannot explain the full paradigm.

(9)   Mary would [$_{VP_3}$ ALSO [$_{VP_2}$ [$_{VP_1}$ have won \$100] [if she had opened Box B]]].

In this configuration 'also' c-commands the antecedent and all material inside it.

It has been argued for independently that sentence-initial *if*-clauses are derived syntactically from underlying sentence-final *if*-clauses by movement. For example, it has been suggested as a solution for a very similar problem in which c-command relationships are required between a reflexive or a quantifier in the consequent and its antecedent inside the *if*-clause (Bhatt and Pancheva 2006:650).

(10)   a.   If pictures of himself$_i$ are on sale, John$_i$ will be happy.
        b.   If her$_i$ child is late from school, every mother$_i$ is upset.

Since (9) involves two VP-adjuncts, they can combine with the main clause VP in two different ways. This gives rise to different c-command domains of 'also'. The difference between local and non-local 'also' can thus be characterized syntactically as a difference in adjunction height. This is illustrated for the sentence-initial *if*-clauses in (5), together with the movement steps by which they are derived:

(11)   a.   [if . . .]$_i$ [Mary would [$_{VP_3}$ ALSO [$_{VP_2}$ [$_{VP_1}$ have won \$100] t_i]]]
              (underlying form of (5a); non-local 'also'; 'also' c-commands antecedent)
        b.   [if . . .]$_i$ [Mary would [$_{VP_3}$ [$_{VP_2}$ also [$_{VP_1}$ have won a LAPtop]] t_i]]
              (underlying form of (5b); local 'also'; 'also' only c-commands consequent)

I now move on to discussing how the presence of non-local 'also' is highly informative in terms of the information structure that surrounds the conditional utterance. This will involve the notion of *contrastive topic*.

3   Contrastive topic

Contrastive topic (CT) is an information-structural entity that has a certain prosodic realization, and a special semantic contribution. Prosodically, CT is marked by a B-accent (Jackendoff 1972), or a rise-fall-rise (L+H* L-H%) contour (this is a somewhat simplified statement, see Constant 2014 for phonetic details). Semantically, CT is used to convey that an answer is only a partial answer to a question (Büring 2003; Constant 2014). For example, in (12), the CT-marking on 'John' expresses that it is only a partial answer, i.e. that the normal exhaustive interpretation (that Mary did not drink wine) is not obtained.

(12)   Q: Did John and Mary drink wine at the party?
        A: [John]$_{CT}$ did. . .

The use of CT-marking comes with a constraint on the repetition of the same focus-marked constituent in a CT-Focus construction (Krifka 1999). This is illustrated in (13a):

(13)   Q: Where do John and Mary live?
        a. #A : [John]$_{CT}$ lives in [France]$_F$, and [Mary]$_{CT}$ lives in [France]$_F$.
        b.   A′: [John]$_{CT}$ lives in [France]$_F$, and [Mary]$_{CT}$ [also]$_F$ lives in France.

Although there is no established name for this constraint as far as I know, I will refer to it as the 'Repeated Focus Constraint' (RFC) for ease of reference.

(14)   *Repeated Focus Constraint* (RFC)   (informal)
         $\#[\ldots \text{CT}_1 \ldots \text{Foc}_1 \ldots] \land [\ldots \text{CT}_2 \ldots \text{Foc}_2 \ldots]$ with $\text{Foc}_1 = \text{Foc}_2$

The RFC is of interest to our purposes because the violation of it in (13a) can be rescued by using a stressed additive particle such as 'also', as in (13b). Because in these cases 'also' receives focal stress, deviating from the canonical pattern in which (only) the associate is focus-marked (as in (7)), these are known as **stressed postposed additive particles** (Krifka 1999; Sæbø 2004).[5] Non-local 'also' as in (5a) is also stressed, so (5a) and (13a) deviate in the same way from the canonical intonation pattern of additive focus particles. This suggests that they are instances of the same phenomenon, and that non-local 'also' should be analyzed in the theory of stressed additive particles. Below I give additional arguments that further support this idea.

First, we observe that RFC as exemplified for simple monoclausal sentences in (13) extends to more complex syntactic structures such as conditionals. Here are conditional parallels to (13), with the exact same judgments as (13):

(15)   What would have happened if I had opened Box A or Box B?
         a. #If you had opened [Box A]$_{\text{CT}}$, you'd have won [\$100]$_{\text{F}}$, and if you had opened [Box B]$_{\text{CT}}$, you'd have won [\$100]$_{\text{F}}$.
         b. If you had opened [Box A]$_{\text{CT}}$, you'd have won [\$100]$_{\text{F}}$, and if you had opened [Box B]$_{\text{CT}}$, you'd [also]$_{\text{F}}$ have won \$100.

Second, Krifka (1999:125) points out that postposed stressed focus particles have to be additive, they cannot be exclusive or scalar:

(16)   John lives in France ALSO/TOO/*ONLY/*EVEN.

The same restriction holds for non-local 'also' in conditionals: in (5a), 'ALSO' cannot be replaced by stressed 'ONLY' or 'EVEN'.

Another relevant and important property is that the RFC is specific to CT-Focus constructions, since the exact same string of words as in (13a) is licensed when it answers a different question (and does not have CT-marking):

(17)   Q: Who live in France?
         A: [John]$_{\text{F}}$ lives in France, and [Mary]$_{\text{F}}$ lives in France.

This has led to the hypothesis that stressed postposed additive particle associate with CT:

(18)   *Contrastive topic hypothesis*   (Krifka 1999)
         The associated constituent of stressed postposed additive particles is the contrastive topic of the clause in which they occur.

Thus besides the empirical parallels described above, we need to show that non-local 'also' in (5a) associates with a contrastive topic.[6]

First, in order to argue that sentences like (5a) indeed involve CT-marking, we need to appeal to the information-structural notion of CT. Observe that just like a conjoined question can have a partial answer by answering one of its component questions (as we saw in (12)), we can construct a conjoined conditional question with 'or' in the antecedent. A question of the

---

[5]They are called 'postposed' because it was thought that the focus particle receives stress whenever it linearly follows its associate, see Krifka (1999). My examples of non-local 'also' in sentence-final *if*-conditionals (see (9)) provide a direct counterexample to this alleged generalization.

[6]For reasons of space I cannot go into how Krifka's (1999) theoretical account can be extended to conditionals. In Tellings (2016) I show how some modifications are required, but the spirit of his analysis can be maintained.

form 'What if p or r?' is a consequent-based *wh*-question on two alternative antecedents (p and r). Partial answers take the form 'If $[\ldots p_{CT} \ldots], [\ldots q^1{}_F \ldots]$' and 'If $[\ldots r_{CT} \ldots], [\ldots q^2{}_F \ldots]$' to express $p \rightarrow q^1 \wedge r \rightarrow q^2$. This is illustrated in (19) (a direct parallel to (12)):

(19)   What would have happened if you opened Box A or Box B?
         If I had opened [Box A]$_{CT}$, I would have [won \$100]$_F$ ...

This is what happens in (5a) too. In the context it is given that Mary won \$100 by opening Box A. The question under discussion then is what would have happened if she had opened Box B. Sentence (5a) (with non-local 'also') is then a congruent answer, with CT-accent on 'Box B'. Stressed 'also' is required to rescue the RFC: another conditional with the same focus-marked consequent (winning \$100) was already salient.

Second, it has been argued elsewhere on independent grounds that conditionals set up a natural CT-inducing discourse. For example, Constant (2014) writes that "considering one hypothetical possibility almost inevitably leads to questions about contrasting possibilities" (Constant 2014:321), i.e. it is natural to contrast a hypothetical situation with alternatives. These contrast with factive clauses, such as for example *because*-clauses, as "there is no corresponding option of contrasting polar opposite *because*-clauses" (p. 323). These intuitions are supported by data of the following sort (p. 324):

(20)   a.   Because it is raining, we'll have to cancel the picnic. #And because it is not?
         b.   If it is raining, we'll have to cancel the picnic. And if it is not?

Indeed, using stressed 'also' in a *because*-clause is infelicitous:

(21)   #Because it's really warm today, we go swimming. And because the pool is open, we ALSO go swimming.

These data illustrate that asking about alternative or additional antecedents in a conditional is possible, but asking about other reasons in a *because*-clause is not.

The conclusion of the above is that non-local 'also' as in (5a) indicates the presence of a particular discourse structure involving contrastive topic in the antecedent. We have seen that this corresponds to alternative causes for the consequent q.[7] Therefore I will refer to such contexts that make multiple causes salient as **multiple-cause contexts**. I pointed out in (17) that a repetition of foci can be licensed if there is no CT-marking involved. This predicts that in the domain of conditionals, there are multiple-cause contexts that do not require the presence of stressed 'also'. This is important because, anticipating somewhat, my theory will correctly predict that in those contexts CF_q gets cancelled too. I will mention two cases here. First, (17) itself has a conditional parallel:

(22)   [context: A incorrectly assumes that only box A contains \$100]
         A: I didn't win \$100. But if I had opened Box A, I would have won \$100.
         B: If you had opened [Box A]$_F$, you would have won \$100, and if you had opened [Box B]$_F$, you would have won \$100.

Second, the string in (17) can be licensed in a 'listing context', a list in which each conjunct has a H-L% intonation contour. In (23) I give both the simple and the conditional case:

---

[7]The criteria for when causes count as distinct are the same as those for when focus alternatives for 'also' are distinct. So for example the presupposed alternative should not entail the associate of 'also':

(i)   #If you'd eaten fruit, you'd have been healthy, and if you'd eaten an apple, you'd have ALSO been healthy.

(23)   a.   John lives in France$_{\text{H-L\%}}$, Bill lives in France$_{\text{H-L\%}}$, Mary lives in France$_{\text{H-L\%}}$, . . .
         b.   [context: A played a quiz in which 8 of the ten boxes contain $100]
               A: I won $100! I am so happy.
               B: Well, that wasn't so hard really: if you had opened Box A, you would have won
               $100, and if you had opened Box B, you would have won $100, and if you had
               opened Box C, you would have won $100, . . .   [with listing intonation, H-L%]

4   The source of CF$_q$

Before we can talk about the link between focus-association, multiple-cause contexts and the
generation of counterfactual inferences, we need to spend a few words on how these inferences
are generated in the first place. As I noted in the introduction, most attention has been directed
toward CF$_p$. For English, a language that does not have a dedicated counterfactuality marker,
two broad groups of accounts for the source of CF$_p$ can be distinguished: accounts in which
CF$_p$ is the result of a modal interpretation of a past tense morpheme in the antecedent (a 'fake
past', e.g. Iatridou 2000), and accounts in which CF$_p$ arises from a Gricean competition of
presuppositions that conditionals have (e.g. Ippolito 2006; Leahy 2011).

One might hope that these accounts for CF$_p$ extend to CF$_q$ with only minor modifications,
but for both types of theories some serious technical and empirical problems arise. For reasons of
space I cannot discuss these issues here, but see Tellings (2016) for more details. An alternative
option is to say that CF$_q$ is the result of CF$_p$ plus the special semantic relationship between p
and q that is conveyed by the conditional. This is what Karttunen (1971) suggests, and it is
the approach I will adopt. In particular, Karttunen relates CF$_q$ to conditional perfection (i.e.,
strengthening $p \rightarrow q$ to $\neg p \rightarrow \neg q$, discussed further in section 5). Then, CF$_p$ plus conditional
perfection results in CF$_q$ via Modus Ponens.

(24)   Utterance: $p \rightarrow q$ (Karttunen 1971)
         Implicatures:

$$\frac{\begin{array}{ll} \neg p & \text{(counterfactuality of } p) \\ \neg p \rightarrow \neg q & \text{(conditional perfection on } p \rightarrow q) \end{array}}{\neg q \qquad\quad (=\text{CF}_q; \text{ by Modus Ponens})}$$

An important prediction that this account makes, not mentioned by Karttunen, is that when
conditional perfection is not triggered in a context for some reason or other, we predict that
there is no CF$_q$ either in that context. This is exactly the type of explanation I will provide.

5   Conditional perfection

Conditional perfection (CP) refers to the pragmatic phenomenon that conditionals are strength-
ened to biconditionals, i.e. the strengthening of $p \rightarrow q$ to $\neg p \rightarrow \neg q$:

(25)   If you mow the lawn, I will give you $5.   (Geis and Zwicky 1971)
         implicature: if you don't mow the lawn, I will not give you $5

CP is restricted in the sense that various groups of conditionals have been shown in the liter-
ature not to trigger CP (e.g. Horn 2000). These include for example semifactuals and biscuit
conditionals. My account predicts that these types of conditionals do not trigger CF$_q$, and this
is borne out. Both semifactuals and biscuit conditionals convey the truth of their consequent,
and we find that the lack of CP coincides with the lack of CF$_q$:

(26)   a.   Even if John had studied all day, he wouldn't have passed.
              no CP: $\not\rightarrow$ If John hadn't studied all day, he would have passed
              no CF_q: $\not\rightarrow$ John did pass.
        b.   If you had been hungry, there would have been cookies in the cupboard.
              no CP: $\not\rightarrow$ If you hadn't been hungry, there wouldn't have been cookies
              no CF_q: $\not\rightarrow$ There are no cookies in the cupboard

Besides these groups of conditionals, we are most interested in the focus-sensitive behavior of 'also' in blocking CF_q. I showed that non-local 'also' indicates a multiple-cause context, but that there are other ways to realize such a context as well. It remains to be shown that multiple-cause contexts block CP. It is an intuitive idea that when multiple causes $p_1 \rightarrow q$, $p_2 \rightarrow q$, ... are salient, the inference $\neg p_1 \rightarrow \neg q$ is no longer made. Moreover, at various places in the literature the context-sensitivity of CP has been addressed, and often these claims are based (without explicitly mentioning so) on there being additional causes for the consequent; for example, CP fails in (25) in contexts in which there are several ways to earn \$5 (von Fintel 2001). There is also experimental evidence, for instance Politzer (2003) discusses studies that show that speakers are less likely to make the fallacy of denying the antecedent when other antecedents with the same consequent are salient.

In addition to these informal pieces of corroboration, we would like to derive the result within existing pragmatic theories for conditional perfection. CP has a long history in pragmatic theory (van der Auwera 1997). Many of the earlier accounts attempt to derive CP as a Gricean quantity or relevance implicature, and although these theories have mostly been superseded by theories that incorporate discourse structure (to be discussed below), it is worth pointing out that some of them already incorporated the intuition that multiple-cause contexts block CP (for example Van der Auwera suggests CP is a scalar implicature based on the scale $\langle \ldots,$ 'if p, q and if r, q and if s, q', 'if p, q and if r, q', 'if p, q'$\rangle$, which explicitly contains multiple-cause contexts).

More recent theories of CP not only aim to explain the mechanism that generates the implicature but also the contextual restrictions that it is subject to. Summarizing a lot of work, many proposals suggest that perfection arises when the conditional is an *exhaustive* answer to the question under discussion (von Fintel 2001; Herburger 2015). For example, Herburger suggests that an exhaustive conditional answer is interpreted as 'if p, q and only if p, q' (with the second conjunct not pronounced). Her semantics for 'only if' then yields the CP interpretation. In these types of accounts for CP, it readily follows that multiple-cause contexts block CP: as discussed in section 3, in such contexts the conditional utterance is only a partial, and not an exhaustive answer to the question under discussion.

I closed section 3 with (22) and (23b) as examples of multiple-cause contexts without 'also'. My theory predicts that they nevertheless block CF_q, and this prediction is borne out: in both cases the context makes it true that the speaker won \$100, yet the conditional utterance is felicitous. The analysis of multiple-cause contexts has thus given us a wider characterization of contexts that cancel CF_q than just the data containing non-local 'also'.

6   Conclusion

The relation between the focus-association of 'also' in the consequent of a conditional and the cancellation of CF_q is an indirect one: when 'also' associates with material in the antecedent (i.e., non-local 'also') the alternatives are distinct causes for the consequent. Such a multiple-cause context blocks CP. When 'also' on the other hand associates locally, the alternatives (in my example) are different prizes. The context then presents one cause for different prizes, so

we do not have a multiple-cause context, and CF_q is triggered in the normal fashion. Because the presence of 'also' is only one indicator of a multiple-cause context, my analysis gave way to a wider characterization of contexts that block CF_q. These include both unperfectable conditionals such as semifactuals, and other types of multiple-cause contexts such as the ones marked by listing intonation or exhaustive focus.

The picture that arises from this is that counterfactuality should not be analyzed at the level of single utterances, but is rather a phenomenon that fundamentally interacts with the surrounding discourse and its topic-focus structure. I have shown how combining insights from theories on the semantics, pragmatics, and information-structure of conditionals can help analyze empirical generalizations. This encourages the development of a more integrated theory of conditionals, which will aid in solving other open problems in the area of conditionals.

References

Anderson, A. (1951). A note on subjunctive and counterfactual conditionals. *Analysis*, 11:35–38.

van der Auwera, J. (1997). Pragmatics in the last quarter century: The case of conditional perfection. *Journal of Pragmatics*, 27:261–274.

Bhatt, R. and Pancheva, R. (2006). Conditionals. In Everaert, M. and van Riemsdijk, H., editors, *The Blackwell Companion to Syntax*, pages 638–687. Blackwell, Oxford.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and Philosophy*, 26(5):511–545.

Constant, N. (2014). *Contrastive topic: meaning and realizations*. PhD thesis, UMass Amherst.

Declerck, R. and Reed, S. (2001). *Conditionals. A Comprehensive Empirical Analysis*. De Gruyter.

Dretske, F. I. (1972). Contrastive Statements. *The Philosophical Review*, 81(4):411–437.

von Fintel, K. (2001). Conditional strengthening: a case study in implicature. Ms., MIT.

Geis, M. and Zwicky, A. (1971). On invited inferences. *Linguistic Inquiry*, 2(4):561–566.

Herburger, E. (2015). Conditional Perfection: The truth and the whole truth. SALT 25 handout.

Horn, L. R. (2000). From *if* to *iff*: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics*, 32:289–326.

Iatridou, S. (2000). The Grammatical Ingredients of Counterfactuality. *Linguistic Inquiry*, 31:231–270.

Ippolito, M. (2006). Semantic Composition and Presupposition Projection in Subjunctive Conditionals. *Linguistics and Philosophy*, 29:631–672.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press, Cambridge, MA.

Karttunen, L. (1971). Counterfactual Conditionals. *Linguistic Inquiry*, 2(4):566–569.

Krifka, M. (1999). Additive Particles under Stress. In Strolovitch, D. and Lawson, A., editors, *Proceedings of SALT 8*, pages 111–128. CLC Publications, Ithaca, NY.

Lassiter, D. (2012). Presuppositions, provisos, and probability. *Semantics & Pragmatics*, 5(2):1–37.

Leahy, B. (2011). Presuppositions and Antipresuppositions in Conditionals. In Ashton, N., Chereches, A., and Lutz, D., editors, *Proceedings of SALT 21*, pages 257–274. Rutgers University.

Ogihara, T. (2000). Counterfactuals, Temporal Adverbs, and Association with Focus. In Jackson, B. and Matthews, T., editors, *Proceedings of SALT 10*, pages 115–131. Cornell University.

Politzer, G. (2003). Premise interpretation in conditional reasoning. In Hardman, D. and Macchi, L., editors, *Thinking: Psychological perspectives on reasoning, judgment, and decision making*, pages 79–93. Wiley, Chichester.

Rullmann, H. (2003). Additive Particles and Polarity. *Journal of Semantics*, 20:329–401.

Sæbø, K. J. (2004). Conversational Contrast and Conventional Parallel: Topic Implicatures and Additive Presuppositions. *Journal of Semantics*, 21(2):199–217.

Tellings, J. L. (2016). *Counterfactuality in discourse*. PhD thesis, UCLA. To appear July 2016.

Locating Hidden Quantifiers in *De Re* Reports[*]

Daniel Tiskin

Saint Petersburg State University
daniel.tiskin@gmail.com

Abstract

The present paper is a contribution to the debate on the nature of quantification over modes of presentation in *de re* attitude reports. I show in Section 3 that the "universal" readings of *de re* reports with the quantifier *no* cannot result from a special "universal" reading of the attitude verb, contra a recent proposal by Charlow and Sharvit [3] reviewed in Section 2. Then I outline an account which has the advantages of Charlow and Sharvit's and of Santorio's [15] accounts, but in addition can handle "universal" readings (Section 4). The account is then shown to extend to another class of examples [3] discusses (Section 5).

The main feature of my account is that the attitude verb does not bind concept generator variables; this work is relegated to the special operator, crucially lower in the structure.

1   The concept generator theory

A prominent problem in the semantics of *de re* attitude reports is that of *double vision*. As already Quine [12] and Kaplan [6] noticed, in case an attitude holder is acquainted with a given object in more than one way, there is no contradiction in (1) and (2) being simultaneously true.

(1)   Ralph believes that Ortcutt is a spy.

(2)   Ralph believes that Ortcutt is not a spy.

Kaplan's solution is to quantify existentially over "vivid names" Ralph has for Ortcutt. Assuming that he has two if he has encountered Ortcutt twice and does not identify the person he met on those two occasions, we indeed get non-contradictory denotations for (1) and (2):

(1′) $\exists n_1(\textsc{VividName}(n_1, \mathbf{r}, \mathbf{o}) \wedge \forall w \in \textsc{dox}(\mathbf{r}, @).\lceil \mathbf{spy}_w(n_1) \rceil \equiv \mathbf{1}$

(2′) $\exists n_2(\textsc{VividName}(n_2, \mathbf{r}, \mathbf{o}) \wedge \forall w \in \textsc{dox}(\mathbf{r}, @).\lceil \neg\mathbf{spy}_w(n_2) \rceil \equiv \mathbf{1},$

where $\textsc{dox}(\mathbf{r}, @)$ is the set of Ralph's doxastic alternatives at the actual world @. Whereas Kaplan's idea can be given a more epistemic twist (in terms of acquaintance functions instead of vivid names) and—at some cost—a compositional implementation [21, 20], there is something it cannot (straightforwardly) account for (e.g. bound *de re*, see below).

1.1   *De re* using concept generators

A more powerful theory of acquaintance-based *de re* ascription which allows for *in situ* compositionality was pioneered by Percus and Sauerland [10]. Their primary goal was to develop a plausible analysis for *de se* and *de re* attitude reports, but later "bound *de re*" readings were discovered, which can be successfully analysed using the means they devised.

Percus and Sauerland presented a strong argument in favour of there being a structural difference between *de se* and *de re* reports. To demonstrate this, they used what I shall call the *only*-test, which will also be employed in a key argument in Section 3 of the present paper; so let me introduce it in some detail. First, consider the following scenario.

---

[*]Thanks to Manuel Križ, Jan Wiślicki and two anonymous reviewers for AC 2015. All errors are my own.

Scenario 1. *John, Bill and Sam are running for president. Once they all get drunk and sit down to watch the debates. John fails to recognise himself on the screen but still thinks: "I will win". Bill does not recognise himself and says, pointing to his own image, "This guy will win and I will lose". Sam recognises only John and thinks: "That guy will win".*

In this scenario each of John and Bill has some kind of acquaintance-based mode of presentation i of himself s.t. he believes the value of f_i[1] will win the election. However, (3) has a true reading in Scenario 1.

(3)   Only John$_i$ thinks that he$_i$ will win the election.

Why is *only* "insensitive" to Bill's also having a relevant belief? The reason is, presumably, that the part of the sentence following *only John* is ambiguous, and one of the LFs corresponding to it gets interpreted as a condition only John satisfies. Therefore, a *de se* report differs from the corresponding *de re* reports not only in the sort of the acquaintance function evoked but also in the LF it has; *only* generates one true reading of (3) for each LF satisfied by its sister DP.

Percus and Sauerland's hypothesis was that the difference lies in the use of a *concept generator* (CG) variable in the case of *de re*. An $(x, w$-acquaintance-based) concept generator G is a function of the type $\langle e, se \rangle$ which, for each individual y s.t. x is acquainted with y in the world w, provides the unique acquaintance function (individual concept) f s.t.:

- $f(w) = y$;
- for all $v \in \text{DOX}(x, w)$, $f(v) = z$, where z is the unique individual to which the Lewisian [7] centre $c(v)$ is acquainted in the same way x is acquainted to y at w.

To bind CG variables in *de re* reports, the lexical entry for an attitude verb should look like

(4) $[\![\text{believe}]\!]^g = \lambda w. \lambda \mathcal{P}. \lambda x. \exists^{\text{for } x \text{ in } w} \langle G_1, \ldots, G_n \rangle. \forall v \in \text{DOX}(x, w) : \mathcal{P}(G_1, \ldots, G_n, v) \equiv \mathbf{1},$

where $\mathcal{P}$ is a function from n-tuples of CGs to propositions. The use of **sequences** of CGs is motivated by the possibility of "multiply *de re*" reports. As can be seen, the length of the CG sequences the verb quantified over depends on the number of *de re* occurrences of DPs within the attitudinal clause. (Without sequences, one would have to say that attitude verbs are type-flexible, taking as arguments functions of the type $\langle \langle e, se \rangle, \ldots, \langle e, se \rangle, \langle s, t \rangle \rangle$ depending on the number of different *de re* occurrences within the attitude clause.)

This much said, the two LFs corresponding to the two readings of (3) may be outlined.[2]

(3′)   a.   *De se* LF (true reading in Scenario 1):
             [Only John] λ_1[x_1 believes λ_2[c(w_2) [will_win-w_2]]]
        b.   *De re* LF (false reading in Scenario 1):
             [Only John] λ_1[x_1 believes λ_3[λ_2[[[G_3 c(w_2)] w_2] [will_win-w_2]]]]

1.2   Bound *de re*

Let us now turn to bound *de re* readings, which, although they "were discovered several years after [Percus and Sauerland]'s concept generator theory was formulated, ...are clearly the strongest piece of evidence in its support" [15]. In several recent publications [2, 18, 3] attention has been brought to cases such as

---

[1]Instead of acquaintance relations, I prefer to speak in terms of acquaintance functions [16, 3], which are functions $f_i : D_s \mapsto D_e$ yielding, for each world $w \in D_s$, the unique individual (if any) of whom the given individual has the ith mode of presentation. Being of the type $\langle s, e \rangle$, acquaintance functions are individual concepts.

[2]I will generally omit the abstractor over worlds in the matrix clause. The treatment of *de se* is simplified.

(5)   Olympia thinks she burgled herself.

(6)   John believes that every$_i$ female student loves her$_i$ mother.

A plain *de re* (or "simple bound") reading for (5) would be true if Olympia thought, looking at the image of a girl in a candid camera video, "That girl burgled into her own apartment!", the girl being actually Olympia herself. Alongside with this reading, another *de re* one is found:

Scenario 2. *Olympia is watching a candid camera video featuring a$_1$ girl breaking into an apartment. The portrait of the$_2$ (female) owner can bee seen on the wall. Olympia thinks: "This$_1$ girl burgled into that$_2$ girl's apartment!" In fact the video documents the drunken Olympia breaking into her own apartment, where her portrait hangs.*

This is the "bound *de re*" reading for (5). Although not without complications (cf. the papers cited above), it can be generated from the following LF:

(5′)   Olympia λ_1[x_1 thinks $\lambda_{\langle 3,4 \rangle}$[λ_2[[[$\underline{G_3 \text{ c}(w_2)}$] w_2] [burgled-w_2 [[$\underline{G_4 \text{ c}(w_2)}$] w_2]]]]]

Here both the subject and the complement of the verb are "CG phrases", and the two CG variables are not coindexed. Therefore, the verb *believes* quantifies over pairs of CGs, to the effect that Olympia-the-burglar may be presented to the attitude holder under a mode of presentation different from how Olympia-the-owner is presented to her. This does not contradict the coindexing of the DPs embedded into CGs.

Likewise, (6) has a true reading in the following circumstances.

Scenario 3. *John is shown two sets of pictures. In each set, every relevant female student (Mary, Sally and Sue) is pictured exactly once. John, who does not know that he is only dealing with three girls and that those girls are students, points to Mary's picture$_1$ from the first set, then to Mary's picture$_2$ from the second one, and says: "This$_1$ girl likes that$_2$ girl's mother". The same then repeats for the other girls.*

The LF for the "bound *de re*" reading, with *every female student* moved by Quantifier Raising (QR), is as in (6′), @ being the dedicated index for the actual world.

(6′)   John λ_1[x_1 believes $\lambda_{\langle 3,4 \rangle}$[λ_2[
             [every female_student-@] λ_5[[[$\underline{G_3 \text{ } x_5}$] w_2] [loves [[$\underline{G_4 \text{ } x_5}$] w_2]'s_mother]]]]]

2   Charlow and Sharvit on "universal" readings

As we have seen, the original version of the CG theory endows attitude verbs with existential quantificational force over (sequences of) CGs, and that's for good reason: not only does (3) on its *de re* reading require no more than there being **some** mode of presentation of John for himself under which he believes himself to be winning; as witnessed by Scenario 4, the same goes for (7), where John has to possess, for each actual female student, **some** mode of presentation under which he takes her to be French [3].

(7)   John believes that every female student is French.

Scenario 4. *John is shown two sets of photographs. In each set, every relevant female student (Mary, Sally and Sue) is pictured exactly once. John, who does not know that he is only dealing with three girls and that those girls are students, points consecutively to Mary's picture from the first set, to Sally's picture from the second one and to Sue's picture from the second set. Then he says: "Those girls are French".*

Things become different, however, once a negative determiner appears in the place of *every*: as Charlow and Sharvit claim, (8) requires for its truth that John have the relevant belief under **all** salient modes of presentation he has for each girl.

(8)   John believes that no female student is French.

The sentence (8) is false in Scenario 4 but would be true if John proclaimed that none of the six photos depicted Frenchwomen. In other words, in certain contexts (see also Section 5) attitude verbs seem to quantify universally over (sequences of) CGs.

Assuming the denotation for *believe* given in (4) and the syntax of *de re* reports as presented in [10, 3], there is no way to generate this "universal" reading. Among possible ways to deal with this empirical challenge, Charlow and Sharvit chose to stipulate the lexical ambiguity of attitude verbs. E.g. *believe* in their analysis becomes the realisation of two different items, which I will label as *believe*$_\exists$ and *believe*$_\forall$. To the former corresponds the entry in (4); the latter has the semantics in

(9) $[\![\text{believe}_\forall]\!]^g = \lambda w.\lambda \mathcal{P}.\lambda x.\forall^{\text{for } x \text{ in } w}\langle G_1, \ldots, G_n\rangle.\forall v \in \text{DOX}(x, w) : \mathcal{P}(G_1, \ldots, G_n, v) \equiv \mathbf{1}$

While this semantics indeed captures the data Charlow and Sharvit present, it is rather stipulative: one would definitely like to derive the "universal" flavour from the fact that it is manifest with downward entailing quantifiers. In what follows, I will argue that the ambiguity approach is not even empirically adequate, and then suggest a more uniform analysis.

3   Against the "universal" *believe*

Even if the universal reading is strongly dispreferred with DPs other than downward-entailing quantifiers, one would expect it to show up in cases where it would be the only true reading of a sentence. My strategy in this section will be to show that this is not the case; accordingly, I will conclude that there is no such thing as the universal version of attitude verbs' denotations.

To see the argument, consider the following scenario.

Scenario 5. *Paul, Ralph and Sam watch a certain talent search show on a daily basis. One day they all take notice of a gifted singer, Alfred, and of the judge called Bill. The next day they are all fascinated by a great juggler, Mr. Adams, and fancy a remarkable strict judge who is called Mr. Brown.*

When asked about their expectations, the three fans respond as follows. Paul says: "Bill will vote for Alfred and Brown for Adams". Ralph says: "Bill will vote for Alfred". Sam says: "Both Bill and Mr. Brown will vote for each of Alfred and Mr. Adams".

Unbeknownst to the three, Alfred is the same person as Adams and Bill is Brown.

Here is a chart summarising the beliefs of the three watchers:

Paul	Alfred	Adams	**Ralph**	Alfred	Adams	**Sam**	Alfred	Adams
Bill	+		Bill	+		Bill	+	+
Brown		+	Brown			Brown	+	+

In Scenario 5, Sam is the only one who has the belief of the form "$\lceil x$ will vote for $y\rceil$" under all relevant pairs of modes of presentation: $\langle$Bill, Alfred$\rangle$, $\langle$Bill, Adams$\rangle$, $\langle$Brown, Alfred$\rangle$ and $\langle$Brown, Adams$\rangle$. The theory which stipulates the existence of *believe*$_\forall$ predicts that Sam will be the only one to verify the reading of (10) where *believes* is resolved as *believes*$_\forall$:

(10)   Only Sam believes Bill Brown will vote for Alfred Adams.

Therefore, despite Paul and Ralph verifying **another** reading of *believe Bill Brown will vote for Alfred Adams*, where *believes* is resolved as *believes*$_\exists$, (10) should have a true reading in Scenario 5. (As mentioned above, the *only* test is assumed to detect all available readings of the structure over which [*only* DP] scopes.) However, no such reading is available.

The conclusion I draw from this observation is that in cases where the universal reading comes to prominence,[3] it is most likely an effect of the interaction between the negative determiner and the quantifier over modes of presentation (or CGs); cf. the equivalence familiar from first-order logic: $\neg \exists x \phi \Leftrightarrow \forall x \neg \phi$. The problem is to demonstrate how such interaction is possible; on the current assumptions concerning the syntax of attitude reports [10], there is no way to put a quantifier over CGs into the scope of a negative determiner within the embedded clause. The next section contains a proposal which provides for this scope ordering.[4]

4   The proposal: an intermediate location

4.1   Desiderata

Summing up the desiderata for a successful account of *de re*, one finds at least the following requirements.

1. To correctly handle "double vision" scenarios. This forces the quantifier over modes of presentation, whatever its concrete realisation, to scope above the embedded negation.

2. To yield both the *de se*/*de dicto* and the *de re* reading of attitude report where *only* associates with focus on the matrix subject. Hence, the quantifier over modes of presentation is restricted to positions below the matrix subject (ruling out e.g. existential closure at the discourse level, as done for indefinites in some dynamic settings).

---

[3] An additional observation supporting this conclusion is that no cases have so far been found where both the "existential" and the "universal" reading are available for a single sentence.

[4] An anonymous reviewer suggests that I should be more explicit about how Maier's [8] DRT-based account fares on the discussed data. Maier's key idea is that a referential expression (such as a proper name or a definite) triggers a presupposition which projects to the global context; whenever this projection involves crossing the boundary of an attitude clause, another presupposition appears to the effect that the attitude holder must stand in an acquaintance relation to the denotation of the definite. If the definite's presupposition is projected, the definite gets a *de re* reading; if the presupposition remains *in situ*, a *de dicto* reading.

In this setting, Percus and Sauerland's [10] examples like (3) are given two representations (p. 403) depending on whether the acquaintance presupposition projects, with the acquaintance relation "bound to equality" (*de se*), or remains *in situ*, which is truth-conditionally equivalent to the *de re* reading. In the case of *de re*, the acquaintance relation should be allowed to covary with the value of the matrix subject (which can be quantificational, as in

(i)   Every candidate believes that he will win the election.

This means that there should be a variable at the acquaintance relation name, bound by the matrix subject, which precludes global accommodation ("trapping" of the presupposition).

Given this, is is unclear how to represent the dependence of quantification over modes of presentation on the **embedded** subject, as in (7). This would be possible if quantified embedded subjects could QR out of the attitude clause and take scope over the whole attitude report. The latter option, however, is shown to be unavailable by Charlow and Sharvit [3]:

(iii)   John is certain that no female student$_i$ likes her$_i$ mother.
         Cannot mean: #'For no female student x: John is certain that x likes x's mother'

Once such long-distance movement is impossible, the acquaintance presupposition accommodated at the embedded level still scopes over the embedded subject. In my account the position of Δ is below the (short-distantly QRed) embedded subject, which solves the problem.

3. To offer an analysis of "bound *de re*" pronouns, either reflexives or possessives. The existence of such cases means that multiple modes of presentation are in principle available for a given DP within a single attitude clause.

4. To provide a replacement for the "universal reading" of attitude verbs in the case of negative quantifiers. The easiest way to do so would be to show that the effective scope ordering is $\neg\exists$, which is equivalent to $\forall\neg$.

The last requirement is *prima facie* puzzling: traditionally, it is the attitude verb that is assumed to quantify over (sequences of) CGs, but the verb outscopes the embedded subject. Therefore there is no straightforward way to switch the order of the quantifier over CGs and the negation built into the denotation of the negative quantifier.

4.2 Satisfying the desiderata

The proposal of the present paper is to split the job done by the attitude verb into two parts. The usual "displacement" part, i.e. binding the possible world (or situation) parameter, will be left to the verb, whereas quantification over CGs will be relegated to a special operator (call it Δ), whose position is **immediately below** the material generated by Quantifier Raising in the embedded clause, i.e. exactly below the lowest of the indices binding the traces of QRed DPs. Here is the LF for (7) we obtain:

(7′)   John $\lambda_1[\ x_1$ believes $\lambda_2[\ $ [every female_student-@] $\lambda_3[\ \Delta \ [\ [[G_\alpha \ x_3] \ w_2]$ is_French-$w_2] \ ] \ ] \ ]$

Naturally, *every female student* undergoes Quantifier Raising to avoid type mismatch. For proper names such as *Ortcutt* this might seem unnecessary, but I will assume they also undergo QR. The reason is that, as Santorio [15] notes, a DP embedded into a "CG phrase" does not c-command anything outside that phrase, so it is unclear why (11) is ungrammatical on any reading that preserves coindexing (as no violation of Condition C is predicted).

(11)   * Ralph believes that he$_i$ likes Ortcutt$_i$.

To formulate the semantics of Δ, assume with Santorio [15] that alongside with the usual assignment parameter g responsible for the evaluation of (individual) variables, there is another assignment parameter h whose sole role is to evaluate CG variables. (This is done for technical reasons; in my formulation nothing important hinges on it but the semantics of Δ as a h-switcher is easier to formulate.) Indices subject to interpretation in terms of h are represented by Greek letters.

(12)   SEMANTICS OF Δ (first take):
$$[\![\Delta]\!]^{g,h} = \lambda\phi.\exists h'.[\![\phi]\!]^{g,h'} \equiv \mathbf{1}$$

(13)   SEMANTICS OF CG VARIABLES:
$$[\![G_\alpha]\!]^{g,h} = \lambda x.\lambda w.[h(\alpha)](x)(w)$$

Given (12) and (13), the interpretation of the LF in (7′) should be as in

(14) $\forall w \in \text{DOX}(\mathbf{j}, @)\big(\forall x\big(\mathbf{female\_student}_@(x) \to \exists h(\mathbf{french}_{w_2}([h(\alpha)](x)(w)))\big)\big)$.

The quantifier over h-assignments scoping below the universal quantifier over students correctly predicts the possibility that for each actual female student, John has a different mode of presentation under which he believes her to be French. On the other hand, the semantics for (8) will look like

(15) $\forall w \in \text{DOX}(\mathbf{j}, @)\big(\neg\exists x\big(\mathbf{female\_student}_@(x) \wedge \exists h(\mathbf{french}_{w_2}([h(\alpha)](x)(w)))\big)\big)$,

precluding John from having, for any of the actual female students, any mode of presentation under which he takes her to be French, as desired.

The proposed modification naturally captures bound *de re* readings: just like in Percus and Sauerland's version, they are generated by using different CG variables for the coreferential DPs. Moreover, a feature the present account shares with that of Santorio is that no type-flexibility of attitude verbs is needed; in the particular implementation offered above, even quantification over sequences is never evoked (as CG variables remain free and are interpreted by the h-assignment, which is in turn manipulated by Δ).

4.3   A complication

As it stands, the proposal outlined above gives rise to the sort of overgeneration noticed (with reference to an earlier paper [14]) in Santorio's [15] (where it is called "long-distance binding"): if allowed to be evaluated by any h they like, CG variables will make reference to acquaintance functions which belong not to the attitude holder herself. E.g. (16) will then be predicted to have a true reading if, in Mary's opinion, John believes that some x, who is the value of **Mary's** acquaintance function for Ortcutt, is a spy. The indexing which leads to the incorrect reading is given in (17).

(16)   Mary thinks that John thinks that Ortcutt is a spy.

(17)   * Mary $\lambda_1[\ x_1$ thinks $\lambda_{2,\alpha}[$ John $\lambda_3[\ x_3$ thinks-$w_2\ \lambda_{4,\beta}[\ [[\underline{G_\alpha}$ Ortcutt$]\ w_4]$ is_spy-$w_4\ ]]\ ]]$

What is needed to avoid the problem is to ensure in some way or other that the value of G_α is among John's acquaintance functions. Additionally, one would like to rule out variants of (17) where the world argument of G_α is @ or w_2 instead of w_4. Note that the latter situation is not unimaginable: once trans-world existence is granted, something which is the value of John's acquaintance function $[h(\alpha)](\llbracket$Ortcutt$\rrbracket)$ in the doxastic alternatives of **Mary's** (or, for that matter, in the actual world) may easily be a spy in **John's** doxastic alternatives.

The latter predicament may be resolved by leaving the world argument of G_α unsaturated and making the verb take such an intensional argument,[5] but this route is undesirable for the reasons of uniformity: we would not be happy with the verb taking an extensional subject in case there is no CG and an intensional subject in case there is one, and uniformly requiring intensionality would perhaps be too complex. I will have to leave this part of Santorio's problem unsolved. (One could however use Santorio's own apparatus to fix it.)

As for the first task, namely to make sure that only the closest attitude predicate contributes a holder whose acquaintance functions may be assigned by h to a CG variable, we may account for it by putting a part of the h-related duty back to the attitude verb. More specifically, let the verb quantify over h-assignments, imposing a restriction on them:

(18) $\llbracket$believes$\rrbracket^{g,h} = \lambda p.\lambda x.\lambda w.\underline{\forall h'}\big[\forall\alpha\forall y.[h'(\alpha)](y)$ is an acquaintance function for $x\big].$
$$\forall w' \in \text{DOX}(x,w).\llbracket p\rrbracket^{g,h'}(w) \equiv \mathbf{1}.$$

Now we have to modify the semantics for Δ given in (12):

(12′)   SEMANTICS OF Δ (second take):
$\llbracket\Delta\rrbracket^{g,h} = \lambda\phi.\exists h' \in \text{PERM}(h).\llbracket\phi\rrbracket^{g,h'} \equiv \mathbf{1},$
where PERM(h) is the set of permutations of h, i.e. the assignments differing from h at most in which acquaintance function they assign to which index, but crucially **not** in their definedness conditions.

---

[5]Cf. the explanation of a similar restriction on transparent readings of predicates, known as Generalisation X [9], in [17] (pp. 446–447).

With the help of this definition, we retain a quantifier over h-assignments within the scope of QRed DPs (which is needed to handle "universal" readings) and at the same time transmit the restrictions from the attitude verb to this lower quantifier.

4.4 Related proposals

It is not as if the stipulation of silent operators in positions similar to that of Δ were unprecedented in the literature. For instance, Podobryaev [11] suggests a pair of covert operators $\oplus$ and $\oplus$ in order to handle the behaviour of pronouns in the presence of *imposters*, i.e. nonpronominal expressions which are able to obtain 1st- or 2nd-person semantics. Such an operator, with the semantics as in

(19) $[\![\oplus]\!]^g = \lambda\phi.[\![\phi]\!]^{g'}$,
where g' is like g except that all values of g' for the indices of the form $\langle i, \text{①} \rangle$, where i indicates coreference and ① the first person feature, are set to "undefined",

may occupy the position between the attitude verb and the lexical content of the attitude clause. Moreover, the splitting move I made when I detached quantification over (chains of) CGs from verbs has its precursor in Shklovsky and Sudo's [19], where the monstrous operator in Uyghur was argued to mingle into the embedded clause, thus freeing the matrix verb from the parameter-shifting duty.

Additionally, a somewhat looser analogy may be drawn between my Δ and Asudeh and Giorgolo's [1] silent "unit" operator η, whose function is to abstract over the perspective index (and which is used in perspective-sensitive contexts such as *Lois Lane loves* _____ to generate ambiguities).

5   The behaviour of *only*

We have so far ignored the second point raised by Charlow and Sharvit [3] in support of the existence of universal readings. The point is that a *de re* report with *only X* as the embedded subject requires for its truth that the attitude holder lack the corresponding attitude w.r.t. **any** mode of presentation she has of any of X's alternatives (in the sense of [13]):

(20)   John believes that only Mary is French.

> **Scenario 6.** *John is given six photos: two of Mary, two of Sally and two of Sue. He thinks those may be six different girls and says...*
>
> a.   OK *...pointing to **both** pictures of Mary's, "Those girls are French, the others aren't"*
>
> b.   # *...pointing to **one** of Mary's pictures and **one** of Sue's, "Those girls are French, the others aren't"*

The reason why the problem emerges is that the existential construal of attitude verbs leads to the following (simplified) configuration of the quantifier over (sequences of) CGs and *only*:

(21) $\ldots \exists G \ldots [\text{only Mary}] \ldots$ is French.

The order in (21) that the reading in (20b) should be available: indeed, in (20b) there is an acquaintance function for John s.t. only Mary yields a value x of this function s.t. x is French; although Sue is also believed *de re* to be French by John, she is so believed under a **different** acquaintance function.

Having pointed out the problem, we can now easily see how the re-ordering of quantifiers I have proposed above helps: now it is Δ, not *believes*, that supplies the existential quantifier over CGs, and Δ's syntactic position is below those of QRed DPs in the attitude clause. Therefore, instead of the unwelcome (21), the quantifier over h-assignments (i.e. the denotation of Δ) nevertheless ends up within the scope of *only*:[6]

(22)   John believes $\lambda w[$ [[only C] Mary]$_i$ $\lambda_i[$ Δ [[[G x_i] w] is French]]],

thus yielding the truth conditions in

(23)   'In all worlds compatible with what John believes, no x except for Mary is s.t. there is an acquaintance function f for John s.t. $f(x)$ is French';[7]

thus it cannot be that John also has a mode of presentation, even if a different one, under which he ascribes Frenchhood to Sally.

At least the most disastrous lapses of the existential analysis of attitude verbs are thus remedied by changing the location of the CG-quantifier; some issues remain, however. First, what happens if Scenario 6 is modified as follows?

Scenario 7. *John is given six photos: two of Mary, two of Sally and two of Sue. He thinks those may be six different girls and says...*

c. *...pointing to* **one** *of Mary's pictures,* **"This girl** *is French, the others aren't"*

The difference from Scenario 6 here is, obviously, that John does not ascribe Frenchhood to anyone except for Mary, under any mode of presentation; in contrast to (20a), however, John does not ascribe it to Mary under **all** modes of presentation he has for her. Is (20) true in such a scenario? The question requires further study, and the negative answer would mean that the present account is incomplete: something will then have to be said about the residual universality that remains unexplained by the reordering of quantifiers I propose.

Second, I have assumed that *only* forms a constituent with *Mary* in (20b) and they are QRed together; even granted that this is a plausible option, it is unclear whether one should allow for combinations such as [*only* [[G *Mary*] w]]. If QRed above Δ, such a constituent would have a free variable G in it, so the LF would not be interpretable. But what if *Mary* alone were QRed out of this structure? (Luckily, movement out of focus with which *only* associates may be impossible at all [4].)

6   Conclusion

The aim of the present paper was two-fold. First, I pointed to an empirical problem faced by Charlow and Sharvit's [3] account of the "universal" readings of attitude reports. The problem arises in the cases where we would expect *only* to discriminate between the "existential" and the "universal" reading; the latter unexpectedly fails to arise. Second, I suggested that quantification over modes of presentation takes place not at the level of the attitude verbs but lower in the structure, specifically below the QRed arguments within the subordinate clause.

---

[6]This is so even if, as Charlow and Sharvit assume (p. 34), "the presuppositions triggered by *only* can be accommodated in the scope of the attitude à la Heim [5]". Without postulating Δ, one would have to argue that the presupposition of *only* has to be accommodated globally; I will not, however, go as far as to call that route impossible: it may be, after all, that the *de re* reading of the focus constituent somehow forces global accommodation.

[7]Recall that in Section 4.3 we endowed attitude verbs with a quantifier over h-assignments. It quantifies vacuously, so the preference towards universal force (as opposed to existential) may have seemed unmotivated to the reader. Now that we get to cases involving *only*, existential force would make my analysis vulnerable to Charlow and Sharvit's criticism regarding the readings of (20).

The discussion above left several questions unresolved. For one thing, I did not explain how the empirically desirable restrictions on world pronoun indexing come about; for another thing, more work is needed on the scenarios verifying reports with *only*.

As for other perspectives, I believe, perhaps a bit paradoxically, that a general theory of *de re* ascription may eventually do away with concept generators: applying the CG technique to reflexives leads to binding-theoretic problems [18], and the observed range of readings is hard to account for [2]. This suggests that treating reflexives as arity-reducing operators may be more productive than assigning indices to them. Then "bound *de re*" should be reinterpreted as *de re* readings of the predicate whose arity is reduced and are no longer strong evidence for the CG theory. A more detailed discussion of the issue should wait for another occasion.

References

[1] Ash Asudeh and Gianluca Giorgolo. Perspectives. 2015.

[2] Simon Charlow. *De re* anaphors. SALT 20 poster, 2010.

[3] Simon Charlow and Yael Sharvit. Bound 'de re' pronouns and the LFs of attitude reports. *Semantics and Pragmatics*, 7(1):1–43, 2014.

[4] Michael Yoshitaka Erlewine. *Movement out of focus*. PhD thesis, Massachusetts Institute of Technology, 2014.

[5] Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9(3):183–221, 1992.

[6] David Kaplan. Quantifying in. *Synthese*, 19(1–2):178–214, 1968.

[7] David Lewis. Attitudes *de dicto* and *de se*. *Philosophical Review*, 88(4):513–543, 1979.

[8] Emar Maier. On the roads to *de se*. In *Proceedings of SALT*, volume 21, pages 393–412, 2011.

[9] Orin Percus. Constraints on some other variables in syntax. *Natural Language Semantics*, 8(3):173–229, 2000.

[10] Orin Percus and Uli Sauerland. On the LFs of attitude reports. In *Proceedings of Sinn und Bedeutung*, volume 7, pages 228–242, 2003.

[11] Alexander Podobryaev. Impostrous domains. In *Proceedings of NELS 44*, volume 2, pages 91–104, 2014.

[12] W. V. Quine. Quantifiers and propositional attitudes. *Journal of Philosophy*, 53(5):177–187, 1956.

[13] Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116, 1992.

[14] Paolo Santorio. Modals are monsters: on indexical shift in English. In *Proceedings of SALT*, volume 20, pages 289–308, 2011.

[15] Paolo Santorio. On the plurality of indices, September 2014.

[16] Uli Sauerland. Counterparts block some 'de re' readings. In Luka Crnič and Uli Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 2, pages 65–85. MITWPL, 2014.

[17] Florian Schwarz. Situation pronouns in determiner phrases. *Natural Language Semantics*, 20(4):431–475, 2012.

[18] Yael Sharvit. Covaluation and unexpected BT effects. *Journal of Semantics*, 28(1):55–106, 2011.

[19] Kirill Shklovsky and Yasutada Sudo. The syntax of monsters. *Linguistic Inquiry*, 45(3):381–402, 2014.

[20] Seth Yalcin. Quantifying in from the Fregean perspective. *Philosophical Review*, 124(2):207–253, 2015.

[21] Igor Yanovich. The problem of counterfactual *de re* attitudes. In *Proceedings of SALT*, volume 21, pages 56–75, 2011.

An Interactive Approach to Proof-Theoretic Semantics

James Trafford

University of Creative Arts Epsom, UK
jtrafford2@ucreative.ac.uk

Abstract

In truth-functional semantics for propositional logics, categoricity and compositionality are unproblematic. This is not the case for proof-theoretic semantics, where failures of both occur for the semantics determined by monological entailment structures for classical and intuitionistic logic. This is problematic for inferentialists, where the meaning of logical constants is supposed to be determined by their rules. Recent attempts to overcome these issues have primarily considered symmetric entailment structures, but these are tricky to interpret. Here, I instead consider an entailment structure that combines provability with the dual notion of disproof (or refutation). This is interpreted as a dialogue structure between the roles of prover and denier, where an assertion of a statement involves a commitment to its defence, and a denial of the statement involves a commitment to its challenge. The interaction between the two is constitutive of a proof-theoretic semantics capable of dealing with the above issues.

1 Inferentialism and categoricity

Logical inferentialism is usually taken to rest upon the idea that the meaning of a logical expression is determined by its inferential rules, where those rules have a substantive connection with ordinary inferential practices. Sometimes called proof-theoretic semantics, this approach should not take truth as the primary semantic notion, nor should it require access to truth-conditional semantics in order to fix the meanings of logical constants. So, the (fairly standard) distinction that is often made between formal proofs and semantic models, and upheld by soundness and completeness proofs, must be somehow overcome. This distinction rests upon the idea that formal proofs are purely syntactical entities, and, as such, can not be the sort of things that are meaningful in a true sense. In this respect, it is of central importance that inferentialism shows, instead, that patterns of inference are the sort of things that can be both represented syntactically, and confer meaning upon logical expressions. This coheres with the more general position in philosophy of language that draws upon Wilfrid Sellars' "inferential game of making claims and giving and asking for reasons", developed most prominently by Robert Brandom [3].[1]

One issue that arises for this approach to semantics is the well-known "categoricity" problem, where the standard inferential rules for classical logic fail to rule out non-standard interpretations, so they do not suffice to determine the meaning of logical constants. In particular, that framework is easily shown to be sound and complete with respect to *both* the classical semantic model, and a model in which every formula is interpreted "true". This is also problematic from the point of view of compositionality, since it is simple to show that the rules defining disjunction, for example, do not ensure that the truth-values of the sub-formulas for a formula $\alpha \vee \beta$ always determine the truth-value of that formula.

Say that a "logic", L, is an entailment structure, which is an ordered pair, $(S, \vdash_L)$, where S is a denumerable set of propositional formulas, and $\vdash_L$ is a binary entailment relation defined

---

[1] See also [11] for a defense of games as central to logic.

on $P(S) \times P(S)$ ($P(S)$ is the set of all finite subsets of S) satisfying the properties of reflexivity, transitivity, monotonicity, and finitariness. Entailment structures can be restricted in different ways, which we think of in terms of sequents in a logic. These are just ordered pairs, $\langle \Gamma, \Delta \rangle$ where Γ, Δ are finite (possibly empty) sequences of formulas of S. Then, say that a right-asymmetric sequent is restricted to at most a single formula on the right; a left-asymmetric sequent is restricted to at most a single formula on the left, and a symmetric sequent has no such restrictions. Now, say that a sequent rule $\mathcal{R}$ in any logic L is an ordered pair consisting of a finite sequence of sequent premises and a sequent conclusion $\mathcal{R} = \langle \{ SEQ^P \}, SEQ^C \rangle$. In this way, we may think of a specific logic to be determined by a proof structure (set of axioms and sequent rules), where any collection of sequents $\mathcal{S}$ that is closed under the standard structural rules determines a finitary, normal, logic. For example, it is the case that $\Gamma \vdash \alpha$ iff for some finite $\Gamma_0 \subseteq \Gamma$, we have $(\Gamma_0 \vdash \alpha) \in \mathcal{S}$.[2]

Take the rules ordinarily used to define classical negation (in a right-asymmetric entailment structure):

$$\frac{\Gamma, \neg\alpha \vdash \beta \qquad \Gamma, \neg\alpha \vdash \neg\beta}{\Gamma \vdash \alpha} \ (Reductio) \qquad\qquad \frac{\Gamma \vdash \alpha \qquad \Gamma \vdash \neg\alpha}{\Gamma \vdash \beta} \ (EFQ)$$

Now, consider a standard Lindebaum-Asser construction: For any finite normal logic L, given a formula β, and a theory Γ (where $\Gamma, \beta \in S$) such that $\Gamma \nvdash \beta$, Γ can be extended to Γ' (where $\Gamma \subseteq \Gamma'$ and Γ' is relatively maximal with respect to β (in L) so that for no proper superset Γ'' of Γ' do we have $\Gamma'' \nvdash \beta$. In [2], it is proven that a semantics for a logic, L, is given by taking the characteristic function of relatively maximal theories for L as follows: given a sequent $\alpha_1, ..., \alpha_n \vdash \beta$ in a logic L, and a relatively maximal theory Γ' of L, we say that Γ' satisfies this sequent iff, whenever $\Gamma' \vdash \alpha$, for each $\alpha_1, ..., \alpha_n$, $\Gamma' \vdash \beta$. The issue mentioned above is that there are possible interpretations where both α and $\neg\alpha$ are in Γ', and which are not ruled out by the rules used above to determine the maximal sets of formulas. Similarly, the rules defining disjunction (in right-asymmetric form) ensure that whenever $\alpha \in \Gamma'$ or $\beta \in \Gamma'$, it is the case that $\alpha \vee \beta \in \Gamma'$. But, they fail to rule out a situation in which $\alpha \vee \beta \in \Gamma'$, whilst neither α nor β are.

An obvious fix for this issue, which also sheds light on the root of the problem, is to use a symmetric sequent structure to define the logic. The extension theorem in this context is slightly different since we can say that, for any finite normal logic L, and $\Gamma \nVdash \Delta$, Γ, Δ can be extended to $\Gamma \subseteq \Gamma'$, $\Delta \subseteq \Delta'$, with $\Gamma' \nVdash \Delta'$ (and no proper superset $\Gamma'' \supseteq \Gamma'$, $\Delta'' \supseteq \Delta'$ do we have $\Gamma'' \nVdash \Delta''$). Let us call such maximal Γ', Δ' a quasi-partition. The semantics defined this way also differs: for a symmetric sequent, $\alpha_1, ..., \alpha_n \Vdash \beta_1, ..., \beta_m$, in a logic L, and a quasi-partition Γ', Δ' of L, we say that the quasi-partition satisfies this sequent iff it is not the case that $\Gamma' \Vdash \alpha$, for each $\alpha_1, ..., \alpha_n$ whilst $\Delta' \Vdash \beta$, for each $\beta_1, ..., \beta_m$. Now, we have a situation in which the non-normal interpretations are straightforwardly ruled out, primarily because we have symmetry over the turnstile. So, for example, the scenario in which $\alpha, \neg\alpha$ end up in Γ' is ruled out simply by ensuring that we have $\alpha, \neg\alpha \Vdash \emptyset$ for all formulas of S. Similarly, the trouble with disjunction is now assuaged since we have $\Gamma, \alpha \vee \beta \Vdash \alpha, \beta, \Delta$. The symmetry manifest itself most clearly in the fact that we now have a way of refuting propositions that is symmetrical with the way in which they are proved, and collecting these refuted propositions

---

[2] For symmetric sequents, this is $\Gamma \Vdash \Delta$ iff for some finite $\Gamma_0 \subseteq \Gamma$, $\Delta_0 \subseteq \Delta$ we have $(\Gamma_0 \Vdash \Delta_0) \in \mathcal{S}$. See also [8, p.113]. Note that I use $\Vdash$ rather than typical $\vdash$ for symmetric sequents to highlight that they can be read in both directions.

together in Δ'.

The trouble with this "fix" is that I do not think it gives us any sort of proof-theoretic semantics at all. It is tricky to interpret symmetric entailment structures, and it is unclear that they provide any account of "proof" whatsoever. According to one prominent interpretation, due to Greg Restall [17], a valid sequent $\Gamma \vdash \Delta$ can be interpreted as saying "it is incoherent to simultaneously assert all of Γ, and deny all of Δ". But, then, we do not have that, for example, (classically) $(\alpha \vee \neg \alpha)$ to be provable, rather it is merely incoherent to deny it. Additionally, according to Lafont [6, Appendix B1], the lack of constructive properties, such as disjunction property, means that "classical logic has no denotational semantics except the trivial one which identifies all the proofs of the same type".[3]

By way of response to the above issue, we might take the restriction on sequents to right-asymmetric form to define intuitionistic, rather than classical, logic. In this setting, the validity of a proof is not dependent upon an interpretation, but rather on local constraints such as canonicity and harmony [15, e.g.]. This way of thinking about valid proofs is not appropriate to classical logic because canonicity requires the disjunction property to hold: a proof of a disjunction must be given in (or reduced to) canonical form, which requires that it is also possible to provide a proof of one of the disjuncts. So, whilst, in a formal sense, symmetric classical logic is stronger than intuitionistic logic (the set of theorems of intuitionistic logic is a proper subset of the theorems of classical logic), informally, classical logic is weaker from the point of view of proofs and the determination of meaning. In symmetric classical logic, we do not have proofs at all. Similarly, we do not have refutations. For instance, say that $\alpha \wedge \beta$ is refuted whenever one of the conjuncts is refuted. But, we can refute $\alpha \wedge \beta$ without having a refutation of either of the conjuncts since the "conjunction" property for refutation fails in classical logic. This is just as problematic from the $p.o.v$ of constructing proof-theoretic semantics, since we require an ability to determine, in a fine-grained manner, the manner in which a proof or refutation is "valid".

In [14] it is argued that the problem of categoricity is avoided by this manoeuvre. In this context, the problematic situations would be a scenario in which every statement is provable is compatible with the inferential rules, or it is possible for $\alpha \vee \neg \alpha$ to be provable, whilst neither α nor $\neg \alpha$ are. The reason that these situations can not arise, according to [14], is that there can be no canonical proof of 0 (since there exists only the null introduction rule for 0), so that rules out a scenario in which there exists a proof of both α and $\neg \alpha$ (where $\neg \alpha$ is equivalent to $\alpha \Rightarrow 0$).[4] But, as pointed out in [9], the justification of this rule makes use of the rule for 0-elimination, unlike other rules for connectives, and alternative rules where this is not the case do in fact allow for situations in which proofs for α and $\neg \alpha$ may exist (specifically, where there exists a proof for each atomic formula of the language). The sticking point is that: 'one also needs to regard, controversially, the rule of 0-elimination as justified on the basis of a non-existent rule of 0-introduction, taken as saying that there is no canonical proof of 0' [9]. This is similar to an argument made in [7], where it is pointed out that, whilst the 0-elimination rule only tells us that anything may be inferred from 0, this does not ensure that 0 has the meaning of false. For example, it is possible to consider a language in which all atoms are true, and in which case 0 will be true rather than false, and in which case the 0-elimination rule does not determine the (intended) meaning of 0. Of course, this means that $\neg \alpha$, defined as $\alpha \Rightarrow 0$, must also be true by vacuous discharge, and so, in this language we would have both α and $\neg \alpha$ true.

---

[3]Note that this result holds for any logic in symmetric form. For example, we may construct a paraconsistent or paracomplete logic in which the left or right negation sequent rule is inadmissible, yet is symmetrical in terms of the structure of sequent derivation (see [17] for further details). But, since Lafont's example does not involve negation, there is no reason why such logics should be capable of dealing with this issue.

[4]Note that I use 0 in place of the usual $\perp$.

So, again, the rules that are supposed to determine the meaning of negation, intuitionistically, do not do so. The issue may be generalised. Typically, when considering the relationship between an entailment structure and a semantic model, counter-models play the role of ruling out invalid formulas. So, for example, take Γ^+ as a set of sentential theorems, V as some model, with $\vdash$ a derivability relation, and $\models$ a model-satisfaction relation. Then, we ordinarily require that, for every formula α: Either $\exists\Gamma^+(\Gamma^+ \vdash \alpha)$ or, $\exists V(V \not\models \alpha)$. That is to appeal to the idea that a sequent $\Gamma \vdash \alpha$ is invalid iff some model V makes all the formulas of Γ true, whilst making α false. It is precisely this appeal to counter-models that is not in keeping with a proof-theoretic approach to semantics, yet it is also what it looks as though to be required given the inability of the asymmetric entailment structures to rule out inadmissible cases. This is perhaps more obvious intuitionistically, since there is a built-in asymmetry between proof (and truth) and refutation (and falsity). For example, truth is directly attainable (by constructing a proof), whilst falsity is typically equated with a reduction to a contradiction, so the latter relies on a syntactic feature (negation), but also on truth. The asymmetry is problematic for the development of a semantics of proofs, where we do not want to rely upon a counter-model as stand-in for refutation. Perhaps what we want instead is something more like this: Either $\exists\Gamma^+(\Gamma^+ \vdash \alpha)$ or, $\exists\Delta^-(\Delta^- \dashv \alpha)$ (where Δ^- is a set of sentential counter-theorems, or refutations, and $\dashv$ a "refutation" entailment relation). Whilst this is possible in symmetrical classical logic, we, arguably, do not get any sort of proofs or refutations in that setting. What we require is some way of maintaining symmetry whilst also ensuring that proofs (and, perhaps refutation) are the primary semantic notions.

2 Co-constructive logic

To maintain logical symmetry over proof and refutation, whilst also ensuring that we have a proof-theoretic semantics that is constructive, I propose to harness the fact that de Morgan duality (without negation) holds between intuitionistic and co-intuitionistic logic, as separate entailment structures. As I have argued elsewhere [18], the latter logic is best understood as a logic of refutation. This will give us a combined entailment structure (co-constructive logic) that is symmetric over proof and disproof (or refutation) without collapsing to classical logic, and which can be interpreted as a dialogue structure between the roles of prover and denier. In general, the idea is that an assertion of a statement brings with it a commitment to its defence, and a denial of the statement involves a commitment to its challenge, with the interaction between the two being constitutive of a proof-theoretic semantics capable of dealing with the above issues. In this setting, de Morgan duality will be shown to bring with it a form of dialogical balance, which is maintained by a kind of cut-rule between prover and denier.

First, take two entailment structures that formalize proof and refutation. I shall call these L_I for intuitionistic logic, which I take to deal with proofs, and L_C for co-intuitionistic logic, which I take to deal with refutations. First, define two languages S, S^d over a denumerable set of atomic formulas, for L_I and L_C respectively, in Backus-Naur form (α^+ is any atomic formula of S, α^- is any atomic formula of S^d) as:

$S\ \beta^+ ::= [\alpha^+|(\neg_I\beta^+)|(\beta^+ \wedge \beta^+)|(\beta^+ \vee \beta^+)|(\beta^+ \Rightarrow \beta^+)|0^+]$
$S^d\ \beta^- ::= [\alpha^-|(\neg_C\beta^-)|(\beta^- \wedge \beta^-)|(\beta^- \vee \beta^-)|(\beta^- \Leftarrow \beta^-)|1^-]$

Here, $\neg_I$ and $\neg_C$ denote the negations of the two languages, and $\Rightarrow$ and $\Leftarrow$ denote implication and co-implication, respectively. These are the key distinctions with classical logic, though as I show below, $\Leftarrow$ essentially operates as a kind of "implication for refutations". Note also that we have made a syntactic distinction between atoms of the two languages, to signify

whether they are part of a proof or a refutation. De Morgan duality holds between L_I and L_C, by replacing any use of $\wedge, \vee, \Rightarrow, \neg_I$ in the former, with $\vee, \wedge, \Leftarrow, \neg_C$ in the latter. Also, the dual to a right-asymmetric sequent $(\Gamma \vdash \alpha)$ is a left-asymmetric sequent $(\alpha \vdash \Gamma)$ [19], though we are reading the latter as a refutation read from right to left, which we indicate by a reverse turnstile, $\dashv$. Whilst the two logical structures are syntactically separated, it is instructive to "see" them within the one and the same proof-theoretic structure.[5] Formulas decorated with $(-)^+$ do not interact with formulas decorated with $(-)^-$, apart from in a rule I call TERMINAL-CUT. Sequents of this, co-constructive, logic, are quadruples, composed of two multisets of formulas, Γ^+ and Δ^-, and two distinguished sets (stoups) containing zero or one formula, Π^+ and Λ^-. This may be noted $\Lambda^-;\Gamma^+ \vdash\dashv \Delta^-;\Pi^+$, though ordinarily we will rewrite this as $\alpha^-;\Gamma^+ \vdash\dashv \Delta^-;\beta^+$, making transparent that α^- and β^+ are single formulas; where empty, we will write this as $;\Gamma^+ \vdash\dashv \Delta^-;$. Superscripts denote as above, and Γ^+ is shorthand for $\alpha_1^+,...,\alpha_n^+$ for each $\alpha \in \Gamma$ (similarly for Δ^-).[6]

$$\frac{}{;\alpha^+ \vdash\dashv\, ;\alpha^+}\ (\mathrm{Id}^+) \qquad \frac{}{\alpha^-\,;\vdash\dashv\alpha^-\,;}\ (\mathrm{Id}^-)$$

$$\frac{;\vdash\dashv\Delta^-;}{\alpha^-;\vdash\dashv\Delta^-;}\ (\mathrm{Thin}\text{-}L^-) \qquad \frac{;\Gamma^+ \vdash\dashv\,;\beta^+}{;\Gamma^+,\alpha^+ \vdash\dashv\,;\beta^+}\ (\mathrm{Thin}\text{-}L^+)$$

$$\frac{\beta^-;\vdash\dashv\Delta^-\,;}{\beta^-;\vdash\dashv\Delta^-,\alpha^-\,;}\ (\mathrm{Thin}\text{-}R^-) \qquad \frac{;\Gamma^+ \vdash\dashv\,;}{;\Gamma^+ \vdash\dashv\,;\alpha^+}\ (\mathrm{Thin}\text{-}R^+)$$

$$\frac{\beta^-;\vdash\dashv\Delta^-,\alpha^-,\alpha^-\,;}{\beta^-;\vdash\dashv\Delta^-,\alpha^-\,;}\ (\mathrm{Cont}^-) \qquad \frac{;\Gamma^+,\alpha^+,\alpha^+ \vdash\dashv\,;\beta^+}{;\Gamma^+,\alpha^+, \vdash\dashv\,;\beta^+}\ (\mathrm{Cont}^+)$$

$$\frac{\beta^-;\vdash\dashv\Delta^-,\alpha^-,\sigma^-,\Theta^-;}{\beta^-;\vdash\dashv\Delta^-,\sigma^-,\alpha^-,\Theta^-;}\ (\mathrm{Int}^-) \qquad \frac{;\Gamma^+,\alpha^+,\sigma^+,\Theta^+ \vdash\dashv\,;\beta^+}{;\Gamma^+,\sigma^+,\alpha^+,\Theta^+ \vdash\dashv\,;\beta^+}\ (\mathrm{Int}^+)$$

$$\frac{\beta^-;\vdash\dashv\Delta^-,\alpha^-;\quad \alpha^-;\vdash\dashv\Gamma^-;}{\beta^-;\vdash\dashv\Delta^-,\Gamma^-;}\ (\mathrm{Cut}^-) \qquad \frac{;\Gamma^+ \vdash\dashv\,;\alpha^+ \quad ;\Delta^+,\alpha^+ \vdash\dashv\,;\beta^+}{;\Delta^+,\Gamma^+ \vdash\dashv\,;\beta^+}\ (\mathrm{Cut}^+)$$

$$\frac{;\Delta^+,\alpha^+ \vdash\dashv\,;\sigma^+}{;\Delta^+,\alpha^+ \wedge \beta^+ \vdash\dashv\,;\sigma^+}\ (\wedge L_1^+) \qquad \frac{;\Delta^+,\beta^+ \vdash\dashv\,;\sigma^+}{;\Delta^+,\alpha^+ \wedge \beta^+ \vdash\dashv\,;\sigma^+}\ (\wedge L_2^+)$$

$$\frac{;\Delta^+ \vdash\dashv\,;\alpha^+ \quad ;\Delta^+ \vdash\dashv\,;\beta^+}{;\Delta^+ \vdash\dashv\,;\alpha^+ \wedge \beta^+}\ (\wedge R^+)$$

$$\frac{\alpha^-;\vdash\dashv\Delta^-;}{\alpha^- \wedge \beta^-;\vdash\dashv\Delta^-;}\ (\wedge L_1^-) \qquad \frac{\beta^-;\vdash\dashv\Delta^-;}{\alpha^- \wedge \beta^-;\vdash\dashv\Delta^-;}\ (\wedge L_2^-)$$

$$\frac{\sigma^-;\vdash\dashv\Delta^-,\alpha^-;\quad \sigma^-;\vdash\dashv\Delta^-,\beta^-;}{\sigma^-;\vdash\dashv\Delta^-,\alpha^- \wedge \beta^-;}\ (\wedge R^-)$$

---

[5] This is possible by using a technique similar to that used in [5], and in a similar context to this one in [1] (there are also marked differences, particularly regarding the notion of positive and negative rules in those systems).

[6] Note that Cut^+ and Cut^- are elimimable in the combined structure.

$$\frac{;\Delta^+, \alpha^+ \dashv \vdash ;\sigma^+ \qquad ;\Delta^+, \beta^+ \dashv \vdash ;\sigma^+}{;\Delta^+, \alpha^+ \vee \beta^+ \dashv \vdash ;\sigma^+} \ (\vee L^+) \qquad \frac{;\Delta^+ \dashv \vdash ;\alpha^+}{;\Delta^+ \dashv \vdash ;\alpha^+ \vee \beta^+} \ (\vee R_1^+)$$

$$\frac{;\Delta^+ \dashv \vdash ;\beta^+}{;\Delta^+ \dashv \vdash ;\alpha^+ \vee \beta^+} \ (\vee R_2^+)$$

$$\frac{\alpha^- ; \dashv \vdash \Delta^- ; \qquad \beta^- ; \dashv \vdash \Delta^- ;}{\alpha^- \vee \beta^- ; \dashv \vdash \Delta^- ;} \ (\vee L^-) \qquad \frac{\sigma^- ; \dashv \vdash \Delta^-, \alpha^- ;}{\sigma^- ; \dashv \vdash \Delta^-, \alpha^- \vee \beta^- ;} \ (\vee R_1^-)$$

$$\frac{\sigma^- ; \dashv \vdash \Delta^-, \beta^- ;}{\sigma^- ; \dashv \vdash \Delta^-, \alpha^- \vee \beta^- ;} \ (\vee R_2^-)$$

$$\frac{;\Delta^+, \alpha^+ \dashv \vdash ;\beta^+}{;\Delta^+ \dashv \vdash ;\alpha^+ \Rightarrow \beta^+} \ (\Rightarrow R^+) \qquad \frac{;\Delta^+ \dashv \vdash ;\alpha^+ \qquad ;\Gamma^+, \beta^+ \dashv \vdash ;\sigma^+}{;\Delta^+, \Gamma^+, \alpha^+ \Rightarrow \beta^+ \dashv \vdash ;\sigma^+} \ (\Rightarrow L^+)$$

$$\frac{\beta^- ; \dashv \vdash \Delta^-, \alpha^- ;}{\beta^- \Leftarrow \alpha^- ; \dashv \vdash \Delta^- ;} \ (\Leftarrow L^-) \qquad \frac{\sigma^- ; \dashv \vdash \Delta^-, \alpha^- ; \qquad \beta^- ; \dashv \vdash \Gamma^- ;}{\sigma^- ; \dashv \vdash \Delta^-, \Gamma^-, \alpha^- \Leftarrow \beta^- ;} \ (\Leftarrow R^-)$$

$$\frac{;\Delta^+, \alpha^+ \dashv \vdash ;}{;\Delta^+ \dashv \vdash \neg_I \alpha^+ ;} \ (\neg_I R^+) \qquad \frac{;\Delta^+ \dashv \vdash ;\alpha^+}{;\Delta^+, \neg_I \alpha^+ \dashv \vdash ;} \ (\neg_I L^+) \qquad \frac{; \dashv \vdash \Delta^-, \alpha^- ;}{\neg_C \alpha^- ; \dashv \vdash \Delta^- ;} \ (\neg_C R^-)$$

2.1   Interpreting co-constructive logic

In brief, we will consider a dialogue in the above structure in terms of a relationship between PROOF ATTEMPTS and REFUTATION ATTEMPTS that are "tests" of each other. As such, the turnstile, $\dashv\vdash$ may be read positively, from left to right, as a proof-attempt, and negatively, from right to left, as a refutation attempt. For example, the rules $\wedge L_1^-$ and $\wedge L_2^-$ may together be interpreted in BHK-style as "c is a refutation attempt of $\alpha \wedge \beta$ if c is a pair (l, c') such that c' is a refutation attempt of α or c is a pair (r, c') such that c' is a refutation attempt of β". This coheres with a view presented in [10], which suggests that making an assertion involves a commitment to defend it when challenged. So, assertoric norms should not be understood to restrict what an agent ought to assert, instead they are constraints on how agents respond to challenge and dialogue. This is related to inferentialism in [11], where it is argued that "an act of asserting a statement brings with it a commitment to defend the assertion, if challenged, so to make an assertion is to make a move in a game [...] the 'game of asking for and giving reasons' is embedded in the very nature of assertions". Here, we will say that an interaction consists of an assertion of a statement, which brings with it a commitment to its defence, or a denial of the statement, which involves a commitment to its challenge. It is this interaction that we will say provides an interpretation of the statement through its testing. For example, a "test" for a proof-attempt of a conjunction is disjunctive, so is more fine-grained than any counter-model since we test each formula used as a premise for the proof-attempt.

Take the following interaction. "Prover" asserts a conjunction, $\alpha \wedge \beta$, putting $\alpha \wedge \beta$ "into the game", and it is incumbent upon prover, to provide a proof attempt giving some sort of reason or evidence in support of both α and β. Refuter, on the other hand, challenges $\alpha \wedge \beta$ by providing a refutation attempt giving some sort of reason or evidence refuting either α or β. Exactly the reverse is the case if the formula in question is a disjunction. So, we can interpret the relationship between L_I and L_C in terms of tests, where a test of α^+ is just a refutation-attempt of form α^-, and a test of α^- is just a proof-attempt of form α^+: Testing $\alpha^+ \wedge \beta^+$

involves testing α^+ or testing β^+; Testing $\alpha^+ \vee \beta^+$ involves testing α^+ and testing β^+; Testing $\alpha^- \wedge \beta^-$ involves testing α^- and testing β^-; Testing $\alpha^- \vee \beta^-$ involves testing α^- or testing β^-; Testing $\alpha^+ \Rightarrow \beta^+$ involves testing for a function that maps each test of α^+ into a test of β^+; Testing $\beta^- \Leftarrow \alpha^-$ involves testing for a function that maps each test of α^- into a test of β^-.

Whereas this is fairly intuitive for conjunction and disjunction, that a "test" of an attempted proof of $\alpha \Rightarrow \beta$ is an attempted refutation of $\beta \Leftarrow \alpha$ is less obvious. However, understanding why this is the case also sheds light on the more general issue that we also introduced here the idea that proofs (and refutations) may not be valid, since proof-attempts are open to testing, counter-examples, and challenge. Note, first, that this rules out a conception of proofs as tenseless objects. Rather, following [4], we will take the view that proofs (and refutations) are acts: "a proof is a process whose result may be represented or described by means of linguistic symbols". According to [12], this alters the standard intuitionistic definition of conditional so that $\alpha \Rightarrow \beta$ may be valid only in case an agent has an actual proof of α to hand. A proof of a conditional is a function that maps actual proofs of the antecedent into actual proofs of the consequent since: 'as long as no proof of α is known, [the function] f has nothing to map. So we can still define f as the constant function which, once a proof π of α is known, maps every proof of α into the proof of β'.[7] So, a valid proof of a conditional $\alpha \Rightarrow \beta$ requires a proof of α in addition to a proof that there exists a function mapping every proof of α into a proof of β.

This can be usefully mapped on to a distinction between proof, or refutation, *attempts*, and *valid* proofs and refutations. Consider this in terms of the detachable forms of the entailment relations $\alpha \vdash \beta$ and $\alpha \dashv \beta$. Intuitionistically, implication is a detachable operator that embeds provability such that $\alpha \vdash_I \beta \leftrightarrow \vdash_I \alpha \Rightarrow \beta$. In co-intuitionistic logic, co-implication is a detachable operator that embeds refutability such that $\alpha \dashv_C \beta \leftrightarrow \dashv_C \beta \Leftarrow \alpha$.[8] So, by deduction (and dual-deduction) theorem, we can read a sequent $\Gamma \vdash \alpha$ as "under the assumption that there exists a proof of each $\alpha \in \Gamma$, there exists a proof of β also" (similarly for refutations and $\Gamma \dashv \alpha$). This is a reading in the form of a conditional, and so deserves to come under the rubric of proof-attempt. A valid proof, according to this distinction, requires, in addition, that we can also provide proofs for each $\alpha \in \Gamma$.[9] A proof-attempt is analogous with a conditional for which a function does not yet exist since we do not know if there is anything for it to map; a valid proof is one in which deduction theorem holds since we have valid proofs for each premise (so they can be introduced as axioms), so the conclusion may be taken as a theorem. Co-implication is symmetrical. This way of considering implication and co-implication makes things more transparent since we can not detach on "both sides" as it were: a formula α can not have a valid proof and a valid refutation (simultaneously at least). So, whilst we can consider the hypothetical proof and refutation of implication and co-implication in one and the same moment, only one can "win".

Negation is altered by this approach since the distinction between proof-attempts and valid proofs is at odds with intuitionistic negation [13]. Simply put, the negation of α can never be valid if it is defined as $\alpha \Rightarrow 0$, whilst also requiring that in order for the conditional to be valid, there must also exist a proof of α. Whilst this is an issue for intuitionism, it provides grist to the mill for the account offered here, in which there must be a balance between negative and positive

---

[7]There is an obvious analogy to Ramsey's [16] argument that to assert a conditional is not asserting a conditional proposition, but to make a conditional assertion of the consequent.

[8]Note that one of the reasons for syntactic separation is due to the fact that L_I can not be extended by co-implication, nor can L_C be extended by implication, and any attempt to write them in by definition collapses to classical logic.

[9]Prawitz [15] makes a similar distinction between open and closed arguments, where open arguments involve undischarged assumptions, or unbound variables, whilst closed arguments contain no assumptions, and are valid just in case it is either a canonical argument (i.e. an argument that ends with an instance of an introduction rule), or it can be reduced to a canonical argument for the conclusion.

logics, and the role played by $\neg\alpha$ in constructive logic is played by $\dashv \alpha^-$. This also means that the negation defined via implication is an expressive advance on the usual interpretation of intuitionistic negation, which enables us to simply say that a proof (or refutation)-attempt *does not go through.*

2.2   Dialogical coherence and termination

This brings us to the relationship between a valid proof or refutation, and the termination of dialogue. In addition to the above rules, we have a rule that is admissible under certain circumstances, which we call TERMINAL-CUT.[10] First, define counterpart formulas between the two logics as follows: for any formula α with superscript $^{+,-}$, the counterpart formula, denoted $\alpha_\circ$, is simply the same formula with the opposite superscript, apart from $\Rightarrow$, whose counterpart is $\Leftarrow$.[11] So, for example, the counterpart of $\alpha^+ \wedge \beta^+$ is just $\alpha^- \wedge \beta^-$. With this, we define terminal-cut as follows.

$$\frac{\alpha_\circ^-;\Gamma^+ \dashv\vdash \Delta^-; \qquad ;\Gamma^+ \dashv\vdash \Delta^-;\alpha_\circ^+}{;\Gamma^+ \dashv\vdash \Delta^-;} \text{ (Terminal-cut)}$$

Terminal cut is admissible for dialogues that terminate with either prover or denier "winning", and may be understood as inducing a kind of harmony between the two logics. Harmony ordinarily operates across introduction and elimination rules inside the same logical structure, the idea being that there must be a balance between the grounds for the possible assertion of a formula with the consequences of accepting it.[12] Here, there is also a form of balance between positive and negative formulas of the logics by the fact that rules for counterpart formulas are related by de Morgan duality, and so are secondary to this balance between proof- and refutation-attempts over the course of an interaction. We define a dialogue as COHERENT iff it is the case that *if terminal-cut were to be applied*, then either: empty sequent would be "derived" in the course of a dialogue; or a canonical introduction of a proof or refutation is derived if the dialogue is terminated (in both cases we should add the clause that the proof or refutation can be reduced to a canonical proof or refutation if it is not already). Stipulating that terminal-cut must be admissible requires that a formula with a dominant operator can not be canonically introduced as both a proof and a refutation *at the termination of dialogue.*

With this in hand, we can provide a definition of validity that ensures the semantic role of proofs and refutations in a localised manner (i.e. without reference to idealised or global constraints upon soundness and completeness): Say that a proof or refutation is VALID iff: (a) It is closed; (b) It is canonical, or can be reduced to a canonical proof or refutation; (c) It is the result of a termination of a coherent dialogical interaction such that terminal-cut is admissible (without removing the formula with dominant operator). Whilst (a) and (b) are familiar from the literature mentioned above, (c) reflects the fact that we are concerned with an interaction between hypothetical proofs and tests, such that a valid formula is one that is introduced at the *end* of that interaction that terminates with an agreement between agents involved in the dialogue. The semantic role played by a valid proof or refutation is, therefore, dependent

---

[10]We need to be careful about calling this a rule, since it really just formalises termination of dialogue.

[11]The reason for this is simply due to notation since the same role is being played for proofs and $\Rightarrow$, and refutations and $\Leftarrow$

[12]Though, because we see proofs as acts, unlike in [15], harmony follows from the prior requirement that cut is eliminable. For example, we know that, for any proof of $\alpha \wedge \beta$, it must be possible to extract proofs of α, and proofs of β, since if this were not the case, then it would not be possible to form a cut-free proof with that conclusion. Cut-elimination is, therefore, a key requirement in defining the validity of proofs, which, of course, ensures the transitivity of deduction so that validity is preserved.

upon the success of either, but not both. It is not the case that rules of inference determine meaning by themselves, but rather that the correctness of these rules is already dependent upon dialogical coherence. Sloganised, the meaning of a proposition is built-up through this process of interaction between proof and refutation attempts.

As is obvious, problems of categoricity and compositionality do not arise for valid formulas, since it is no longer possible for both α^+ and α^- to be valid, and neither is it possible for $(\alpha^+ \vee \alpha^-)$ to be valid whilst neither α^+ nor α^- is (for any formula, α). Furthermore, far from requiring access to idealised Lindenbaum chains, or semantic models, a local completeness theorem arises naturally from the constraints on the validity of proofs and refutations given above: Let $|P|$ denote a terminating dialogue ending with a valid proof, $|R|$ a terminating dialogue ending with a valid refutation, and α^+, α^-, denote proof and refutation attempts of a formula α, respectively. Then, for any α^+, α^- whose interaction has terminated there exists either a valid proof P of α^+ such that $P = |P|$, or a valid refutation R of α^- such that $R = |R|$. Then, $|P|$ or $|R|$, are said to interpret the formula α. It is important to note that this theorem refers only to interactions that have terminated, it does not require that each interaction will terminate, and is localised in that it makes no reference to proof (and refutation) attempts that are outside the current context of the dialogue in question. For example, take the case of a dialogue involving a proof-attempt of $\alpha \wedge \beta$, and with an attempted refutation through α (the notation $[A_1]$ just indicates subproofs (or subrefutations)):

$$
\dfrac{\dfrac{[A_1] \qquad\quad [A_2]}{\;;\Gamma^+\vdash;\alpha^+ \quad\; ;\Gamma'^+\vdash;\beta^+}{\;;\Gamma^+,\Gamma'^+\vdash;\alpha^+\wedge\beta^+}\,(\wedge R^+) \qquad \dfrac{\dfrac{[B]}{\alpha^-;\vdash\Delta^-;}}{\alpha^-\wedge\beta^-;\vdash\Delta^-;}\,(\wedge L_1^-)}{\;;\Gamma^+,\Gamma^{+\prime}\vdash\Delta^-;}\,\text{(Terminal-cut)}
$$

Equally, "denier" could attempt to refute via β, with the same result. As is obvious, this also means that the dominant formulas can be eliminated through the usual process of cut-elimination by pushing cuts upwards. For example, for α:

$$
\dfrac{\dfrac{[A_1]}{\;;\Gamma^+\vdash;\alpha^+} \qquad \dfrac{[B]}{\alpha^-;\vdash\Delta^-;}}{\;;\Gamma^+\vdash\Delta^-;}\,\text{(Terminal-cut)}
$$

What about the case in which, say, the refutation-attempt is valid? This requires that the subrefutation $[B]$ of all $\sigma \in \Delta^+$ is valid, which we denote just by $[B_V]$. Then, the case where there exists a valid refutation of $\alpha \wedge \beta$ looks like this:

$$
\dfrac{\dfrac{[A_1] \qquad\quad [A_2]}{\;;\Gamma^+\vdash;\alpha^+ \quad\; ;\Gamma'^+\vdash;\beta^+}{\;;\Gamma^+,\Gamma'^+\vdash;\cancel{\alpha^+\wedge\beta^+}}\,(\wedge R^+) \qquad \dfrac{\dfrac{[B_V]}{\alpha^-;\vdash\Delta^-;}}{\alpha^-\wedge\beta^-;\vdash\Delta^-;}\,(\wedge L_1^-)}{\alpha^-\wedge\beta^-;\Gamma^+,\Gamma^{+\prime}\vdash\Delta^-;}\,\text{(Terminal-cut)}
$$

Where we use $\cancel{\alpha^+\wedge\beta^+}$ to indicate that the formula $\alpha^+ \wedge \beta^+$ is not introduced due to the existence of a valid refutation of α^- so the subproof attempt, $[A_1]$, of α^+ fails, and there can be no introduction of $\alpha^+ \wedge \beta^+$. Then, the above refutation R can subsequently be written just like this:

$$\frac{\alpha^-; \vdash \Delta^-;}{\alpha^- \wedge \beta^-; \vdash \Delta^-;} (\wedge L_1^-)$$

with $[B_V]$ above.

Moreover, this valid refutation R is equivalent with the version that went via a terminating dialogue, so $R = |R|$ holds.

References

[1] Gianluigi Bellin, Massimiliano Carrara, Daniele Chiffi, and Alessandro Menti. Pragmatic and dialogic interpretations of bi-intuitionism. part i. *Logic and Logical Philosophy*, 2014.

[2] Jean-Yves Béziau. Sequents and bivaluations. *Logique Et Analyse*, 44(176):373–394, 2001.

[3] Robert Brandom. *Articulating Reasons: An Introduction to Inferentialism*, volume 110. Harvard University Press, 2000.

[4] Jacques Dubucs and Mathieu Marion. Radical anti-realism and substructural logics. In A. Rojszczak, J. Cachro, and G. Kurczewski, editors, *Philosophical Dimensions of Logic and Science*, pages 235–249. Kluwer Academic Publishers, 2003.

[5] Jean-Yves Girard. A new constructive logic: classic logic. *Mathematical Structures in Computer Science*, 1(03):255–296, 1991.

[6] Jean-Yves Girard, Yves Lafont, and Paul Taylor. Proofs and types, volume 7 of. *Cambridge tracts in theoretical computer science*, 7, 1989.

[7] Michael Hand. Antirealism and falsity. In *What is Negation?*, pages 185–198. Springer, 1999.

[8] Lloyd Humberstone. *The Connectives*. MIT Press, 2011.

[9] Luca Incurvati and Peter Smith. Rejection and valuations. *Analysis*, 70(1):3–10, 2010.

[10] John MacFarlane. What is assertion? In Jessica Brown and Herman Cappelen, editors, *Assertion*. Oup Oxford, 2011.

[11] Mathieu Marion. Why play logical games? In Ondrej Majer, Ahti-Veikko Pietarinen, and Tero Tulenheimo, editors, *Games: Unifying Logic, Language, and Philosophy*, pages 3–26. Springer Verlag, 2009.

[12] Enrico Martino and Gabriele Usberti. Temporal and atemporal truth in intuitionistic mathematics. *Topoi*, 13(2):83–92, 1994.

[13] Julien Murzi. Knowability and Bivalence: Intuitionistic Solutions to the Paradox of Knowability. *Philosophical Studies*, 149(2):269–281, 2010.

[14] Julien Murzi and Ole Thomassen Hjortland. Inferentialism and the categoricity problem: Reply to raatikainen. *Analysis*, 69(3):480–488, 2009.

[15] Dag Prawitz. Remarks on Some Approaches to the Concept of Logical Consequence. *Synthese*, 62(2):153–171, 1985.

[16] Frank Ramsey. General propositions and causality. In D. H. Mellor, editor, *Philosophical Papers*, pages 145–163. Cambridge University Press, 1990.

[17] Greg Restall. Multiple conclusions. In *Logic, methodology and philosophy of science: Proceedings of the twelfth international congress*, pages 189–205. Kings College Publications, 2005.

[18] J. Trafford. Co-constructive logic for proofs and refutations. *Studia Humana*, 3(4):22–40, 2015.

[19] Igor Urbas. Dual-intuitionistic logic. *Notre Dame Journal of Formal Logic*, 37(3):440–451, 1996.

Discourse Structure and Syntactic Embedding:
The German Discourse Particle ja[*]

Yvonne Viesel

University of Konstanz, Germany
`yvonne.viesel@uni-konstanz.de`

Abstract

German discourse particles (DiPs) do not add truth-conditionally relevant meaning but are elements of speaker attitude and indicate a relation between the information in their scope (p) and another piece of information (q) in the context. The DiP ja (literally 'yes') was claimed to be felicitous with a proposition p that the speaker believes common to both speaker and hearer, or immediately verifiable. However, formalizations modeling this into the use conditions of ja fall short on the DiP's discourse function, which is to indicate that p is not used to address the current Question under Discussion but stands in a relation to q (pRq), where q is the information that the speaker makes another context update, pRq is intuitively explanatory, and p is not necessarily known to anyone but the speaker. Regarding prerequisite grammatical properties of the DiP's host constructions, data show that ja is not restricted to assertive, root-like environments and defies predictions about not being able to appear in the scope of descriptive operators. Instead the data suggest that the DiP's licitness in surprising positions depends on information-structural factors.

1   Introduction

German discourse particles (DiPs) are elements of speaker attitude and usually appear clause-medially in their scope position above vP[1], the propositional core of the clause (Bayer 2012 [1]). Combinations of DiPs show them to be subject to a rather strict hierarchical order. Characteristically, they refer to the Common Ground (CG; Egg 2012 [5]) and relate the information in their scope to (another piece of information in) the discourse. Their distribution shows some DiPs to be sensitive to the clause type and illocutionary force of their hosts, most notably $denn$ (roughly 'I wonder') and ja (literally 'yes', roughly 'uncontroversially'), which are geared to interrogatives and declaratives, respectively.

DiPs differ regarding their ability to appear in non-matrix constructions. Those found in assertions are readily found in utterance-modifying ('peripheral') adverbial clauses (ACs) and non-restrictive relative clauses (RCs). This is expected insofar as such constructions are analyzed as conventional implicatures (CIs), which have "the same semantic force as a main clause assertion" (Potts 2005: 68 [19]). With regard to proposition-modifying ('central') adverbials, restrictive modifiers, and complement clauses, however, the situation is less clear. In earlier approaches, restrictions are proposed on principled grounds, e.g. Thurmair (1989, [23]) excludes DiPs from complement clauses except those of verba dicendi, and Coniglio (2008, [4]) still excludes them from factive clauses and central adverbials as well as restrictive modifiers. His analysis is in line with observations by Liliane Haegeman on root phenomena in general

---

   [1] This also holds for cases in which topical material has scrambled to a position above the DiP.

(e.g. Haegeman 2002 [10]) and suggests the dichotomy between embedded[2] clauses to be due to a syntactic difference, with DiPs being illicit in clauses with a truncated left periphery accompanying their lack of illocutionary force. In time, more and more 'exceptions' have been recognized. Zimmermann (2011: 2023 [27]) allows for certain DiPs (*doch*, roughly 'as you should know', and *wohl*, roughly 'presumably') to appear in complements of "appropriate matrix predicates", for instance the factive verbs *vergessen* ('forget') and *bedauern* ('regret'). However, *ja*, one of the most frequent DiPs, is especially rare in embedded constructions of the non-CI type. Accordingly, it was claimed to be principally illicit in embedded position (except under verba dicendi; cf. Zimmermann 2011 [27], Kratzer 1999 [15]).

Moreover, "many" DiPs, like *schon* (roughly 'nevertheless'), *doch*, *auch* (literally 'also'), but explicitly not *ja*, have been argued to "contribute to discourse-level semantics [...] by relating *discourse segments*" (Egg 2012: 300 [5]). The present approach argues that the function of *ja* is to indicate that p is not used to answer to the current Question under Discussion (QuD) in the sense of Roberts (2012 [21]) but stands in a relation to q (pRq), where q is the information that the speaker makes another context update and pRq is intuitively explanatory. It must be acknowledged that the DiP *ja* contributes to discourse-level semantics.

Having established that *ja* is a discourse particle proper after all, original data is provided illustrating that even *ja* may be felicitously used in embedded constructions. Maintaining that the meaning and function of the DiP are the same as in main clauses, its rare occurrence in semantically integrated environments is more plausibly a mere side effect of its severely restricted use conditions. At any rate, the empirical situation of DiPs cannot be fully explained with recourse to the syntax of clausal left peripheries or the presence of matching illocutionary operators in types of constructions. The question will briefly be addressed if there are nevertheless grammatical prerequisites to be met by the host constructions for the successful employment of *ja*. The tentative answer at this point is that, yes, the role of *ja*-marked information in discourse is mirrored in the information structure of the DiP's complement, with phonological focus being the least sufficient reflex thereof in reduced constructions like adjectival phrases (APs).

2   The DiP *ja* contributing to discourse-level semantics

The DiP *ja* is perceived to mark information as part of the CG (cf. Zimmermann 2011: 2012 [27]). But formalizations as by Gutzmann (2009: 53 [9]; cf. Kratzer 2004: 128 [14]) do not capture the 'use conditions' of a *ja*-utterance. As argued subsequently, it is neither sufficient nor necessary to fulfill the requirements in (1). In (1a) as well as the present account, p refers to the first semantic argument of *ja*, usually the content proposition expressed in the vP of a declarative clause:

(1)   a. $[\![\lambda p.\mathbf{ja}(p)]\!]^c =$ that function $f \in \{f | f : D_{\langle s,t \rangle} \mapsto D_{\langle s,u \rangle}\}$ such that $f(p)(w) = \sqrt{}$, if c_S believes that p is common knowledge of c_S and c_H in w, or that it is verifiable on the spot that p, else $f(p)(w) = \frac{1}{4}$.   (cf. Gutzmann 2009: 53 [9])

       b. $[\![$David ist ja ein Zombie$]\!] = \langle 1, \sqrt{} \rangle$, if it is true that David is a zombie and if c_S believes that it is common knowledge that David is a zombie.

Similar intuitions have been expressed by Zeevat (2004: 181–182 [25]) – "*ja* presents [p] as common ground between speaker and hearer" – and Kratzer & Matthewson (2009: 6 [16]) –

---

[2] 'Embedded' is used as an umbrella term for all 'subordinate' constructions, semantically and syntactically 'integrated' ones (cf. Potts 2005 [19]), e.g. restrictive relatives, as well as their 'unintegrated' counterparts.

the speaker "takes p to be an established fact". But as (2) proves, "repeating old information is not useful by itself" (Zeevat 2004: 181 [25]): p being shared knowledge is insufficient to make *ja*(p) felicitous:

(2) At the bus stop, Anna meets a new neighbor, Bela, who is holding by the leash what is obviously his dog.

    Bela: *Das ist **ja** mein Hund.*
               'This [JA] is my dog.'
    Anna: ...   (*ja   und*)?
                  yes   and
             '... (so what)?'
    Bela: a. # ∅
              b. √ 'Do I need a bus ticket for him?'

Bela's *ja*-assertion that the dog is his will leave Anna puzzled (2a) unless it supplements a second context update, e.g. Bela's question if a bus ticket is needed for the dog (2b; cf. Viesel, to appear [24]).[3] As to the necessity for p to comply with (1), cf. (3):

(3) Anna: 'I have not heard of lonesome Cedrick in ten years. What about you?'
    Bela: *Also – Cedrick hat **ja** vor fünf Jahren Didi geheiratet,*
           'Well – Cedrick [JA] married Didi five years ago,
           und jetzt haben sie schon zwei Kinder und wohnen in einer Kommune!
           and now they have two children already and live in a commune!'

Bela uses *ja* in (3) even though he knows Anna not to know p. The effect is that Anna will not mistake p for a complete answer to her question. The *ja*-assertion does not convey Bela's main point, but merely provides the background for what is to come. Thus, Anna is less likely to interrupt him at the first piece of joyous news.

Of course, the use of *ja* in contexts in which it is obvious to all that p is hearer-new is rather exceptional.[4] But it shows that speakers may use *ja* even if they know the addressee not to know p. Refraining, therefore, from an attempt to classify *ja*-marked propositions as CG or established information, what can be maintained is that the speaker "doesn't consider the question whether or not p to be an issue for either the current or any future inquiry" (Kratzer & Matthewson 2009: 6 [16]). This is our first general conclusion and a significant specification of (1): a *ja*-assertion does not answer to the current QuD.[5]

This is precisely the reason why the felicity of *ja*(p) hinges on the second semantic argument of *ja*, q, which does not feature in (1) at all. Both q and the relation between p and q (pRq) are notoriously hard to characterize. In Kratzer & Matthewson's (2009: 8 [16]) terms, "there must always be some suitable connection to another salient fact". Similarly, in light of classic constellations such as (4), it is tempting to suggest that p "serve[s] as background for the

---

[3] Note that in the absence of *ja*, Anna would take the introduction of the dog as Bela's main point and proceed accordingly:

    (i)   Bela: *Das ist mein Hund.*
          Anna: 'He's cute.'

[4] However, cf. Viesel (to appear [24]) for remarks on a use of the DiP as an argumentative trick.

[5] Along with both Zeevat (2006 [26]) and Hinterhölzl & Krifka (2013 [11]), the present analysis rejects the idea that information under DiPs may be presupposed or have to be accommodated despite being provided in the form of an assertion. The presuppositional flavor associated with such assertions is a by-product of the function of *ja*, according to which p is added to the CG without discussion (if it is not already there anyway).

evaluation of the proposition [q] that is to be put on the table and that actually is at-issue[6] [at a specific point in the discourse]" (Hinterhölzl & Krifka 2013: 12 [11]):

(4) **Die Schere ist ja rund** (*wie Sie sehen*). *Sie ist für Kinder geeignet.*
'The scissors [JA] are blunt (as you can see). They are suitable for children.'

Here, proposition p (*The scissors are blunt*) incidentally adheres to the definition in (1), but, what is more, it appears to supplement the associated utterance. Based on both Kratzer & Matthewson's (2009: 8 [16]) and Hinterhölzl & Krifka's (2013 [11]) descriptions, the "salient fact" q, itself to be evaluated in light of p and 'at-issue in the current discourse', would be the proposition that the scissors are suitable for children. In the style of Egg (2012 [5], on *schon*, *doch* and *auch*), *ja* might then be assumed to indicate a relation between two content propositions in terms of defeasible inference:

(5) Preliminarily:
$\llbracket ja \rrbracket(p)(q)$ iff p and q are true, p is not used to answer to the QuD, and p defeasibly entails q ($p > q$).

However, an entailment relation as in (4) is absent in (6) and (7):

(6) *Die Schere ist **ja** rund* (*wie Sie sehen*). *Trotzdem ist sie für Kinder ungeeignet.*
'The scissors [JA] are blunt (as you can see). Nevertheless they are unsuitable for children.'

(7) After TV footage just showed the winning couple, a reporter is interviewing a witness on a dance competition.
Reporter: 'What colors were in fashion this year?'
Witness: *Naja, das Gewinnerpärchen war **ja** dunkel gekleidet,*
'Well, the winning couple [JA] was dressed in dark colors,
bei den neuen Paaren waren Herbstfarben beliebt, [...].
among the new couples, fall colors were popular, [...].'

The relation between (the content propositions) p and q is the exact opposite in (4) and (6) – p>q vs. p> ¬q in the concessive case. In (7), no entailment relation whatsoever holds between the content propositions p and q. Finally, recall that in (2b), p justifies the speaker to ask a question.

It is worth comparing *ja* to *then*. Modulo general differences between DiPs and discourse markers and the fact that the first and second arguments of the two lexical items swap roles, the relations the items establish between context updates are much alike. According to Biezma (2014: 373 [3]), "*then* requires that two propositions enter into a 'causal explanatory claim'-relation in which one (the antecedent) provides the 'reasons' for the other (the consequent)." These propositions need not be content propositions: *then* "signals that the utterance of the embedded clause is in some sense motivated by the preceding discourse move" (Biezma 2014: 380 [3]). Similarly, *ja* signals that the content proposition p of the embedded clause in some sense motivates another discourse move.

This allows for a revision of (5), cf. (8). By the definition suggested, q is not a content proposition and, by nature, not put on the table or under discussion. Rather than allowing for

---

[6] The latter point helps to grasp most cases, like (4), although actually, a *ja*-utterance can be used to provide support for, say, another *ja*-utterance. It is not necessary that the associated context update addresses the QuD (i.e. "is at-issue" in the discourse). This minor point may be made without illustration, as q will be argued not to coincide with the content proposition of the associated context update anyway.

q to be either a content proposition or something else, (8) generalizes over all cases without recourse to relations that may or may not hold between content propositions:

(8) $[\![ja]\!](p)(q)$ iff p is true, p is not used to answer to the QuD, and p explains q, where q is the proposition that **the speaker asks a question or makes an assertion or a request**.

Thus, under *ja*, content proposition p justifies that another context update is made, be it an assertion, question or request.[7] To illustrate, (4), (6), (7) and (2) can be roughly paraphrased in a completely parallel fashion as in (9):

(9) a. Since (unquestionably) the scissors are blunt, **the speaker asserts** that they are suitable for children / nevertheless unsuitable for children.   (cf. (4), (6))

b. Since (unquestionably) the winning couple was dressed in dark colors, **the speaker asserts** that among the new couples, fall colors were popular.   (cf. (7))

c. Since (unquestionably) the dog is the speaker's, **the speaker asks** whether a bus ticket is required for the dog.   (cf. (2))

The matter of causal explanations or explanatory claims is complicated (cf. the discussion in Biezma (2014 [3]). A final analysis of pRq along the lines of (9) will depend on the definition of the discourse relation *Explanation*. But the discourse function of *ja* clearly pertains to this area of research, so that pRq as specified above can for instance be shown to correspond to an extent to a suggestion by Lascarides & Asher (1991: 58 [17]), according to whom the relation *Explanation* normally holds if p causes q:

(10) a. **Explanation:** $\langle \alpha, \beta \rangle \land cause(me(\beta), me(\alpha)) > Explanation(\alpha, \beta)$

b. In words, if α and β are discourse-related and the event described in β caused the event described in α, then $Explanation(\alpha, \beta)$ normally holds.

Following (8), *ja*, just like *then* (cf. Biezma 2014 [3]), relates propositions and not events. But if in (10), $me(\beta)$ and $me(\alpha)$, the 'main eventualities' described by propositions (Lascarides & Asher 1991 [17]), are replaced by the context updates in which *ja*(p) (β) and q (α) are introduced, then the first arguably causes the second, so that the fact that the winning couple was unquestionably dressed in dark colors in (9b) explains that the speaker asserts that among the new couples, fall colors were popular.

Even if the argumentation so far needs refinement, the present proposal agrees with Rojas-Esponda (2013: 131 [22]), who considers *ja* one of the 'roadsigns of communication' which "speakers can use to signal their views and preferences about the QUDs". The present account has argued that *ja* indicates information to justify another utterance. Therefore, the fact that *ja*(p) does not answer to the QuD is not "informative in itself", and *ja* does contribute to discourse-level semantics, contrary to Egg (2012: 299-300 [5]). The difficulties in unifying *ja* with the other members of its class apparently arise from a meaning that seems "relatively unspecific" (Zimmermann 2011: 2012 [27]), but only as long as just the first argument of the DiP is considered.

---

[7] For the sake of completeness, an imperative case:

(i) *Frag einfach Norman!* **Lena ist ja grad nicht da.**
     'Just ask Norman! Lena [JA] is not available right now.'

3 Embedded *ja* and discourse structure

Why is *ja* so rare in some embedded environments? Since the DiP occasionally does show up in unpredicted places, this cannot be fully explained by structural difference in the CP-system of subordinate clauses or hinge on the presence of illocutionary operators in, e.g., central ACs and restrictive modifiers as opposed to their peripheral and non-restrictive counterparts.[8]

So far, the present account is in line with recent research by Jacobs (2015 [12]). He proposes three contextual restrictions for *ja*, of which only one will be addressed in greater detail below. First, *ja* is banned from non-veridical environments. This restriction seems to hold apart from some quirky occurrences of *ja* in event-conditionals, showing just how fuzzy the boundaries of acceptability really are.[9] Second, *ja* is banned if p is activated. This restriction has to be evaluated for its interaction with information structure in future research. Third, Jacobs claims *ja* to be banned from appearing, syntactically, in the scope of descriptive operators because it is not in their scope semantically.

Jacobs' (2015 [12]) ban of *ja* from the scope of descriptive operators is argued to hold for cases of *ja* scoping over (the base positions of) quantifier-bound pronouns supposedly causing intervention effects (Kratzer 1999 [15]), restrictive RCs, and, e.g., temporal ACs. These are analyzed as containing a time variable bound by a definiteness operator scoping over the DiP. But even though *ja* is not 'scoped out' by such operators, it may materialize in a position in their scope domain:

(11) a. *sagen will keiner$_{[i]}$ was, weil **er$_{[i]}$ ja** [t$_i$] **die Jäger selber kennt.***
   'nobody wants to say anything because [JA] he knows the hunters himself.'
   (http://www.griechenland-blog.gr/2013/12/allgemeines-jagdverbot-in-griechenland
   /79429/, 11/26/2015)

   b. *Mit Herrn K. bekommt die Firma einen Angestellten, **der ja immer pünktlich ist.*** (cf. Viesel, to appear [24])
   'With Mr. K., the firm gets an employee who [JA] is always on time.'

   c. *ist es ihm einfach egal, weil er eh der "letzte" Jedi ist bevor **er ja die Akademie auf Yavin eröffnet?***[10]
   'does it just not matter to him because he is anyway the "last" Jedi before [JA] he opens the academy on Yavin?'
   (http://phoenix-hort.de/joomla/component/kunena/26-taverne/58761-aw-luke-
   bricht-den-codex, 03/31/2015)

Whatever the details[11], in all cases, the scope of the DiP is restricted to the content proposition of the containing subordinate clause (or to the information that a discourse referent has the

---

[8] As Jacobs (2015: 5 [12]) points out, the assumption of licensing illocutionary elements would lead to an inflation of structural ambiguities which are nevertheless context-dependent. Moreover, it is empirically implausible, too, that a language like German, while distinguishing by means of surface features over 30 conventional combinations of sentence form and illocutionary type in main clauses, should not once mark overtly whether countless numbers of embedded clauses and APs even have a defined illocutionary type at all.

[9] (i) *Falls **man ja wirklich mal unnötige Dinge schon länger mit sich rumschleppt***
   lohnt es sich zwischendurch mal ein Paket nach Hause zu schicken.
   'In case [JA] one really has lugged around unnecessary things for a while
   it is worth sending a parcel home intermittently.'
   (http://weltreise-info.de/forum/index.php?topic=1936.0, 03/31/2015)
Space forbids to do the finding justice.

[10] The potential scope domains of *ja* are indicated in boldface. Recall that scrambled elements are interpreted in their base positions below the DiP. The position of '[JA]' in the translations is intended to capture this fact best possible.

[11] Additional factors influence the acceptability of *ja*. E.g. restrictive modifiers (11b) may only contain

properties denoted by the AP and NP). Accordingly, the illocutionary force of the embedding (matrix) clause does not matter. (11) also shows that it is as yet unclear why especially *ja* is very rare in certain environments.

Plausibly, certain syntactic environments are just very frequent in contexts incompatible with the specific meaning of this particle. As specified in section 2, *ja* marks information which does not answer to the QuD. Indeed, the DiP cannot be used in the answer to a corresponding question. This major limitation of its host's employment options is an idiosyncrasy of *ja* not shared for instance by *doch* although the two are close in meaning (cf. Grosz 2014 [7]):

(12) Anna: Where is Constantin?
      Bela: *Er ist (*ja/$^{\surd}$doch) in Delhi.*
             'He ([JA]/[DOCH]) is in Delhi.'

In a question-based discourse-model (cf. Roberts 2012 [21]), this explains naturally why *ja* is so rarely encountered in complement clauses, central ACs and restrictive modifiers. In (13), information in the potential scope of *ja* answers to the QuD, *ja* is out:

(13) a. [...] *das sind Menschen wie wir.*   (*Die Zeit*, 12/30/2014)
         'these are people like us.
         (QuD: What about people like us? Subquestion: What about one such person?)
         Wir bringen unser Geld zu einem Mann, der [*ja] grüne Wollpullover trägt.
         We take our money to a man who [JA] wears green woolen pullovers.'

      b. Anna: 'What do you know about the main suspect?'
         Bela: *Wir wissen dass er (*ja) grüne Wollpullover trägt.*
                'We know (that) [JA] he wears green woolen pullovers.'

Thus there is a good pragmatic reason why *ja* might be illicit in most restrictive subordinate constructions, as opposed to supplementary constructions of the CI type that are generally taken to be semantically and syntactically non-integrated. However, in the right contexts, information in restrictive subordinate clauses merely supplements the matrix assertion or question, which itself answers or comprises the QuD. The lexical content p could be taken away and asserted in isolation, and the containing utterance would still be informative with the dependent clause replaced by an anaphor, contrary to the situation in (13). This is illustrated in (14a), where *ja* appears in a factive clause:

(14) a. *Haben sie erkannt, daß sie ja auf genau der gleichen Linie liegen? Weit gefehlt!*
         'Have they realized that [JA] they are exactly on the same page? Far from it!'
         (http://bitflow.dyndns.org/german/FranzGrafStuhlhofer/Das_Ende_Naht.html, 03/14/2015)

---

the DiP in DPs with specific reference, and different adverbs (*leider* – 'unfortunately', *immer* – 'always', ...) improve the DiP's acceptability. The acceptability of *ja* in very marked places, e.g. between prepositions and determiners, or in DPs without adnominal modifiers, depends on being accompanied by other DiPs and adverbs:

(i) [*die Krise.* ...] *diese* **ja** **auch** **fast** *Auflösung* *der* *Band*
     ['the crisis. ...]   this   [JA]   [AUCH] nearly   dissolution the.GEN band'
     (http://www.arte.tv/de/ja-panik-sind-die-laessigsten-poptheoretiker-der-berliner-indieszene/ 7825136,CmC=7822214.html, 08/29/2014)

(ii) *nach* **ja** **auch** **wieder** *der* *internationalen* *Einbindung* (*Berliner Zeitung*, 07/31/2004)
      'after   [JA]   also   again   the   international   involvement'

b. **Sie liegen ja auf genau der gleichen Linie.** *Haben sie das erkannt? Weit gefehlt!*

'They [JA] are exactly on the same page. Have they realized that? Far from it!'

c. Since (unquestionably) they are exactly on the same page, the speaker asks whether they have realized that.

According to Hinterhölzl & Krifka (2013: 12 [11]), central adverbials and restrictive modifiers may occur in a "non-default use", contributing 'background assertions', the content propositions of which are "directly added to the CG without discussion (without being put on the table [cf. Farkas & Bruce 2009 [6]])". Background-asserted information is "needed for evaluating the assertion of the main clause" (Hinterhölzl & Krifka 2013: 12 [11]) – a question in (14a) – with the assertiveness of the dependent clause being accompanied by focus and focus particles.

There is independent evidence from phonological research that information that provides "an answer [...] to a supplemental question [to the immediate QUD]" (Riester & Baumann 2013: 233 [20]) is indeed prominent. For instance, factives as in (15) may be "not marked by means of a nuclear pitch accent but by some pre- oder postnuclear prominence" indicative of a 'secondary' focus (Riester & Baumann 2013: 217 [20]):

(15) 'Everyone knew that Mary only eats VEgetables$_{F1}$. If even PAUL$_{F2}$ knew that Mary only eats VEgetables$_{SOF}$, then he should have suggested a different restaurant.'[12]

Instances of capitalization in informal written German show how some speakers implement focus to enforce the background-assertive reading for the integrated clause under *ja*:

(16) *Aber seit **er ja nun sich ÜBERALL hochzieht**, warte ich nur auf den Moment, wo er den Wasserhahn in der Wanne aufdreht.*   (original capitalization)
'But ever since [JA] he now pulls himself up EVERYWHERE, Im only waiting for the moment where he runs the water tap in the bathtub.'
(http://www.babyforum.de/threads/1214818-43269829432698294326-maerzbabys-2014-43269829432698294326/page798?p=8770370#post8770370, 03/31/2015)

Moreover, a 2014 corpus study conducted in the DWDS ('Digital Dictionary of the German Language', cf. Klein & Geyken 2010 [13]) has shown APs with *ja* to be almost always lexically and structurally complex.[13] DPs with a single simple, non-modified adjective under *ja* are judged inelegant to inacceptable in isolation – a problem for the common view that any non-restrictive adjective could license DiPs. But exceptional cases are found (3.3% of corpus findings) and are perfectly fine in the right context when heavily stressed, as in (17b):

(17) a. *Neulich habe ich meinen **ja** *(**schwer**)**reichen** Onkel gesehen.* (constructed)
recently have I   my   JA (heavily)rich   uncle seen

b. *Könnte es sein, dass die an **den ja giftigen Blättern** gestorben sind?*
'Could it be that they died from the [JA] poisonous leaves?'
(http://www.koi-live.de/viewtopic.php?t=36499&sid=
3f8195eed3ae23e0edec30c5ca98c7b2, 03/29/2015)

---

[12] Cf. Riester & Baumann (2013: 217 [20]) for discussion of their example (15) and the literature on 'second occurrence focus' (SOF).

[13] Cf. Viesel (to appear [24]). The reasons are unclear and could be of structural or pragmatic nature, e.g. to prevent a simple attributive reading by enforcing a clausal/predicational structure or to maximize activation.

In (17b), the causal relation between the dying event and the leaves is itself currently under discussion, and p is the information that the leaves are poisonous. An intonation where the adjective receives more than regular phrase stress suggests itself (a 'not-at-issue focus', Arndt Riester, p.c.). Similarly, *ja* is fine with simple adjectives in contrastive focus. Finally, some speakers implement marked intonation by making creative use of capital letters, hyphens, commas and brackets in the adjectival domain.

In sum, *ja* can appear in genuinely embedded non-root constructions as long as it relates pieces of information as in section 2. It is not banned from any syntactic environment even though it is not semantically embeddable under descriptive operators. But much positive evidence indicates that the right environment for the DiP is distinguished information-structurally. If the DiP's immediate host lacks root properties, focus is the last remaining indicator of the additional supplementary discourse function of p.

4   Conclusions

The DiP *ja* has been argued to fulfill a very specific and clearly relational discourse function. Prototypically, the two propositions related are associated with two independent utterances. The content proposition of a *ja*-assertion is not directly relevant to the QuD but justifies another assertion, question or request. In restrictive modifiers, central adverbials, factive clauses and many more environments, the DiP is rare, but not impossible. Its use is dependent on a discourse relation between pieces of information, just as with *ja* in root clauses, and whenever such a relation is expressed, the independent discourse status of p is reflected information-structurally.

While it is too early to speculate about *ja* being conventionally focus sensitive in the sense of Beaver & Clark (2008 [2]), efforts must be taken to further elucidate the relation between focus and *ja*. Earlier statements expressly denying the focus sensitivity of this DiP must be handled with care (e.g. Grosz 2014 [7][14]). The DiP patterns with other members of its class regarding its relational discourse function, and difficulties in embedding *ja* have been reduced to its specific meaning instead of peculiarly strict grammatical requirements. If in the end *ja* should prove to be focus-sensitive, this would only further unify the item and other notorious members of its class which have been argued to display (conventional) focus sensitivity (e.g. Egg 2012 [5] on *schon*, Grosz to appear [8] on *doch*, Mursell 2013 [18] on *wohl*).

References

[1] Josef Bayer. From modal particle to interrogative marker: A study of German 'denn'. In Laura Brugé, Anna Cardinaletti, Giuliana Giusti, Nicola Munaro, and Cecilia Poletto, editors, *Functional Heads*, volume 7 of *Oxford studies in comparative syntax*, pages 13–28. Oxford University Press, 2012.

[2] David I. Beaver and Brady Clark. *Sense and Sensitivity: How Focus Determines Meaning*, volume 5 of *Explorations in semantics*. Blackwell, Malden, 2008.

[3] María Biezma. The grammar of discourse: The case of 'then'. In *Proceedings of SALT 24*, pages 373–394, New York, 2014.

[4] Marco Coniglio. *Die Syntax der deutschen Modalparktikeln: Ihre Distribution und Lizensierung in Haupt- und Nebensätzen*. PhD thesis, University of Venice, 2008.

[5] Markus Egg. Discourse particles at the semantics-pragmatics interface. In Werner Abraham and Elisabeth Leiss, editors, *Modality and Theory of Mind*, pages 297–333. Gruyter, Berlin, 2012.

---

[14] I thank Patrick Grosz for a recent discussion of this issue.

[6] Donka F. Farkas and Kim B. Bruce. On reacting to assertions and polar questions. 2009.

[7] Patrick Grosz. German 'doch': An element that triggers a contrast presupposition. In *Proceedings of the Chicago Linguistic Society 46*, pages 163–177, University of Chicago, 2014.

[8] Patrick Grosz. Information structure and discourse particles. In Caroline Féry and Shin Ishihara, editors, *OUP Handbook of Information Structure*. Oxford University Press, Oxford, to appear.

[9] Daniel Gutzmann. Hybrid semantics for modal particles. *Sprache und Datenverarbeitung*, 33(1-2):45–59, 2009.

[10] Liliane Haegeman. Anchoring to speaker, adverbial clauses and the structure of CP. *Georgetown University Working Papers in Theoretical Linguistics*, 2:117–180, 2002.

[11] Roland Hinterhölzl and Manfred Krifka. Modal particles in adverbial and adnominal clauses: Ms., Università Ca'Foscari Venezia, Humboldt-Universität Berlin. Ms., Università Ca'Foscari Venezia, Humboldt-Universität Berlin, 2013.

[12] Joachim Jacobs. Deutsche Modalpartikeln als Hauptsatzphänomene: Handout, Universität Konstanz, 04.06.2015.

[13] Wolfgang Klein and Alexander Geyken. Das Digitale Wörterbuch der Deutschen Sprache (DWDS). *Lexicographica: International Annual for Lexicography*, 26:79–96, 2010.

[14] Angelika Kratzer. Interpreting focus: Presupposed or expressive meanings? A comment on Geurts and van der Sandt. *Theoretical Linguistics*, 30:123–136, 2004.

[15] Angelika Kratzer. Beyond 'Oops' and 'Ouch': How descriptive and expressive meaning interact: Handout, Cornell Conference on Theories of Context Dependency, 26.03.1999.

[16] Angelika Kratzer and Lisa Matthewson. Anatomy of two discourse particles: Handout, SULA 5, 15.-17.05.2009.

[17] Alex Lascarides and Nicholas Asher. Discourse relations and defeasible knowledge. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 55–63, 1991.

[18] Johannes Mursell. Discourse particles and their interactions in the clause: Presentation, Wien, 02.10.2013.

[19] Christopher Potts. *The Logic of Conventional Implicatures*. Oxford University Press, Oxford, 2005.

[20] Arndt Riester and Stefan Baumann. Focus triggers and focus types from a corpus perspective. *Dialogue and Discourse*, 4(2):215–248, 2013.

[21] Craige Roberts. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69, 2012.

[22] Tania Rojas-Esponda. The roadsigns of communication. In Raquel Fernández and Amy Isard, editors, *SEMDIAL 2013 DialDam*, pages 131–139, Amsterdam, 2013.

[23] Maria Thurmair. *Modalpartikeln und ihre Kombinationen*, volume 223 of *Linguistische Arbeiten*. Niemeyer, Tübingen, 1989.

[24] Yvonne Viesel. Discourse particles 'embedded': German 'ja' in adjectival phrases. In Josef Bayer and Volker Struckmeier, editors, *Discourse Particles*, Linguistische Arbeiten. Gruyter, Berlin, to appear.

[25] Henk Zeevat. Particles: Presupposition triggers, context markers or speech act markers. In Reinhard Blutner and Henk Zeevat, editors, *Optimality Theory and Pragmatics*, Palgrave studies in pragmatics, language and cognition, pages 163–199. Palgrave Macmillan, Basingstoke, 2004.

[26] Henk Zeevat. A dynamic approach to discourse particles. In Kerstin Fischer, editor, *Approaches to Discourse Particles*, pages 133–148. Elsevier, Amsterdam, 2006.

[27] Malte Zimmermann. Discourse particles. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: Volume 1*, volume 33.1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, pages 2011–2038. Gruyter, Berlin, 2011.

Simplifying Counterfactuals

Malte Willer

University of Chicago
`willer@uchicago.edu`

Abstract

The fact that counterfactuals in general license simplification of disjunctive antecedents is a familiar problem for the traditional Lewis-Stalnaker variably strict analysis of counterfactuals. This paper argues that recent semantic attempts to solve this problem in a variably strict setting do not address related simplification patterns and demonstrates that the data are well explained by a dynamic strict analysis of counterfactuals using ideas from the inquisitive semantic tradition.

1 The Plot

On a textbook variably strict analysis ([11],[18]) a *would*-counterfactual $\ulcorner \phi \,\square\!\!\rightarrow \psi \urcorner$ is true at some possible world w iff $f_c(w, \llbracket \phi \rrbracket) \subseteq \llbracket \psi \rrbracket$, where $f_c(w, \llbracket \phi \rrbracket)$ denotes the ϕ-worlds closest to w given some contextually provided similarity relation between worlds. The fact that counterfactuals in general license Simplification of Disjunctive Antecedents (SDA) is a familiar problem for this analysis ([5],[13]):

SDA $(\phi \vee \psi) \,\square\!\!\rightarrow \chi \models \phi \,\square\!\!\rightarrow \chi, \psi \,\square\!\!\rightarrow \chi$

Given classical disjunction, SDA inferences are unexpected in a variably strict setting since $f_c(w, \llbracket \phi \vee \psi \rrbracket) \neq f_c(w, \llbracket \phi \rrbracket) \cup f_c(w, \llbracket \psi \rrbracket)$ unless the closest ϕ-worlds are just as close to w as are the closest ψ-worlds, leaving it unexplained why (1) entails that the party would have been fun if Alice had come *and* that the party would have been fun if Bert had come:

(1) If Alice or Bert had come to the party, it would have been fun.
    a. ⤳ If Alice had come to the party, it would have been fun.
    b. ⤳ If Bert had come to the party, it would have been fun.

The first observation of this paper—discussed in the remainder of this section—is that the trouble with simplification goes beyond SDA in interesting ways. The second observation is that the data are well explained by a dynamic strict analysis of counterfactuals that uses ideas from the inquisitive semantic tradition.

Semantic attempts to deal with the SDA trouble in a variably strict setting exist but do not address the full range of simplification data. One idea is to go with a Hamblin-style analysis of disjunction and then let *if*-clauses be universal quantifiers so that $\ulcorner \phi \,\square\!\!\rightarrow \psi \urcorner$ is now true at some possible world w iff $f_c(w, p) \subseteq \llbracket \psi \rrbracket$ for all propositions p in the set of alternatives denoted by the antecedent ϕ—a singleton in case of non-disjunctive antecedents; a set containing all the atomic propositional disjuncts if the antecedent is a disjunct (see [2]). Another idea is to adopt an existential analysis of disjunction—the antecedent of (1), for instance, would be of the form '$\exists x.\, Cx \wedge (x = \text{Alice} \vee x = \text{Bert})$' with 'C' denoting the property of coming to the party—and a treatment of indices of evaluation as world-assignment pairs. If we now say that two such pairs are unconnected (and hence none of them more similar to the index of evaluation than the other) if their assignments differ, we predict that the counterfactual selection function includes indices at which Alice comes to the party as well as indices at which Bert comes to the party (see [16]).

Both approaches explain why (1) simplifies but are hand-tailored to handle simplifications of *disjunctive antecedents*: what does the explanatory lifting in each approach is the special interaction between a non-classical analysis of disjunction with whatever is

involved in interpreting *if*-clauses. And this cannot be the whole story since a counterfactual like (2) with a *negated conjunction* as antecedent also simplifies (see [14]) and since a *might*-counterfactual such as (3) allows for simplification of its disjunctive *consequent*:

(2) If Nixon and Agnew had not both resigned, Ford would never have become president.
    a. ⤳ If Nixon had not resigned, Ford would never have become president.
    b. ⤳ If Agnew had not resigned, Ford would never have become president.

(3) If Mary had not gone to Pisa, she might have gone to Lisbon or Rome.
    a. ⤳ If Mary had not gone to Pisa, she might have gone to Lisbon.
    b. ⤳ If Mary had not gone to Pisa, she might have gone to Rome.

It will not do to just stipulate that (2) is evaluated by checking its consequent against the union of the closest worlds in which Nixon and Agnew do not resign, respectively: this fact calls for an explanation in terms of negation and conjunction just as much as SDA called for an explanation in terms of disjunction.[1] Explaining the simplification pattern exhibited by *might*-counterfactuals such as (3) is also challenging in a variably strict setting under the reasonable assumption that '$\diamondsuit\!\!\rightarrow$' and '$\square\!\!\rightarrow$' are duals ($\phi \diamondsuit\!\!\rightarrow \psi =_{\mathrm{def}} \neg(\phi \square\!\!\rightarrow \neg\psi)$). For observe that given some context c, the truth of $\ulcorner\neg(\phi \square\!\!\rightarrow \neg(\psi \vee \chi))\urcorner$ at some possible world w only requires that $f_c(w, \llbracket\phi\rrbracket) \cap \llbracket\psi\rrbracket \neq \varnothing$ or $f_c(w, \llbracket\phi\rrbracket) \cap \llbracket\chi\rrbracket \neq \varnothing$, not both. In any case, a comprehensive story about how and why counterfactuals simplify cannot be dependent on the interpretation of *if*-clauses (or on considerations about the similarity relation) alone since simplification is also a feature of disjunctive counterfactual consequents.

The Lewis-Stalnaker analysis has a problem with simplification that goes beyond the familiar observations about SDA. One may, of course, couple such an analysis with a powerful pragmatic supplement (see for instance [7],[17]) but in the following §2 I demonstrate that the simplification data already receive a straightforward explanation in a suitably elaborated strict analysis of counterfactuals ([6],[8],[24]). §3 shows how to avoid the most pressing problems with any theory that treats SDA as valid, including the observation that certain counterfactuals seem to resist simplification.

2   Basic Framework

The proposal to be developed here is that a *would*-counterfactual $\ulcorner\phi \square\!\!\rightarrow \psi\urcorner$ is a strict material conditional $\ulcorner\square(\phi \supset \psi)\urcorner$ over a contextually determined but dynamically evolving domain of quantification presupposing that its antecedent ϕ is a possibility in that domain. Assuming duality and that presuppositions project out of negation, a *might*-counterfactual $\ulcorner\phi \diamondsuit\!\!\rightarrow \psi\urcorner$ amounts to an assertion of $\ulcorner\diamondsuit(\phi \wedge \psi)\urcorner$ under the presupposition that ϕ is possible. Once we predict that $\ulcorner\diamondsuit(\phi \vee \psi)\urcorner$ and $\ulcorner\diamondsuit\neg(\neg\phi \wedge \neg\psi)\urcorner$ entail $\ulcorner\diamondsuit\phi\urcorner$ *as well as* $\ulcorner\diamondsuit\psi\urcorner$ via a semantic free choice effect, we predict that (1)–(3) simplify in the way they do.

The target language $\mathcal{L}$ contains a set of sentential atoms $\mathcal{A} = \{p, q, ...\}$ and is closed under negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), the modal possibility operator ($\diamondsuit$), and the *would*-counterfactual ($\square\!\!\rightarrow$). Other connectives are defined in the usual manner. I begin by stating a simple semantics for $\mathcal{L}$ that predicts the free choice effect and the key simplification observations. The semantics does not give us everything one might hope for but—as I will show in the next section—many potential shortcomings are avoided once we add some bells and whistles to the basic story.

---

[1]Not all counterfactuals with negated conjunctions simplify, as an anonymous reviewer helpfully points out: 'If John had not had that terrible accident last week and died, he would have been here today' does not license 'If John had not died, he would have been here today' since he still might have had that accident. What underlies this observation, I suggest, is that negated conjunctions sometimes give rise to a 'neither' rather than a 'not both' reading (see [20] for detailed discussion) and that, unsurprisingly, only the latter licenses simplification.

2.1 Modals

The analysis of modals is in the spirit of Veltman's approach ([21]) but here we do not treat input contexts as sets of possible worlds but as sets of consistent propositions (which I label here *alternatives*).

Definition: Possible Worlds, Propositions. w is a *possible world* iff $w \colon \mathcal{A} \to \{0,1\}$. W is the set of such w's, $\mathcal{P}(W)$ is the powerset of W. The function $[\![\cdot]\!]$ assigns to nonmodal formulas of $\mathcal{L}$ a *proposition* in the familiar fashion. $\bot$ is the contradictory proposition (the empty set of possible worlds) while $\not\leq$ is any consistent proposition.

Definition: States, Alternatives.   A *state* $s \subseteq \mathcal{P}(W)\backslash \bot$ is any set of consistent propositions (*alternatives*). S is just the set of all such states.   The *information* carried by a state s is the set of possible worlds compatible with it so that $\texttt{info}(s) = \{\bigcup \sigma \colon \sigma \in s\}$. We refer to $\varnothing$ as the *absurd* state and speak of $s_0 = \mathcal{P}(W)\backslash \bot$ as the *initial* state.

States thus have informational content in the sense that they rule out certain ways the world could be. In addition, they encode this information as a set of alternatives (which do not have to be mutually exclusive).

   States are updated by updating each of their alternatives. Updates on an alternative σ are sensitive to the state s containing it since modals perform tests on the state's informational content. Furthermore, I will think of update rules as *relations* between alternatives to capture the inquisitive effect of disjunction and distinguish between a *positive* acceptance inducing update relation $[\cdot]_s^+$ and a *negative* rejection inducing update relation $[\cdot]_s^-$ to allow for inquisitive negation (inspired by [1],[9]). So for instance we shall say:

$$(\mathcal{A}) \quad \begin{array}{l} \sigma[p]_s^+\tau \text{ iff } \tau = \sigma \cap [\![p]\!] \\ \sigma[p]_s^-\tau \text{ iff } \tau = \sigma\backslash[\![p]\!] \end{array} \qquad (\neg) \quad \begin{array}{l} \sigma[\neg\phi]_s^+\tau \text{ iff } \sigma[\phi]_s^-\tau \\ \sigma[\neg\phi]_s^-\tau \text{ iff } \sigma[\phi]_s^+\tau \end{array}$$

A positive update with a sentential atom p eliminates from an alternative all possible worlds at which p is false while a negative update with p eliminates all possible worlds at which p is true. A positive update with $\ulcorner\neg\phi\urcorner$ is just a negative update with ϕ and a negative update with $\ulcorner\neg\phi\urcorner$ is just a positive update with ϕ.

   The basic idea about possibility modals is that they test whether their prejacent relates the information carried by a state to a contradiction $\bot$ or to a consistent proposition $\not\leq$:

$$(\diamond) \quad \begin{array}{l} \sigma[\diamond\phi]_s^+\tau \text{ iff } \tau = \{w \in \sigma \colon \langle \texttt{info}(s), \bot\rangle \notin [\phi]_s^+\} \\ \sigma[\diamond\phi]_s^-\tau \text{ iff } \tau = \{w \in \sigma \colon \langle \texttt{info}(s), \not\leq\rangle \notin [\phi]_s^+\} \end{array}$$

For an alternative in a state s to pass the test imposed by a *positive* update with $\ulcorner\diamond\phi\urcorner$, the information carried by s must not be related to the *inconsistent* proposition via a positive update with ϕ. For an alternative in a state s to pass the test imposed by a *negative* update with $\ulcorner\diamond\phi\urcorner$—that is, a positive update with $\ulcorner\square\neg\phi\urcorner$—the information carried by s must not be related to a *consistent* proposition via a positive update with ϕ.

   So far we only have a rewrite of classical Update Semantics but the present setup allows us to combine a test semantics for modals with an inquisitive analysis of disjunction and negated conjunction. Start by coupling each update rule for alternatives with a corresponding update procedure for states:

Definition: Updates on States.   Define two update operations $\uparrow, \downarrow \colon (\mathcal{L} \times S) \mapsto S$:

1. $s \uparrow \phi = \{\tau \neq \bot \colon \exists \sigma \in s. \, \sigma[\phi]_s^+\tau\}$
2. $s \downarrow \phi = \{\tau \neq \bot \colon \exists \sigma \in s. \, \sigma[\phi]_s^-\tau\}$

A positive/negative update of some state s with ϕ delivers all the alternatives that are positively/negatively related to some element of s via ϕ.

   The proposal for disjunction is then the following one:

$$(\vee) \quad \begin{array}{l} \sigma[\phi \vee \psi]_s^+ \tau \text{ iff } \sigma[\phi]_s^+ \tau \text{ \underline{or} } \sigma[\psi]_{s\downarrow\phi}^+ \tau \\ \sigma[\phi \vee \psi]_s^- \tau \text{ iff } \exists \nu : \sigma[\phi]_s^- \nu \text{ \underline{and} } \nu[\psi]_{s\downarrow\phi}^- \tau \end{array}$$

This analysis captures two important intuitions about disjunctions: first, in addition to ruling out certain possibilities they raise each of their disjuncts as an issue in discourse. We capture this by letting a disjunction relate an input alternative to two potentially distinct alternatives: the result of updating with the first and the result of updating with the second disjunct. Moreover, in a sentence such as "Mary is in Chicago or she must be in New York" the modal in the second disjunct naturally receives a modally subordinated interpretation: it is interpreted under the supposition that Mary is not in Chicago. We achieve this result by saying that whenever a disjunction is processed in light of some state s, its second disjunct is processed in light of a negative update of s with the first disjunct.

Given some state s, a positive update with a conjunction $\ulcorner\phi \wedge \psi\urcorner$ proceeds via a positive update with ϕ light of s and then via a positive update with ψ in light of $s \uparrow \phi$:

$$(\wedge) \quad \begin{array}{l} \sigma[\phi \wedge \psi]_s^+ \tau \text{ iff } \exists \nu : \sigma[\phi]_s^+ \nu \text{ \underline{and} } \nu[\psi]_{s\uparrow\phi}^+ \tau \\ \sigma[\phi \wedge \psi]_s^- \tau \text{ iff } \sigma[\phi]_s^- \tau \text{ \underline{or} } \sigma[\psi]_{s\uparrow\phi}^- \tau \end{array}$$

The rules for negative updates with disjunctions and conjunctions enforce the validity of De Morgan's Laws.

Let me highlight some predictions before moving on to conditionals. As a preparation, define the notions of support, entailment, and consistency in the familiar dynamic fashion:

Definition: Support, Entailment, Consistency. Take any $s \in S$ and formulas of $\mathcal{L}$:

1. s supports ϕ, $s \Vdash \phi$, iff $s \uparrow \phi = s$
2. $\phi_1, \ldots, \phi_n$ *entails* ψ, $\phi_1, \ldots, \phi_n \vDash \psi$, iff for all $s \in S$, $s \uparrow \phi_1 \ldots \uparrow \phi_n \Vdash \psi$
3. $\phi_1, \ldots \phi_n$ is *consistent* iff for some $s \in S$: $s \uparrow \phi_1 \ldots \uparrow \phi_n \neq \varnothing$

A state supports ϕ just in case a positive update of s with ϕ idles. Entailment is just guaranteed preservation of support and the consistency of a sequence requires that a positive update with it sometimes results in a non-absurd state. It would, of course, be possible to define the notions of entailment and consistency on the basis of $\downarrow$ but I set an exploration of this interesting avenue aside for now.

Disjunctions embedded under a possibility operator then exhibit the free choice effect:

Fact 1. $\Diamond(p \vee q) \vDash \Diamond p \wedge \Diamond q$

The underlying observation here is that $s \uparrow \Diamond(p \vee q) \neq \varnothing$ only if $\langle \mathtt{info}(s), \bot \rangle \notin [p \vee q]_s^+$. But suppose that $\mathtt{info}(s)$ fails to contain both p- and q-worlds: then $[p]_s^+$ or $[q]_s^+$ *does* relate $\mathtt{info}(s)$ to $\bot$, hence $\mathtt{info}(s)[p \vee q]_s^+ \bot$ and thus $\langle \mathtt{info}(s), \bot \rangle \in [p \vee q]_s^+$ after all. So if $s \uparrow \Diamond(p \vee q) \neq \varnothing$ then $s \uparrow \Diamond p = s$ and $s \uparrow \Diamond q = s$.

Note that $\Diamond(p \vee q) \not\vDash \Diamond(p \wedge q)$ since passing the test conditions under consideration does not require the presence of a $p \wedge q$-world in $\mathtt{info}(s)$. Furthermore, it is easy to see that the free choice effect also arises if $\Diamond$ scopes over a negated conjunction since for all choices of $s \in S$ we have $[\neg(\neg\phi \wedge \neg\psi)]_s^+ = [\phi \vee \psi]_s^+$ by design.

We also account for the observation (see [1] and references therein) that embedding a disjunctive possibility under negation reverts disjunction to its classical behavior:

Fact 2. $\neg\Diamond(p \vee q) \vDash \neg\Diamond p \wedge \neg\Diamond q$

Observe that $s \uparrow \neg\Diamond(p \vee q) \neq \varnothing$ only if $\langle \mathtt{info}(s), \not\leq \rangle \notin [p \vee q]_s^+$. But suppose that $\mathtt{info}(s)$ contains a p- or a q-world: then $[p]_s^+$ or $[q]_s^+$ *does* relate $\mathtt{info}(s)$ to $\not\leq$, hence $\mathtt{info}(s)[p \vee q]_s^+ \not\leq$ and thus $\langle \mathtt{info}(s), \not\leq \rangle \in [p \vee q]_s^+$ after all. So if $s \uparrow \neg\Diamond(p \vee q) \neq \varnothing$ then $s \uparrow \neg\Diamond p = s$ and $s \uparrow \neg\Diamond q = s$.

We may also observe that the framework developed here preserves key insights from the dynamic analysis of modals, including the internal dynamics of conjunction:

Fact 3. $\neg p \wedge \Diamond p$ is inconsistent

Here it pays off that updates are defined relative to a shifty state parameter s. Clearly $\mathtt{info}(s \uparrow \neg p)$ does not contain any p-worlds and so any update with with '$\Diamond p$' in light of $s \uparrow \neg p$ is guaranteed to result in the absurd state.

Finally, let me just state some observations about the material conditional and the necessity operator that are of relevance for the upcoming discussion:

Fact 4. For all $s \in S$: $[\Box(\phi \supset \psi)]_s^+ = [\neg \Diamond(\phi \wedge \neg \psi)]_s^+$ and $[\Box(\phi \supset \psi)]_s^- = [\Diamond(\phi \wedge \neg \psi)]_s^+$

These identities follow immediately if we treat '$\Diamond$' and '$\Box$' as duals and adopt the standard analysis of the material conditional in terms of conjunction and negation.

In sum, the proposal developed so far combines a test semantics for modals with an inquisitive approach to disjunction and negated conjunction in a way that captures the scope as well as the limits of the free choice effect. Let me now turn to counterfactuals.

2.2 Counterfactuals

A *would*-counterfactual is a strict material conditional presupposing that its antecedent is possible. Following standard protocol I treat *might*- and *would*-counterfactuals as duals and presuppositions as definedness conditions on updating ([10],[3]):

$$(\Box\!\!\rightarrow) \quad \begin{array}{l} \sigma[\phi \Box\!\!\rightarrow \psi]_s^+ \tau \text{ iff } \sigma[\Diamond\phi]_s^+ \sigma \ \underline{\text{and}} \ \sigma[\Box(\phi \supset \psi)]_s^+ \tau \\ \sigma[\phi \Box\!\!\rightarrow \psi]_s^- \tau \text{ iff } \sigma[\Diamond\phi]_s^+ \sigma \ \underline{\text{and}} \ \sigma[\Box(\phi \supset \psi)]_s^- \tau \end{array}$$

Given some state s, a positive or negative update with $\ulcorner \phi \Box\!\!\rightarrow \psi \urcorner$ fails to relate an input alternative σ to any output in case the information carried by s is incompatible with ϕ (the presupposition thus projects out of negation). Assuming that the presupposition is satisfied, a positive update with $\ulcorner \phi \Box\!\!\rightarrow \psi \urcorner$ then tests whether s supports $\ulcorner \phi \supset \psi \urcorner$ while a negative update effectively asks whether $\ulcorner \phi \wedge \neg \psi \urcorner$ is compatible with s.[2]

For convenience, let me state explicitly the update rules for *might*-counterfactuals:

$$(\Diamond\!\!\rightarrow) \quad \begin{array}{l} \sigma[\phi \Diamond\!\!\rightarrow \psi]_s^+ \tau \text{ iff } \sigma[\Diamond\phi]_s^+ \sigma \ \underline{\text{and}} \ \sigma[\Diamond(\phi \wedge \psi)]_s^+ \tau \\ \sigma[\phi \Diamond\!\!\rightarrow \psi]_s^- \tau \text{ iff } \sigma[\Diamond\phi]_s^+ \sigma \ \underline{\text{and}} \ \sigma[\Diamond(\phi \wedge \psi)]_s^- \tau \end{array}$$

These update rules are an immediate consequence of treating '$\Box\!\!\rightarrow$' and '$\Diamond\!\!\rightarrow$' as duals.

It is of course uncontroversial that this analysis predicts that counterfactuals simplify if their antecedents involve a disjunction or a negated conjunction:

Fact 5. $(p \vee q) \Box\!\!\rightarrow r \models p \Box\!\!\rightarrow r, q \Box\!\!\rightarrow r$ and $\neg(p \wedge q) \Box\!\!\rightarrow r \models \neg p \Box\!\!\rightarrow r, \neg q \Box\!\!\rightarrow r$

This is an immediate consequence of analyzing *would*-counterfactuals as strict material conditionals. The claim that counterfactuals presuppose the possibility of their antecedents, however, immediately predicts that $s \uparrow (p \vee q) \Box\!\!\rightarrow r = \varnothing$ unless $\mathtt{info}(s)$ includes p- as well as q-worlds due to the free choice effect. I will come back to this fact momentarily, but we can already at this stage observe the following fact about *might*-counterfactuals:

---

[2]Bringing presuppositions into the picture also raises the question of how they project and a proper answer requires minor modifications to some of our original update rules. For instance, in order to predict that presuppositions project out of the possibility operator one would need to say that $\sigma[\Diamond\phi]_s^+ \tau$ holds just in case $\tau = \{w \in \sigma : \langle \mathtt{info}(s), \bot \rangle \notin [\phi]_s^+\}$ and, moreover, $\exists \nu. \sigma[\phi]_s^+ \nu$. Likewise for the negative entry: $\sigma[\Diamond\phi]_s^- \tau$ holds just in case $\tau = \{w \in \sigma : \langle \mathtt{info}(s), \not\bot \rangle \notin [\phi]_s^+\}$ and, moreover, $\exists \nu. \sigma[\phi]_s^- \nu$. I set these additional complexities, which would also affect the update rules to disjunction, aside to streamline the notation and since getting all the facts about presupposition projection right goes beyond the scope of this investigation.

Fact 6. $p \mathbin{\Diamond\!\!\!\rightarrow} (q \vee r) \models p \mathbin{\Diamond\!\!\!\rightarrow} q, p \mathbin{\Diamond\!\!\!\rightarrow} r$

Clearly a state s supports '$p \mathbin{\Diamond\!\!\!\rightarrow} (q \vee r)$' just in case it supports '$\Diamond p$' and '$\Diamond(q \vee r)$' and hence—due to the free choice effect—both '$\Diamond q$' and '$\Diamond r$'. So we predict that *might*-counterfactuals such as (3) simplify in the way the do. In contrast *would*-counterfactuals with disjunctive consequents rightly fail to simplify: a nonempty state supporting '$p \wedge (q \wedge \neg r)$', for instance, supports '$p \mathbin{\Box\!\!\!\rightarrow} (q \vee r)$' without supporting '$(p \mathbin{\Box\!\!\!\rightarrow} r)$'.

This is all good news but there remain some open problems. Fine and Warmbrōd worry that SDA entails the validity of Antecedent Strengthening (AS) assuming substitution of logical equivalents in conditional antecedents (see [5],[22]). While AS fails to be valid here since a state may support the possibility presupposition carried by $\ulcorner \phi \mathbin{\Box\!\!\!\rightarrow} \chi \urcorner$ without supporting the one carried by $\ulcorner (\phi \wedge \psi) \mathbin{\Box\!\!\!\rightarrow} \chi \urcorner$, the framework developed so far fails to leave room for the consistency of Sobel sequences such as (4):

(4)   If Mary had come to the party, it would have been fun. But if Bert had come too, it would not have been fun.

No consistent state compatible with Alice and Bert coming to the party can support the asserted contents of both counterfactuals, and every state that is incompatible with that possibility inevitably fails to satisfy the presuppositions carried by the first or the second member of the sequence. In any case updating with the sequence in (4) is guaranteed to result in the absurd state and thus counts as inconsistent, which is not a good result.

The second worry is that there are familiar cases in which simplification of disjunctive antecedents seems to fail. McKay and van Inwagen ([12]) consider the following case:

(5)   If Spain had fought for the Axis or the Allies, she would have fought for the Axis.
   a. $\rightsquigarrow$ If Spain had fought for the Axis, she would have fought for the Axis.
   b. ??? If Spain had fought for the Allies, she would have fought for the Axis.

(5a) is of course trivial but (5b) is objectionable, contrary to what SDA predicts.

Both problems can be solved by adding a few complexities to our basic story about presupposition and assertion. Presuppositions in general and possibility presuppositions in particular are normally *accommodated* as discourse proceeds, allowing counterfactual domains of quantification to evolve dynamically—this is what underlies the consistency of Sobel sequences. Assertions sometimes conflict with other bits of information taken for granted in discourse and thus may *pragmatically* trigger modifications of the input context so that the proposed update may be processed—this is what explains why SDA is semantically valid but occasionally defeated by intervening pragmatic factors.

3   Hyperstates

The twist to the basic story is the idea that modals and counterfactuals are quantifiers over a *minimal* and *dynamically evolving* domain of quantification. To make sense of this idea we let context determine a slightly more complex background for processing counterfactuals: instead of thinking of a context as providing a single state, we will think of it as providing a *set* of (nonempty) states. Intuitively, the information carried by each state can be understood as a domain of quantification, and counterfactuals then pertain to whatever states come with the strongest informational content.

Definition: Hyperstates.   A *hyperstate* $\pi \subseteq \mathcal{P}(S) \backslash \varnothing$ is any set of nonempty states. We say that $s' \leq_\pi s$, s' is *at least as strong as* s in π, iff $s, s' \in \pi$ and $\texttt{info}(s') \subseteq \texttt{info}(s)$. Π is the set of all hyperstates. We refer to $\varnothing$ as the *absurd* hyperstate and treat $\pi_0 = \mathcal{P}(S) \backslash \varnothing$ as the *initial* hyperstate.

Thinking of contexts as hyperstates requires some modifications to our update system. Fortunately, our update procedures for alternatives can stay the same. But states are now updated in light of a hyperstate and updates with modals now pertain to those elements of a hyperstate whose informational content is strongest. We achieve this by slightly modifying the update functions for states in the following manner:

Definition: Hyper-updates on States Define update functions $\uparrow_\pi, \downarrow_\pi \colon \mathcal{L} \mapsto (S \mapsto S)$ as follows:

1. $s \uparrow_\pi \phi = \{\tau \neq \perp \colon \exists \sigma \in s \; \exists s' \preceq_\pi s. \; \sigma[\phi]^+_{s'} \tau\}$
2. $s \downarrow_\pi \phi = \{\tau \neq \perp \colon \exists \sigma \in s \; \exists s' \preceq_\pi s. \; \sigma[\phi]^-_{s'} \tau\}$

To see why these modifications matter, consider how a modal now interacts with some $s \in \pi$: an update of s with '$\Diamond p$' now tests whether there is some $s' \preceq_\pi s$ whose informational content $\mathtt{info}(s')$ includes a p-world. Clearly, this is so just in case $\mathtt{info}(s)$ includes a p-world as well, and so possibility modals work exactly as before. But the twist does matter when it comes to an update with '$\Box p$': this now tests whether there is some $s' \preceq_\pi s$ whose informational content $\mathtt{info}(s')$ exclusively consists of p-worlds, and that may be so even if $\mathtt{info}(s)$ itself includes a possible world at which p is false (though of course no state stronger than s' may contain a $\neg p$-world). So in this sense '$\Box$' becomes a strict quantifier over the informational content of the strongest members of a hyperstate. Updating with nonmodal formulas of $\mathcal{L}$ stays the same.

We now define what it takes for a context understood as selecting a hyperstate to accept and admit ϕ and define updates on hyperstates on that basis:

Definition: Acceptance, Admission, Updates on Hyperstates. Consider arbitrary $\pi \in \Pi$ and $\phi \in \mathcal{L}$:

1. π *accepts* ϕ, $\pi \Vdash \phi$, iff for all $s' \in \pi$ there exists some $s \preceq_\pi s' \colon s \uparrow_\pi \phi = s$
2. π *admits* ϕ, $\pi \rhd \phi$, iff $\pi \nVdash \neg\phi$
3. $\pi + \phi = \{s \uparrow_\pi \phi \colon s \in \pi \;\&\; \pi \rhd \phi\} \backslash \varnothing$

Acceptance of ϕ amounts to support by the strongest states in a hyperstate. An update with ϕ is admitted as long as its negation is not accepted. And finally, a hyperstate is updated with ϕ by updating each of its elements with ϕ and collecting the nonempty results, provided that an update with ϕ is admissible.

Entailment and consistency are again understood in the familiar dynamic fashion:

Definition: Entailment and Consistency (Hyperstates). Take any $\pi \in \Pi$ and formulas of $\mathcal{L}$:

1. $\phi_1, \ldots, \phi_n$ *entails* ψ, $\phi_1, \ldots, \phi_n \models \psi$, iff for all $\pi \in \Pi$, $\pi + \phi_n \ldots + \phi_n \Vdash \psi$
2. $\phi_1, \ldots \phi_n$ is *consistent* iff for some $\pi \in \Pi \colon \pi + \phi_1 \ldots + \phi_n \neq \varnothing$

This setup preserves everything said in §2 but in addition allows for a Sobel sequence like '$p \; \Box\!\!\rightarrow \; r$' followed by '$(p \wedge q) \; \Box\!\!\rightarrow \; \neg r$' to be consistent. To see why, assume that $w_1 \in [\![p \wedge \neg q \wedge r]\!]$ and that $w_2 \in [\![p \wedge q \wedge \neg r]\!]$, and let $\pi = \{s, s'\}$ be such that $\mathtt{info}(s) = \{w_1\}$ while $\mathtt{info}(s') = \{w_1, w_2\}$. Then clearly both s and s' satisfy the presupposition carried by the first counterfactual in the sequence and since $s \preceq_\pi s'$ and $s \uparrow_\pi p \; \Box\!\!\rightarrow \; r = s$, we have $\pi + p \; \Box\!\!\rightarrow \; r = \pi$. Notice furthermore that $\pi \nVdash \neg((p \wedge q) \; \Box\!\!\rightarrow \; r)$ since we have $s \uparrow_\pi \neg((p \wedge q) \; \Box\!\!\rightarrow \; r) \neq s$: here the underlying observation is that s is the strongest state in π but fails support the counterfactual's possibility presupposition '$\Diamond(p \wedge q)$'. So π admits an update with the second member of the Sobel sequence, resulting in a consistent hyperstate $\pi' = \{s'\}$, as desired. More complex Sobel sequences can be consistently processed in more complex hyperstates. I conclude that the validity of SDA is compatible with the fact that counterfactuals resist AS so that Sobel sequences are consistent.

It remains to comment on the fact that counterfactuals with disjunctive antecedents do not seem to simplify across the board. Here I maintain that SDA is semantically valid and that there is a principled story for why pragmatic factors sometimes intervene. Part of the picture is that the problematic counterfactual (repeated here for convenience) intuitively communicates that Spain would never have fought for Allies.

(5) If Spain had fought for the Axis or the Allies, she would have fought for the Axis.

In fact, explicitly acknowledging the possibility of Spain joining the Allies renders (5) unacceptable (see also [19]):

(6) Spain might have fought for the Allies. ???But if Spain had fought for the Axis or the Allies, she would have fought for the Axis.

Given this communicative effect it is not surprising that a context that has been strengthened with (5) fails to support 'If Spain had fought for the Allies, she would have fought for the Axis' since it fails to satisfy the counterfactual's possibility presupposition.

One may think that the previous observation immediately undermines the proposal that counterfactuals with disjunctive antecedents presuppose the possibility of each disjunct. But this is not so: an indicative conditional such as 'If John wins the competition, then I am the Flying Dutchman' communicates—in ordinary circumstances anyway—that John will not win the competition, but there is no serious doubt that indicative conditionals presuppose that their antecedent is a possibility in the common ground. What is needed is a general story about how conditional assertions may at times remove worlds at which their antecedents are true from the domain of quantification.

Here is how such a story might go. Consider the asserted content of (5) in a context in which the counterfactual's possibility presupposition has been accommodated: updating with that content would result in a context in which Spain might have fought on the Axis *and* the Allies side, but of course it is a background assumption that Spain would not have fought on both sides, and hence clear to all discourse participants that updating with the asserted content of (5) results in the absurd state. Yet speakers are in general cooperative and thus do not propose to add information to the common ground that is incompatible with what is taken for granted. In these lights, it makes good sense to say that a speaker may communicate ψ by asserting ϕ in case updating the discourse with ϕ results in the absurd state but updating that context with ψ and then with ϕ does not. More precisely:

Suppose that $\pi + \phi = \varnothing$ but $\pi + \psi + \phi \neq \varnothing$ and for all χ such that $\pi + \chi + \phi \neq \varnothing$, $\pi + \chi \Vdash \psi$. Then an utterance amounting to an assertion of ϕ in π by default pragmatically implies a proposal to update $\pi + \psi$ with ϕ.

Suppose then that $\pi \models \Diamond \texttt{Ax} \land \Diamond \texttt{Al}$ yet also $\pi \models \neg(\texttt{Ax} \land \texttt{Al})$, and consider $\pi + \Box((\texttt{Ax} \lor \texttt{Al}) \supset \texttt{Ax})$, that is, the result of updating π with the asserted content of (5). Clearly $\pi + \Box((\texttt{Ax} \lor \texttt{Al}) \supset \texttt{Al}) \Vdash \Diamond(\texttt{Ax} \land \texttt{Al})$ and hence the update results in the absurd state. But we can also observe that $\pi + \neg \texttt{Al}$ *is* consistent and *can* be consistently updated with the asserted content of (5) and so we predict—given the pragmatic twist just proposed— that the assertion that comes with an utterance of (5) implies that Spain would not have sided with the Allies. Since this is an implicature, we expect it to be cancellable, which is just what we have in (6). Furthermore, $\pi + \neg \texttt{Ax}$, while consistent, *cannot* be consistently updated with the asserted content of (5), and so we do not make the wrong prediction that there is an implicature that Spain would not have sided with the Axis.

I thus conclude that there is a principled pragmatic explanation for why certain counterfactuals resist simplification. The reason, in brief, is that certain counterfactuals imply information in addition to asserting a strict material conditional, and that this additional information may in fact eliminate possibilities that have been brought into view by presupposed content. While this account taps into pragmatic resources to account for the problematic data, the needed assumption appears to be fairly modest and well-motivated.

4 Conclusion

A dynamic strict analysis that exploits insights from inquisitive semantics predicts why counterfactuals simplify in the way they do. The explanation is compatible with the consistency of Sobel sequences and with the fact that certain counterfactuals resist simplification for principled pragmatic reasons.

Let me conclude the discussion by pointing to two remaining tasks left for another day. First, simplification is not a feature particular to counterfactual conditionals: indicative conditionals are just as amenable to simplification as are their counterfactuals cousins. What is needed then is a story about how the framework developed here can be generalized so that it covers other conditional constructions as well. While this is not a trivial task, the basic idea is clear: all conditionals are strict and come with a possibility presupposition; their differences amount to differences in the domain of quantification.

Second, the attempt to combine a dynamic treatment of modals with an inquisitive treatment of disjunction brings to mind the question, discussed by [4] and [15], of how to distinguish between informative, inquisitive, and attentive content. Here I want to briefly observe that the notion of a hyperstate is fine-grained enough to keep track of various kinds of discourse information. First, there is the *informational content* of a hyperstate π, understood as the possible worlds compatible with what is taken for granted: $\mathtt{Info}(\pi) = \bigcup\{\mathtt{info}(s) : s \in \pi\}$. Second, we may associate with π an *issue* understood as the set of its maximal alternatives: $\mathtt{Issue}(\pi) = \{\sigma : \sigma \in \mathtt{Alt}(\pi) \ \& \ \neg\exists\tau \in \mathtt{Alt}(\pi). \sigma \subset \tau\}$, where $\mathtt{Alt}(\pi) = \bigcup\{s : s \in \pi\}$. Looking at the initial hyperstate π_0 we can then say that an atomic sentence p has $[\![p]\!]$ as its *informational* content in the sense that the informational content of $\pi_0 + p$ is just $[\![p]\!]$. For parallel reasons we can say that '$p \vee q$' has $[\![p]\!] \cup [\![q]\!]$ as its informational content and $\{[\![p]\!], [\![q]\!]\}$ as its *inquisitive* content since $[\![p]\!]$ and $[\![q]\!]$ are in the issue of $\pi_0 + p \vee q$. The distinction between the informativeness of a formula and its inquisitiveness is thus analyzed in terms of the potential to eliminate possibilities and to raise issues in discourse. And this strategy also allows us to look at hyperstates in a way that identifies yet another kind of content. Let me explain.

Earlier I said that each element of a hyperstate is a potential domain of quantification. If we think of a domain of quantification as a candidate for the region of logical space that is *relevant* for the modal discourse under consideration, it makes sense to think of a hyperstate π as identifying what I have called elsewhere a set of *serious* or *live possibilities* (see [23]), that is, the possibilities compatible with every state in π: $\mathtt{Live}(\pi) = \{\sigma : \forall s \in \pi \ \exists w \in \mathtt{info}(s). w \in \sigma\}$. We can then say that a sentence has *attentive* content in virtue of its potential to bring hitherto ignored possibilities into view: '$\Diamond p$', for instance, has $\{[\![p]\!]\}$ as its attentive content since $\pi_0 + \Diamond p$ treats $[\![p]\!]$ as a live possibility, and for parallel reasons '$\Diamond(p \vee q)$' has $\{[\![p]\!], [\![q]\!]\}$ as its attentive content.

A comprehensive discussion would explore in more detail what predictions the setup sketched here makes about the interaction between informational, inquisitive, and attentive content, and how it predictions differ from those of other frameworks. For now, let me just conclude that the framework is of general semantic interest beyond its capacity to handle some empirical challenges pertaining to inferences licensed by counterfactual conditionals: it combines a very attractive dynamic semantic treatment of possibility modals as highlighting the significance of certain possibilities in discourse with a—no less attractive— inquisitive treatment of disjunction as refining issues in discourse. The fact that combining these treatments allows us to make substantial progress toward a better understanding of the free choice effect gives us all the more reason to think that the dynamic inquisitive story told here deserves further exploration.

References

[1] Martin Aher. Free choice in deontic inquistive semantics (dis). *Lecture Notes in Computer Science*, 7218:22–31, 2012.

[2] Luis Alonso-Ovalle. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, 32(2):207–244, 2009.

[3] David I. Beaver. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications, Stanford, 2001.

[4] Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. Attention! *Might* in inquisitive semantics. *Proceedings of SALT XIX*, pages 91–108, 2009.

[5] Kit Fine. Critical notice. *Mind*, 84(335):451–458, 1975.

[6] Kai von Fintel. Counterfactuals in a dynamic context. In Michael Kenstowicz, editor, *Ken Hale: A Life in Language*, pages 123–152. MIT Press, Cambridge, MA, 2001.

[7] Michael Franke. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, 4(1):1–82, 2011.

[8] Anthony S. Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360, 2007.

[9] Jeroen Groenendijk and Floris Roelofsen. Towards a suppositional inquisitive semantics, 2015. Manuscript, University of Amsterdam.

[10] Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst, 1982.

[11] David K. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, MA, 1973.

[12] Thomas McKay and Peter van Inwagen. Counterfactuals with disjunctive antecedents. *Philosophical Studies*, 31(5):353–356, 1977.

[13] Donald E. Nute. Counterfactuals and the similarity of worlds. *Journal of Philosophy*, 72(21):773–778, 1975.

[14] Donald E. Nute. *Topics in Conditional Logic*. Reidel, Dordrecht, 1980.

[15] Floris Roelofsen. A bare bone attentive semantics for *Might*. In Maria Aloni, Michael Franke, and Floris Roelofsen, editors, *The Dynamic, Inquisitive, and Visionary Life of ϕ, $?\phi$, and $\diamondsuit\phi$*, pages 190–215. ILLC Publications, Amsterdam, 2013.

[16] Robert van Rooij. Free choice counterfactual donkeys. *Journal of Semantics*, 23(4):383–402, 2006.

[17] Robert van Rooij. Conjunctive interpretation of disjunction. *Semantics and Pragmatics*, 3(11):1–28, 2010.

[18] Robert C. Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory*, pages 98–112. Blackwell, Oxford, 1968.

[19] William B. Starr. A uniform theory of conditionals. *Journal of Philosophical Logic*, 43(6):1019–1064, 2014.

[20] Anna Szabolsci and Bill Haddican. Conjunction meets negation: A study in cross-linguistic variation. *Journal of Semantics*, 21(3):219–249, 2004.

[21] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

[22] Ken Warmbrōd. Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic*, 10(2):267–289, 1981.

[23] Malte Willer. Dynamics of epistemic modality. *Philosophical Review*, 122(1):45–92, 2013.

[24] Malte Willer. Indicative scorekeeping. *Proceedings of the 19th Amsterdam Colloquium*, pages 249–256, 2013.

Mention-some readings of questions:
Complexities from number marking [*]

Yimei Xiang

Harvard University, Massachusetts, USA
`xiang.yimei@gmail.com`

Abstract

Despite the well-known fact that $\diamond$-questions systematically accept mention-some answers, number marking on *wh*-complements adds two constraints to the distributional pattern of mention-some. First, a singular question admits a unique true answer and hence disallows mention-some. Second, a plural question disallows a mention-some answer that names only a singularity. I argue that the former constraint comes from a closure requirement on short answers, and the latter from the anti-presupposition of the plural morpheme.

1 Introduction

Generally speaking, good answers are exhaustive answers. For example, to properly answer (1), one needs to specify all the actual attendants to the party, as in (1a). We call exhaustive answers like (1a) "mention-all (MA) answers". If an addressee can name only some of the attendants, then to be cooperative, he would ignorance-mark his answer. For instance, he could state the ignorance inference explicitly as in (1b), or mark his answer with a prosodic rise-fall-rise couture as in (1c). If a non-exhaustive answer is not properly marked, it would be interpreted exhaustively.

(1) Who came the party?
    (*w*: only John and Mary came to the party.)

    a. John and Mary did.

    b. John did. I'm not sure who else did.

    c. JOHN did.
       L H* L-H%

Contrary to (1), as firstly observed by Groenendijk & Stokhof (1984), *wh*-questions with an existential modal, called "$\diamond$-questions" henceforth, also admit non-exhaustive answers. For instance, one can naturally answer the $\diamond$-question (2) by specifying one or all of the qualified candidates. The MA answer (2b) can be either a conjunction or a disjunction; in the disjunctive form, the MA answer takes a free choice interpretation. Crucially, the non-exhaustive answer (2a), unlike (1b-c), does not need to carry any ignorance-mark. To highlight this difference, we call (2a) "mention-some (MS) answers", while (1b-c) "partial answers".

(2) Who can chair the committee?
    (*w*: the committee can and can only be chaired by John or Mary.)

    a. John.   [MS answer]

    b. John and/or Mary. [MA answer]

Moreover, number-marking on *wh*-complements adds two more pieces of complexities to the distributional pattern of MS. First, a singular question, namely a *wh*-question where the *wh*-complement is marked as singular, can have only one true answer. For instance, (3b) is incoherent because the second clause contradicts the uniqueness inference that *only one of the professors can chair*.

(3)   a.   Who can chair the committee? $\sqrt{}$ I know that several professors can.

       b.   Which professor can chair the committee? # I know that several professors can.

       c.   Which professors can chair the committee? $\sqrt{}$ I know that several professors can.

Second, a plural $\diamond$-question rejects an MS answer that names only an atomic individual. Compare (4) and (5). If the committee should have one chair but multiple members, an MS answer of (4) and (5), if it is available, names an atomic individual (e.g. *John*) and a sum (e.g. *John+Mary*), respectively. With this difference, while (5) admits both MS and MA answers, (4) accepts only the MA answer (4b) and requires the non-exhaustive answer (4a) to be ignorance-marked. The contrast between (4-5) is more salient in embedding contexts: (6a) requires John to know the MA answer of (4); while (6b) only requires John to know an MS answer of of (5).

(4)   Which professors can chair the committee?
      (*w*: the committee can and can only be chaired by either John or Mary.)

     a.   John. # (I'm not sure who else can.)

     b.   John and/or Mary.

(5)   Which professors can form the committee?
      (*w*: the committee can only be formed by any two professors among John, Mary, and Sue.)

     a.   John+Mary.

     b.   John+Mary, John+Sue, and/or Mary+Sue.

(6)   a.   John knows which professors can chair the committee.

       b.   John knows which professors can form the committee.

2   Previous Studies

2.1   Dayal (1996)

Uniqueness: the strongest true answer exists   Dayal (1996) adopts the Hamblin-Karttunen semantics of questions and defines a presuppositional Ans_D-operator to derive good answers from the Hamblin denotation: $\mathrm{Ans}_D(Q)(w)$ returns the unique strongest true answer of Q in w and presupposes its existence. The strongest true answer is the true answer that entails all the true answers.

(7) $\mathrm{Ans}_D(Q)(w) = \exists p[w \in p \in Q \land \forall q[w \in q \in Q \rightarrow p \subseteq q]].$
       $\iota p[w \in p \in Q \land \forall q[w \in q \in Q \rightarrow p \subseteq q]]$

The presupposition of Ans_D captures the uniqueness requirement of singular questions: in a singular question, the presupposition of Ans_D is satisfied iff this question has a unique true answer which names a singularity. First, following Link (1983), Dayal assumes that a singular NP denotes a set of atomics, while a plural NP ranges over both atomic and sum domains. For instance, with two professors j and m taken into considerations, we have *professor′* $= \{j, m\}$ and *professors′* $=$ **professor′* $= \{j, m, j \oplus m\}$. Thus, the Hamblin set denoted by the plural question (8a) includes plural propositions, while the one denoted by the singular question (8b) does not. For simplicity, I use Q_w to represent the full set of true answers of Q in w: $Q_w = \{p : w \in p \in Q\}$. In a world where multiple professors came to the party, (8a)

has a strongest true answer, namely the plural answer, but (8b) does not.Thus, finally, exercising Ans$_D$ in (8b) gives rise to a presupposition failure.

(8) (*w*: among the professors, only John and Mary came to the party.)

a. Which professors came to the party?

$Q = \{came'(x) : x \in *professor'\}$

$Q_w = \{came'(j), came'(m), came'(j \oplus m)\}$

Ans$_D(Q)(w) = came'(j \oplus m)$

b. Which professor came to the party?

$Q = \{came'(x) : x \in professor'\}$

$Q_w = \{came'(j), came'(m)\}$

Ans$_D(Q)(w)$ is undefined

To sum up, Dayal's account of uniqueness predicts that a question is always defined iff its answer space is closed under conjunction, which ensures the existence of the strongest true answer.[1]

MS answers are partial answers   Dayal's (1996) analysis predicts that a question primarily has only one good answer, namely the strongest true answer. But an MS answer cannot be the strongest, why is it acceptable in $\diamond$-questions? Dayal (in prep: chapter 3) attributes the acceptability of MS to pragmatic factors and assumes that MS answers are special partial answers that are sufficient for the goal behind the question.[2] This idea follows the lines of Groenendijk & Stokhof (1984), van Rooij (2004), van Rooij & Schulz (2006), and among the others: a non-exhaustive answer does not have to carry an ignorance-mark as long as it suffices for the conversational goal of the question. For example, consider the question *where can I get gas?*. If the goal for the questioner is just to find a place to get gas, the addressee only needs to name one accessible gas station; in contrast, if the goal is to investigate the gas market in the considered area, the addressee needs to list out all the gas stations in the considered area. For simplicity, let us call the former goal a "mention-one" goal, while the latter a "mention-all" goal.

I agree that pragmatics plays a role in distributing MS; for instance, if a question is semantically ambiguous between MS and MA, a goal that calls for an exhaustive answer can block MS. But, I doubt that MS is derived pragmatically from the start. This claim has already been reached by George (2011) and Fox (2013). I provide two more empirical arguments against the pragmatic account of MS.

First, intermediate answers, which are more informative than MS answers, are partial answers and must be ignorance-marked. For instance, under a "mention-one" goal, the intermediate answer (9b), which names more than one but not all of the candidates, should be sufficient for the goal; nevertheless, (9b) must to be ignorance-marked, otherwise would be interpreted exhaustively.

(9) Who can chair the committee?
    (*w*: the committee chair can and can only be either John, Mary, or Sue.)

   a. John. (I'm not sure who else can.)
   b. John and/or Mary. # (I'm not sure who else can.)
   c. John, Mary, and/or Sue.

More generally, the obligatory ignorance-mark on intermediate answers suggests that whether an answer admits a non-exhaustive reading is primarily determined by its grammatical structure: an unmarked individual answer can be non-exhaustive; while an unmarked disjunctive or conjunctive answer cannot.

---

[1]Note that closing the quantificational domain of the *wh*-item under sum does not guarantee the existence of the strongest true answer. In (1), *ab* and *cd* each formed a team does not entail *abcd* together formed one. To avoid overly predicting a presupposition failure, Dayal should include the conjunctive proposition $form'(a \oplus b) \wedge form'(c \oplus d)$ as a possible answer.

(1) Who formed a team?   # $Q_w : \{form'(a \oplus b), form'(c \oplus d)\}$

[2]Dayal (in prep: chapter 3) adds the following condition to capture the distributional pattern of MS in plural questions: an MS answer should be able to be used as an MA answer of the given question in a compatible world. She claims that there is no world where a proposition naming a singularity could be used as an MA answer of a plural question. Accordingly, (4a) is not qualified to be an MS answer because it cannot be an MA answer of (4).

First, in embedding contexts, good answers are always mention-one or mention-all, as in (10a) and (10b), respectively. But a conversational goal can be naming any amount of chair candidates. For instance, the following scenario has a "mention-three" goal: the dean wants to make plans for a committee and wants to meet three people who can chair this committee. In this scenario, a pragmatic account incorrectly predicts (10) to take the odd reading (10c): (10) is true iff John knows three or more candidates, and false if John knows less than three candidates. A semantic account, on the other hand, can easily handle the unacceptability of (10c): good answers generated from logical forms are either mention-one or mention-all, not intermediate.

(10)   John knows who can chair the committee.

    a. For some individual x who can chair, John knows that x can chair.   √

    b. For every individual x who can chair, John knows that x can chair.   √

    c. For some three individuals xyz who each can chair, John knows that xyz each can chair.   ×

2.2 Fox (2013)

MS answers are maximally informative true answers   To capture MS grammatically, Fox (2013) proposes a weaker Ans_F-operator: $\text{Ans}_F(Q)(w)$ returns the set of *maximally informative* (MaxI) true answers of Q in w, each of which is a good answer. A true answer is MaxI iff it is not asymmetrically entailed by any true answers. A cross-categorical definition for MaxI is given in (12).

(11) $\text{Ans}_F(Q)(w) = \{p : w \in p \in Q \wedge \forall q[w \in q \in Q \rightarrow q \not\subset p]\} = \text{MaxI}(Q_w)$ (First version)

(12) $\text{MaxI}(\alpha_{\langle \tau, t \rangle}) = \{a_\tau : a \in \alpha \wedge \forall b[b \in \alpha \rightarrow b \not\subset a]\}$

Next, Fox proposes that a question admits MS iff the output set of exercising Ans_F can be non-singleton. Compare (13-14). Underlining highlights their MaxI true answers. The basic *wh*-question (13) can have only one MaxI true answer; while the ◇-question (14) can have multiple ones, which are all MS answers.

(13)   Who came to the party last night?
    (*w*: only John and Mary came to the party yesterday.)

    a. $Q_w = \{came'(j), came'(m), \underline{came'(j \oplus m)}\}$

    b. $\text{Ans}_F(Q)(w) = \{came'(j \oplus m)\}$

(14)   Who can chair the committee?
    (*w*: the committee can and can only be chaired by either John or Mary.)

    a. $Q_w = \{\underline{\Diamond chair'(j)}, \underline{\Diamond chair'(m)}\}$

    b. $\text{Ans}_F(Q)(w) = \{\Diamond chair'(j), \Diamond chair'(m)\}$

Compared to Dayal (1996), Fox's analysis predicts that a non-exhaustive answer can be a good answer, and that a question can have multiple good answers.

Uniqueness: the IE-exhaustification of some answer is true   The Ans_F-operator defined in (11), however, cannot capture the uniqueness requirement of singular questions. For instance, both the true answers in (15a) are MaxI. Thus, (11) incorrectly predicts that a singular question is an MS question.

(15)   Which professor came to the party last night?
    (*w*: among the professors, only John and Mary came to the party yesterday.)

    a. $Q_w = \{\underline{came'(j)}, \underline{came'(m)}\}$

    b. $\text{Ans}_F(Q)(w) = \{came'(j), came'(m)\}$ Problem!

Noticing this problem, Fox (2013) adds two more assumptions to his initial proposal. First, based on Spector's (2007, 2008) observations on □-questions, Fox adds higher-order disjunctive and conjunctive answers to the answer spaces of number-neutral and plural *wh*-questions. Spector (2007, 2008) observes that an elided disjunctive answer can completely answer a □-question. For instance in (16b), the disjunction can be interpreted as scoping below the universal modal, which yields a free choice interpretation that *John can read Syntax or MP, and he has to read one of them.*

(16)   a. What does John have to read?

      b. *Syntax* or *MP*.   ($^{\text{OK}}$*or > have to*; $^{\text{OK}}$ *have to > or*)

(17)   a. Which books does John have to read?

      b. The French books or the English books.   ($^{\text{OK}}$*or > have to*; $^{\text{OK}}$ *have to > or*)

To capture this reading, Spector proposes that a *wh*-question is semantically ambiguous between an individual reading and a higher-order reading. Under the latter reading, the *wh*-item lives on a set of upward monotone generalized quantifiers like $s \vee m$ (i.e. an existential generalized quantifier over the set $\{s, m\}$), producing answers like $\Box read'(j, s \vee m)$. Moreover, contrary to the cases in (16) and (17), Fox observes that a disjunctive answer of a singular □-question can only take an ignorance reading. For instance, (18b) can only mean *John either has to read Syntax or has to read MP.*

(18)   a. Which book does John have to read?

      b. *Syntax* or *MP*.   ($^{\text{OK}}$*or > have to*; # *have to > or*)

Given the contrast between (16-17) and (18), Fox assumes that singular *wh*-phrases live on a set of atomic individuals, while bare *wh*-words and plural *wh*-phrases are semantically ambiguous: they either live on a set of individuals *A*, or a set of conjunctions and disjunctions (i.e. generalized universal or existential quantifiers over subsets of *A*). Accordingly, the true answer set of (19a) includes a higher-order disjunctive answer, while that of (19b) contains only individual answers.

(19)   (*w*: the committee can and can only be chaired by either John or Mary.)

      a. Which professor can chair the committee? $Q_w = \{\Diamond chair'(j), \Diamond chair'(m)\}$

      b. Who can chair the committee? $Q_w = \{\Diamond chair'(j), \Diamond chair'(m), \Diamond chair'(j \vee m)\}$

Second, Fox (2013) proposes a weaker presupposition for the Ans$_F$-operator: Ans$_F(Q)(w)$ presupposes that there exists a possible answer whose *innocently exclusive (IE)-exhaustification* is true in *w*.

(20) $\text{Ans}_F(Q)(w) = \exists p \in Q[w \in \text{IE-Exh}(p, Q)].\text{MaxI}(Q_w)$ (Final version)

Unlike the traditional exhaustification (21) which negates all the non-weaker alternatives (see Chierchia et al. 2012 for a review), IE-exhaustification negates only innocently excludable alternatives, as defined in (22) (Fox 2007). An alternative *q* is innocently excludable to *p* iff affirming *p* and negating *q* is consistent with negating any other non-weaker alternatives of *p*.

(21) $\text{Exh}(p, Q) = p \wedge \forall q \in \text{Excl}(p, Q)[\neg q]$, where $\text{Excl}(p, Q) = \{q : q \in Q \wedge p \nsubseteq q\}$

(22)   a. $\text{IE-Exh}(p, Q) = p \wedge \forall q \in \text{IExcl}(p, Q)[\neg q]$

      b. $\text{IExcl}(p, Q) = \{q : q \in Q \wedge \neg \exists q' \in \text{Excl}(p, Q)[p \wedge \neg q \rightarrow q']\}$

In (19b), among the three true answers, the individual ones are not innocently excludable to the disjunctive one, because $\neg \Diamond chair'(j) \wedge \neg \Diamond chair'(m)$ contradicts $\Diamond chair'(j \vee m)$; the IE-exhaustification of $\Diamond chair'(j \vee m)$ does not negate any of the true answers and thus is true. In contrast, (19b) has no answer whose IE-exhaustification is true: IE-exhaustifying $\Diamond chair'(j)$ yields the negation of the other true answer $\Diamond chair'(m)$, and vice versa.

Challenges from quantified questions   The presupposition of Ans_F, however, is still too strong to rule in individual MS readings of questions with a universal quantifier. The question (23) has two types of MS readings: (i) an individual MS reading, namely that *tell me one of the places where everyone can get gas*; and (ii) a pair-list MS reading, namely that *for each individual, tell me one of the places where he can get gas*. Let us focus on the individual MS reading. Note that the answers where *every* scopes below *can* are all false in w.

(23)   Where can everyone get gas?
    (w: everyone can get gas from station A, and everyone can get gas from station B. But both A and B have very limited stock, and thus it is impossible that everyone gets gas)

   a. $Q = \{\forall y \in man'[\Diamond \textit{get-gas}'(y, x)] : x \in {}^{*}\textit{place}'\}$

   b. $Q_w = \left\{ \begin{array}{l} \forall y \in man'[\Diamond \textit{get-gas}'(y, a)] \\ \forall y \in man'[\Diamond \textit{get-gas}'(y, b)] \\ \forall y \in man'[\Diamond \textit{get-gas}'(y, a \vee b)] \end{array} \right\}$

   c. $\text{Ans}_F(Q)(w)$ is undefined   Problem!

Unlike the case in (19a), here the true individual answers are innocently excludable to the true disjunctive answer. Thus IE-exhaustifying the disjunctive answer negates the individual ones, yielding a false inference that *some but not all of the people can get gas from A, the others can get gas from B*. Therefore, the new definition of Ans_F in (20) predicts a presupposition failure in (23), contra the fact. This problem extends to other quantified questions, for instance:

(24)   a.   Where can half of your friends get gas?
       b.   Where can most of your friends get gas?

2.3   Interim summary

To sum things up for this section, both Dayal (1996) and Fox (2013) attribute the uniqueness requirement of singular questions to an existential presupposition of the Ans-operator. In particular, the Ans_D-operator by Dayal requires the existence of the strongest true answer. This requirement leaves no space for MS. The Ans_F-operator by Fox requires the existence of a possible answer that has a true IE-exhaustification inference. This requirement is still too strong to allow for individual MS readings in several quantified questions.

3   My Analysis

Uniqueness requirement: a closure requirement on true short answers   We are now in a dilemma between uniqueness and MS: Dayal's (1996) account of uniqueness predicts that a question is not subject to uniqueness if the answer space is closed under conjunction; while Fox's (2013) MaxI-based account of MS predicts that MS is available only in an answer space that is not closed under conjunction. These two predictions are clearly contradictory.

   To solve this dilemma, I will keep the basics of Fox's account of MS and revise Dayal's account of uniqueness as the following: **the conjunction of any true *short answers* must be a possible short answer.** This condition predicts that a question is not subject to uniqueness iff the set of short answers is closed under conjunction, regardless of whether the set of full answers is also closed.

   The intuition is quite simple. (25a) cannot take both *J* and *M* as true short answers because it cannot take *J and M* as a possible short answer (cf. (25b)); likewise, (26a) cannot take both *J+M* and *J+S* as true short answers because it cannot take *J+M and J+S* as a possible short answer (cf. (26b)).

(25)   a. Which professor came?   (26)   a. Which two professors formed a committee?

   b. Which professors came?   b. Which professors formed a committee?

Conjunction and disjunction are defined cross-categorically as in (27), à la *meet* and *join* in Inquisitive Semantics (Ciardelli & Roelofsen 2014).

(27)   a. $\alpha_\tau \wedge \beta_\tau = \{\langle P_{\langle \tau,st \rangle}, w \rangle : P_w(\alpha)\} \cap \{\langle P_{\langle \tau,st \rangle}, w \rangle : P_w(\beta)\} = \lambda P_{\langle \tau,st \rangle}.\lambda w.P_w(\alpha) \wedge P_w(\beta)$

   b. $\alpha_\tau \vee \beta_\tau = \{\langle P_{\langle \tau,st \rangle}, w \rangle : P_w(\alpha)\} \cup \{\langle P_{\langle \tau,st \rangle}, w \rangle : P_w(\beta)\} = \lambda P_{\langle \tau,st \rangle}.\lambda w.P_w(\alpha) \vee P_w(\beta)$

There are various linguistic methods to retrieve the short answers. Syntactically, short answers can be retrieved from full answers by ellipsis. Semantically, we can firstly extract a *topical property* out of the Hamblin set, as schematized in (28): $P_{\langle \tau,st \rangle}$ is a topical property of Q iff **(a)** P has the same set of partitions as Q; and **(b)** some possible answer of Q equals to some $P(\alpha)$.

(28)   TP$\langle Q, P_{\langle \tau,st \rangle} \rangle$ iff

  (a) $\forall w \forall w'[[\lambda \alpha_\tau.P_w(\alpha) = \lambda \alpha_\tau.P_{w'}(\alpha)] \leftrightarrow [Q_w = Q_{w'}]]$
    (P has the same set of partitions as Q: for any two worlds w and w', the items having the property P are the same in w and w' iff the true answers of Q are the same in w and w'.)

  (b) $\exists p \exists \alpha_\tau[p \in Q \wedge p = P(\alpha)]$
    (For some p that is a possible answer of Q, there is some α s.t. p equals to $P(\alpha)$.)

Condition **(a)** makes use of Partition Semantics (Groenendijk & Stokhof 1984). But since a property and its negative counterpart (e.g. $\lambda x.\lambda w.came_w(x)$ vs. $\lambda x.\lambda w.\neg came_w(x)$) have the same set of partitions, we also need the Hamblin style condition **(b)** to rule out undesired properties.

Given a topical property P, a true short answer of Q in w is an item true for P in w (i.e. $\{\alpha : P_w(\alpha)\}$), and a possible short answer of Q is an item true for P in some world (i.e. $\{\alpha : \exists w[P_w(\alpha)]\}$). We can now state the closure requirement as in (29).

(29)   **Closure (C)-Requirement**
  For Q being defined in w, it must have a topical property P s.t. the conjunction of any items that are true for P in w is also true for P in some world.
  $\exists P_{\langle \tau,st \rangle}[\text{TP}\langle Q, P \rangle \wedge \forall \alpha_\tau \forall \beta_\tau[P_w(\alpha) \wedge P_w(\beta) \rightarrow P(\alpha \wedge \beta) \neq \bot]]$

First, C-requirement predicts that a question requires uniqueness iff the conjunction of any two distinct short answers is not a possible short answer. The singular question (25a), under a *de re* or *de dicto* reading, provides the topical properties in the form of (30a) or (30b), respectively. (30a/b) maps an actual/possible professor to a possible answer. Since non-atomic item is false for *professor′* in every world, and thus both (30a-b) map a non-atomic item to a contradiction. Therefore, the C-requirement is met in w iff (30a/b) holds for a unique atomic item in w.

(30)   Which professor came?

  a. $\lambda \alpha.\lambda w.\alpha \in professor'_{w_@} \wedge came'_w(\alpha)$ *De re* reading

  b. $\lambda \alpha.\lambda w.\alpha \in professor'_w \wedge came'_w(\alpha)$ *De dicto* reading

Second, C-requirement and that a question is not subject to uniqueness iff the set of short answers is closed under conjunction. Given the fact that number-neutral and plural *wh*-questions can take elided conjunctions and disjunctions as complete answers, I assume that bare *wh*-words and plural *wh*-phrases live on a set $^{\text{INT}}*P$, which consists of not only members of the *wh*-complement $*P$, but also disjunctions and conjunctions. For instance, if $*professor' = \{a, b, a \oplus b\}$, then $^{\text{INT}}*professor' = \{a, b, a \oplus b, a \vee b, a \wedge b, a \vee (a \oplus b), a \vee b \vee (a \oplus b), ..., a \vee (a \wedge b), a \wedge (a \vee b), ...\}$. Thus the topical properties of the plural question (25b) are like (31a-b). They map a conjunction of actual/possible professors to a possible answer. Thus the C-requirement is always satisfied. Obviously, this prediction also applies to quantified questions.

(31) Which professors came?

    a. $\lambda\alpha.\lambda w.\alpha \in {}^{\text{INT}}*professor'_{w_@} \wedge came'_w(\alpha)$ *De re* reading

    b. $\lambda\alpha.\lambda w.\alpha \in {}^{\text{INT}}*professor'_{w} \wedge came'_w(\alpha)$ *De dicto* reading

My Ans_X-operator is thus defined as the following: $\text{Ans}_X(Q)(w)$ presupposes the C-requirement and the existence of at least one MaxI true answer of Q in w[3]; when defined, $\text{Ans}_X(Q)(w)$ returns the set of of MaxI true answers of Q in w.

(32) $\text{Ans}_X(Q)(w)$ is defined iff (i) $\exists P_{\langle\tau,st\rangle}[\text{TP}\langle Q, P\rangle \wedge \forall\alpha\forall\beta[P_w(\alpha) \wedge P_w(\beta) \rightarrow P(\alpha \wedge \beta) \neq \bot]]$, and
    (ii) $\exists p \in \text{MaxI}(Q_w)$. When defined, $\text{Ans}_X(Q)(w) = \text{MaxI}(Q_w)$

**MS/MA ambiguity: scope ambiguity of the higher-order *wh*-trace**   Unlike Dayal (1996), the C-requirement leaves space for MS: the conjunctive closure within the short answers can take scope at any propositional level within IP, and thus closing the set of short answers under conjunction does not necessarily close the set of full answers under conjunction. I argue that a $\diamondsuit$-question admits an MS reading when the conjunctive closure of the short answers takes scope below the existential modal.

Recall that bare *wh*-word and plural *wh*-phrases live on a set consisting items of various types. Thus, the *wh*-trace of a bare *wh*-word or a plural *wh*-phrase is type flexible: if *who* is moved directly from a theta position, as in (33a), the *wh*-trace is of type e; if the *who* undertakes one or multiple QRs before moving to the spec of the interrogative CP, as in (33b), the *wh*-trace takes a higher type.

(33) Who came?

    a.
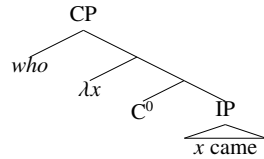
    b.
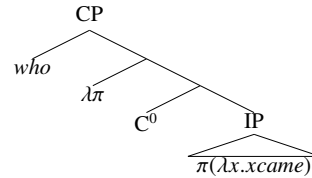
In the case of a number-neutral or plural $\diamondsuit$-question, the higher-order trace π can take scope above or below the weak modal, as illustrated in (34a) and (34b), respectively. I argue that MS is available if π scopes below the existential modal.[45]

(34) Who can chair the committee?
    (*w*: the committee can and can only be chaired by John or Mary.)

    a. Q: $\{\pi(\lambda x.\diamondsuit chair'(x)) : \pi \in {}^{\text{INT}}*person'_{w_@}\}$

    b. Q: $\{\diamondsuit\pi(\lambda x.chair'(x)) : \pi \in {}^{\text{INT}}*person'_{w_@}\}$

---
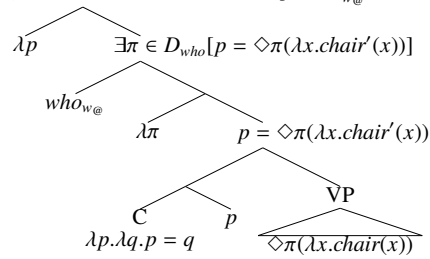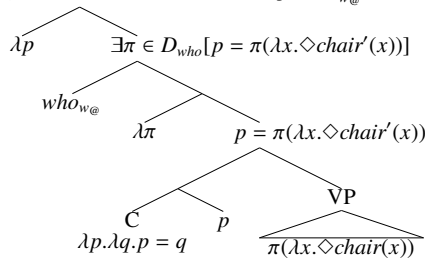
[3] The existential presupposition is to capture the negative island effects of degree questions.

[4] Here I only consider conjunctive MA answers. See Xiang (2015b) for discussions on disjunctive MA answers.

[5] This idea is close to Fox (2013) in several respects. Fox assumes that the *wh*-trace x has a phrase-mate EACH, and argues that a $\diamondsuit$-question can take MS when [x EACH] scopes below $\diamondsuit$. Compared to Fox (2013), the present analysis has advantages in analyzing questions with a collective predicate. For instance, *who formed a team?* needs conjunctive answers like $form'(a{\oplus}b) \wedge form'(c{\oplus}d)$; this answer can be derived from $form'(a \oplus b \wedge c \oplus d)$, but not from $\text{EACH}(a \oplus b \oplus c \oplus d)(form')$.

If the trace π scopes above *can*, exercising Ans$_X$ returns the unique MaxI true answer, namely the conjunctive MA answer, as schematized in (35). Alternatively, if the *wh*-trace π scopes below *can*, the conjunctive and disjunctive closures scope below the weak modal. The set of true answers in (36a) has two MaxI members, both are MS answers. Note that the conjunctive answer $\diamond chair'(j \wedge m)$ is false, since there can be only one chairing-event for the considered committee.

(35) MA reading: $(\pi > \diamond)$

    a. $Q_w = \{\diamond chair'(j), \diamond chair'(m), \diamond chair'(j) \vee \diamond chair'(m), \diamond chair'(j) \wedge \diamond chair'(m)\}$

    b. $\text{Ans}_X(Q)(w) = \{\diamond chair'(j) \wedge \diamond chair'(m)\}$

(36) MS reading: $(\diamond > \pi)$

    a. $Q_w = \{\diamond chair'(j), \diamond chair'(m), \diamond[chair'(j) \vee chair'(m)]\}$

    b. $\text{Ans}_X(Q)(w) = \{\diamond chair'(j), \diamond chair'(m)\}$

Another question arises with the case of local conjunction. In (37), the local conjunctive answer (37a-i) is true and asymmetrically entails the individual answer (37a-ii). But intuitively the individual answer *Mary can help* is a good MS answer. In responding to this problem, I propose that the weak modal *can* optionally embeds an Exh-operator (defined in (21)) associated with the *wh*-trace, given that *Mary can help* intuitively means *Mary alone can help*. The Exh-operator creates a non-monotonic environment with respect to the *wh*-trace, breaking up the entailment relation from (i) to (ii), as schematized in (37b). Thus both the conjunctive answer and the individual answer are preserved as MaxI true answers.

(37) Who can help John?

    (*w*: only Mary helps John in w_1, both Mary and Sue help John in *w*2)

    a. (i) $\diamond[help'(m, j) \wedge help'(s, j)]$ $\Rightarrow$ (ii) $\diamond help'(m, j)$

    b. (i) $\diamond \text{Exh}[help'(m, j) \wedge help'(s, j)] \Rightarrow$ (ii) $\diamond \text{Exh}[help'(m, j)]$

Anti-presuppositions of plural questions   Recall that a plural question rejects an MS answer that names only a singularity: in (38), only (38b) admits MS, which names a sum.

(38) a. Which professors can chair the committee?   (# MS, $^{\text{OK}}$ MA)

     b. Which professors can form the committee?   ($^{\text{OK}}$ MS, $^{\text{OK}}$ MA)

Sauerland et al. (2005) make use of the principle of *Maximize Presupposition* (MP) (Heim 1991) to analyze inferences evoked by plural-markings. This MP principle requires that out of two sentences which are presuppositional alternatives and which are contextually equivalent, the one with the stronger presuppositions must be used if its presuppositions are met in the context. Accordingly, Sauerland et al. (2005) argue that singulars are more presuppositional than plurals, and thus that plural-morphemes implicate an "anti-presupposition" that *the singular counterpart is undefined*.

Following this idea, I propose that the plural-morpheme on the *wh*-complement implicates an anti-presupposition: *the corresponding singular question is undefined*. Further, in spirit of the question-answer congruence, I propose that a proper answer of a plural question entails the anti-presupposition.

(39) A proposition p properly answers Q_{PL} in w iff

    a. $p \in \text{Ans}_X(Q_{\text{PL}})(w)$; and b. $p \subseteq \lambda w. \text{Ans}_X(Q_{\text{SG}})(w)$ is undefined.

Accordingly, the plural question (38a) rejects MS because an MS answer does not entail the anti-presupposition that *the singular question 'which professor can chair the committee' is undefined*. In contrast, (38b) admits MS because its MS answers do entail the anti-presupposition that *the singular question 'which professor can form the committee' is undefined*.

4 Conclusions

This paper investigates the interactions between the distributional pattern of MS answers and the number-markings on *wh*-complements. First, I argue that the presuppositions of Dayal's (1996) and Fox's (2013) Ans-operators are both too strong to capture MS. Alternatively, I propose to account for the uniqueness requirement of singular questions with a closure requirement on short answers: the set of true short answers must be closed under $\cup$ or $\cap$. Second, I show that the MS/MA ambiguity of a $\diamond$-question can be explained by the scope ambiguity of the higher-order *wh*-trace: a $\diamond$-question admits MS when the higher-order *wh*-trace scopes below the existential modal. Last, I argue that a good answer to a plural question needs to entail the anti-presupposition of the plural morpheme.

References

[1] Gennaro Chierchia, Danny Fox, and Benjamin Spector. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *An International Handbook of Natural Language Meaning*, pages 2297–2332. Mouton de Gruyter, 2012.

[2] Ivano Ciardelli and Floris Roelofsen. Alternatives in montague grammar. In *Proceedings of Sinn und Bedeutung*, 2014.

[3] Veneeta Dayal. *Locality in wh quantification*, volume 62. Springer Science & Business Media, 1996.

[4] Danny Fox. Free choice disjunction and the theory of scalar implicatures. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Iimplicature in Compositional Semantics*, pages 71–120. New York: Palgrave Macmillan, 2007.

[5] Danny Fox. Mention-some readings of questions. *MIT seminar notes*, 2013.

[6] Benjamin Ross George. *Question embedding and the semantics of answers*. PhD thesis, University of California Los Angeles, 2011.

[7] Jeroen Groenendijk and Martin Stokhof. On the semantics of questions and the pragmatics of answers. *Varieties of formal semantics*, 3:143–170, 1984.

[8] Irene Heim. Artikel und definitheit. *Semantik: ein internationales Handbuch der Zeitgenössischen forschung*, pages 487–535, 1991.

[9] Godehard Link. The logical analysis of plurals and mass terms: A lattice-theoretic approach. 1983.

[10] Uli Sauerland, Jan Anderssen, and Kazuko Yatsushiro. The plural is semantically unmarked. *Linguistic evidence*, pages 413–434, 2005.

[11] Katrin Schulz and Robert Van Rooij. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, 29(2):205–250, 2006.

[12] Benjamin Spector. Modalized questions and exhaustivity. In *Semantics and Linguistic Theory*, pages 282–299, 2007.

[13] Benjamin Spector. An unnoticed reading for wh-questions: Elided answers and weak islands. *Linguistic Inquiry*, 39(4):677–686, 2008.

[14] Robert Van Rooij and Katrin Schulz. Exhaustive interpretation of complex sentences. *Journal of logic, language and information*, 13(4):491–519, 2004.

[15] Yimei Xiang. Deriving disjunctive answers. Questions in Logic and Semantics, satellite workshop of Amsterdam Colloquium, 2015b.

Existential presupposition projection from *none*? An experimental investigation[*]

Jérémy Zehr[1], Cory Bill[2], Lyn Tieu[3], Jacopo Romoli[4] and Florian Schwarz[1]

[1] UPenn, Philadelphia, U.S.A.
florians@babel.ling.upenn.edu, jeremyz@lab.florianschwarz.net
[2] Macquarie University, Sydney, Australia
cory.bill@students.mq.edu.au
[3] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Paris, France
lyn.tieu@gmail.com
[4] Ulster University, Belfast, U.K.
j.romoli@ulster.ac.uk

Abstract

The question of how presuppositions project from the scope of quantificational sentences, and in particular negative quantificational sentences such as *none* in (1), continues to be controversial, both theoretically and empirically: some theories only predict the existential presupposition projection reading in (1-a) (for example, [2, 3, 26, 13]), while others derive the universal projection reading in (1-b) ([15, 20, 21, 12, 10, 11], among others). In addition, any theory has to account for presupposition suspension, yielding an interpretation without a (global) presupposition (1-c).

(1)   None of the bears won the race.
    a.   *At least one of the bears participated and none of them won.*
    b.   *All of the bears participated and none of them won.*
    c.   *None of the bears both participated and won.*

Previous empirical studies have found evidence for universal projection ([7]), while others have provided evidence for alternatives to universal projection ([24, 14]). To our knowledge, however, there exists no definitive positive evidence for the existential reading in (1-a). We report a study that directly compares the existential, universal, and presuppositionless readings of (1) through the use of a 'covered box' picture selection task [16, 5]. We find clear evidence for existential readings (as well as presuppositionless readings), but no evidence for universal ones. This result challenges theories that predict only universal readings. Our results, taken together with those reported in [7], suggest that any adequate account of presupposition projection must be able to explain all three interpretive options in (1).

1   Introduction

There is a long-standing debate in the presupposition literature concerning the projection of presuppositions from the scope of quantifiers, and in particular of negative quantifiers such as *none* in (2) (the underlined content corresponds to the presuppositional content). While some theories predict

existential presupposition projection (2-a) ([2, 3, 26, 13], among others), others derive universal projection (2-b) ([15, 20, 21, 12, 10, 11]).[1] Additionally, both views assume a mechanism for suspending presuppositions, in order to account for the presuppositionless interpretation in (2-c).

(2)   None of the bears won the race.
        a.   EXISTENTIAL: *At least one of the bears participated* and none of them won.
        b.   UNIVERSAL: *All of the bears participated* and none of them won.
        c.   PRESUPPOSITIONLESS: *None of the bears both participated and won.*

The present study focuses on the prediction of what we will call Universal-Only theories, stated in (3).

(3)   **Prediction of Universal-Only theories:**
        Sentences like (2) only give rise to the universal projection reading (2-b) and the presuppositionless reading (2-c), but not to the existential projection reading (2-a).

[7] reports evidence from an inferential task paradigm for the universal projection reading in (2-b). Such a result is in line with the prediction in (3). More recently, [24] and [14] report evidence from truth value judgment tasks for interpretations of sentences like (2) that do not involve a universal projection reading. However, their designs do not allow us to distinguish between existential and presuppositionless readings. Their results are therefore compatible with the prediction in (3), and do not constitute a challenge for Universal-Only theories, which can capture non-universal responses to (2) through the reading in (2-c). To summarize, there exists evidence for the availability of universal projection readings of sentences like (2), compatible with (3), but there is no direct evidence for existential projection readings. Existing experimental data thus seem to be consistent with Universal-Only theories of presupposition projection in quantificational sentences.

The present study investigates sentences like (2) and directly compares the existential and universal interpretations through the use of a 'covered box' picture selection task [16, 5]. We find clear evidence for existential projection readings and presuppositionless readings, but no evidence for universal projection readings. This result challenges the prediction of Universal-Only theories stated in (3), and, when taken together with the results of [7], indicates that accounts of presupposition projection must be able to capture all three interpretive options in (2).

The rest of this paper is organized as follows. In 1.1, we briefly review the notions of presupposition and presupposition projection in quantificational sentences. In 1.2, we summarize the three previous studies mentioned above, and flesh out the motivation of the present study based on these previous findings. In section 2, we present our experiment and discuss the results. In section 3, we conclude with discussion of possible extensions of this work.

1.1   Theoretical background

Sentences containing expressions like *win*, *stop*, and *know* systematically give rise to inferences with certain characteristic properties, traditionally called 'presuppositions.' For example, sentences like (4-a) and (5-a) presuppose (4-b) and (5-b), respectively.

(4)   a.   Bear won.   (5)   a.   Bear stopped running.
       b.   ⇝ *Bear participated*   b.   ⇝ *Bear was running before*

The first characteristic property of presuppositions is that they are generally inherited by complex sentences. For example, in contrast to standard truth-conditional meaning, the inference (4-b)(=(6-d))

---

[1]Note that our Universal-Projection category encompasses both theories that predict universal projection uniformly for all quantifiers such as [15, 20, 21] and theories that make mixed predictions depending on the quantifier involved ([12, 10, 11]). As will become clearer in the discussion below, this distinction is irrelevant for the case of negative quantificational sentences, as these theories make the same universal-only prediction for these cases. Additionally, there are theories like [8] and [18], which predict both the universal and existential projection readings (see subsection 2.2.2).

'projects' from embedded environments like (6-a)-(6-c). That is, the inference remains present as an inference of the overall sentence, despite the embedding under negation, in a question, or in the antecedent of a conditional, respectively.

(6) a. Bear didn't win.
     b. Did Bear win?
     c. If Bear won, he will celebrate.
     d. $\rightsquigarrow$ *Bear participated*

A second important property of presuppositions is that projection is not strictly obligatory. Presuppositions in embedded sentences can in fact be absent at the global level of the entire sentence. For example, while (6-a) generally gives rise to the inference in (6-d), it can be suspended or cancelled, as illustrated by the felicitous continuation in (7-b).

(7) a. Bear didn't win.
     b. ... He didn't even participate!

Without going into the details of any specific accounts, a common approach to presupposition suspension is to make it possible for the presupposition to be interpreted in the scope of negation. This operation of suspension is commonly referred to as *local accommodation* [15], as the presupposition is 'accommodated' locally in the scope of the relevant operator. If the presupposition of (7-a) is accommodated under negation, the sentence ends up with a meaning that can be paraphrased along the lines of, *It's not true that Bear participated and won*. This meaning is hence compatible with the continuation in (7-b).

While the empirical picture is relatively clear for the cases in (6-a)-(6-c), the situation is more complex and controversial for quantificational sentences like (8). The central question is whether a sentence like (8) has an *existential* projection reading (8-a), a *universal* projection reading (8-b), or both (adopting the schematic notation of [7]).

(8) None of the bears won. $[Qx : R(x)]\ S_p(x)$
     a. EXISTENTIAL: $\rightsquigarrow$ *At least one of the bears participated* $[\exists x : R(x)]\ p(x)$
     b. UNIVERSAL: $\rightsquigarrow$ *All of the bears participated* $[\forall x : R(x)]\ p(x)$

In addition, all theories account for the presuppositionless reading of quantificational sentences (9), parallel to those for negation. As in the case of negation, we can posit a mechanism for interpreting the presupposition in the scope of the negative quantifier, yielding an interpretation that can be paraphrased as *None of the bears is such that it both participated and won*, schematically represented in (10).

(9) None of the bears won ... none of them even participated!

(10) $[Qx : R(x)]\ p(x) \wedge S_p(x)$ PRESUPPOSITIONLESS

In sum, we can distinguish two types of theories of projection from quantifiers: Existential-Only theories predict only the existential inference in (8-a) ([2, 3, 26, 13]) and Universal-Only theories predict only the universal inference in (8-b), either for all quantifiers [15, 20, 21], or for negative quantifiers [12, 10, 11]. All existing theories predict the possibility of presuppositionless readings such as (10).

1.2   Previous studies

While experimental work on presuppositions has grown considerably in recent years (see [22] and [23] for recent overviews), we are aware of only three studies that have specifically targeted presupposition projection in quantificational sentences containing negative quantifiers.[2] We review these below.

---

[2]See [25] for a study that looks at the presuppositions of possessives and *again* in the context of universal and existential quantifiers, with results that are argued to provide evidence for universal and existential presuppositions for the respective quantifiers.

1.2.1 Evidence for universal projection

[7] used an inference task to investigate quantified presuppositional sentences in French. Participants were asked to decide whether certain universal and existential inferences were 'suggested' by a variety of quantified sentences. The example in (11) illustrates the case of a universal inference involving the trigger *know* under a negative quantifier and its potential universal inference.

(11)   "None of these 10 students knows that he is lucky."
        suggests that:
        Each of these 10 students is lucky.
        No?   *Yes?*

[7] compared presupposition projection in sentences like the one in (11) with cases involving scalar implicatures like (12), which involves the potential universal inference of a negatively quantified sentence embedding a strong scalar term, *all*. Moreover, he compared cases like (11) and (12) with corresponding cases involving other quantifiers such as *every*, *some*, and *more than one*.

(12)   "None of these 10 students missed all of their exams."
        suggests that:
        Each of these 10 students missed some of their exams.
        No?   *Yes?*

In the presupposition condition, participants were reported to endorse the universal inference in the case of the negatively quantified sentences more than 80% of the time, a percentage which was significantly higher than in the cases involving other (non-universal) quantifiers. Moreover, in the case of negative quantifiers, no difference was observed between the acceptance of existential and universal inferences. Importantly, in the corresponding scalar implicature condition, the corresponding universal inference was endorsed significantly less often than the existential inference. The difference between the acceptance rate of the universal inference in negatively quantified sentences versus sentences involving other quantifiers, together with the interaction between the force of the projection inference (universal vs. existential) and the type of inference (presupposition vs. scalar implicature), provide evidence for the availability of universal projection readings in cases like (11), compatible with the prediction in (3).

1.2.2 Evidence for existential projection

In a more recent study, [24] used a truth value judgment task to investigate whether participants would judge sentences such as (13-a), involving the presupposition trigger *both* in the scope of *none*, to be good descriptions of pictures that falsified the universal projection reading (13-b), i.e. where one of the depicted circles had only one square in its cell. Nonetheless, [24] found that around half of the participants were happy to accept (13-a) as a description of the picture. This result suggests that at least some speakers allow non-universal projection readings for sentences like (13-a).

(13)   a.   None of these three circles have the same color as both of the squares in their own cell.
        b.   UNIVERSAL: $\rightsquigarrow$ *All of the circles have exactly two squares in their own cell*

In a similar study, [14] report evidence for non-universal projection readings of *none*-sentences. In the relevant condition, participants saw pictures where, for instance, four of five circles were connected to a square (always of a different color). They then had to answer 'True', 'False', or 'Don't know' to the description in (14). Participants accepted such sentences more than 92% of the time (Experiment 1), despite their being false on the universal projection reading in (14-b).[3]

---

[3]Note that [14] were primarily interested in the possibility of restricting the domain of universal quantification to those individuals that satisfy the presupposition of the scope of the quantifier. Domain restriction is indeed an important variable which interacts with (what appears to be) a universal versus an existential projection inference. [7] used an explict partitive to control for domain restriction (*among these ten students, none of them...*). We also explicitly introduced the domain using an overt numeral in the description of the context

(14)   a.   No circle has the same color as the square to which it is connected.
        b.   UNIVERSAL: ↝ *All of the circles are connected to a square*

In sum, both of these studies provide evidence for non-universal projection. However, they do not allow us to distinguish between existential projection and presuppositionless readings, and therefore they leave open the status of the prediction in (3). We turn next to our experiment, which aimed to directly assess whether existential projection readings exist, allowing us to test the prediction in (3).

2   Experiment

Our experiment investigated the possible interpretations of sentences like (15), and in particular focused on the prediction of Universal-Only theories, repeated below in (3).

(15)   None of the bears won the race.
        a.   EXISTENTIAL: *At least one of the bears participated and none of them won.*
        b.   UNIVERSAL: *All of the bears participated and none of them won.*
        c.   PRESUPPOSITIONLESS: *None of the bears both participated and won.*

(16)   **Prediction of Universal-Only theories:**
        Sentences like (15) only give rise to the universal projection reading (15-b) and the presuppositionless reading (15-c), but not to the existential projection reading (15-b).

We presented recordings of sentences like (15), accompanied by pictures that varied in whether *none* or *only some* of the bears participated. Notice that both of these contexts are equally *incompatible* with the reading in (15-b), and equally *compatible* with the reading in (15-c). Based on (16), which predicts (15-b) and (15-c) to be the only possible interpretations of (15), we thus expect no effect of the context manipulation. In contrast, if the existential projection reading in (15-a) exists, a difference based on context could emerge if that reading is more readily available than (15-c), since (15-a) is compatible with the *only some* context but not with the *none* context.

2.1   Methods

2.1.1   Participants

We tested 36 native speakers of English, recruited through Amazon Mechanical Turk. Participants took about 10 minutes to complete the task, and received $1 for their participation.

2.1.2   Materials & procedure

We used a Covered Box paradigm ([16]), adapting the designs of [24] and [5]. Participants were presented with sentences like (15), accompanied by two images, and were instructed to choose the image they thought matched the sentence. Crucially, one image was visible, and one was introduced to participants as 'hidden' (behind a black square). Participants were expected to select the visible image if it was consistent with the sentence, and the covered image otherwise. Each trial included a context followed by a target, as illustrated in (17); see Figure 1 for the accompanying images. The inclusion of the context picture and description was meant to improve felicity of the target sentence, and, more importantly, to control for domain restriction by explicitly introducing the three relevant animals (e.g, *these three bears...*), which are described both in the context and the target picture.

(17)   *Context (Fig. 1a):* In the morning race, these three bears did really well, and in the end one
        of them won. I thought they would do well later in the day as well, but...
        Target (Fig. 1a or 1b): None of the bears won the afternoon race.

immediately preceding the target sentence (see section 2.1.2).

There were two kinds of target pictures, as illustrated by the images in Figures (1b) and (1c). Participants were presented with either (b) or (c) alongside a black box representing the *covered* picture. OnlySome targets (c) were inconsistent with the universal projection reading (that all bears participated), but consistent with the existential projection reading (that at least one bear participated) and with the presuppositionless reading (that no bear both participated and won). Covered picture choices in this condition therefore indicated access to a universal projection reading. The NoRunner targets (b) were inconsistent with the existential and universal projection readings, but consistent with the presuppositionless reading. Visible picture choices in this condition therefore indicated the availability of the presuppositionless reading, while covered picture choices indicated access to one of the other two readings. Since covered picture choices indicate presuppositional readings, we will present the results in terms of the rate of *covered picture* selections.

The prediction of Universal-Only theories is that participants should choose the covered picture equally often in both conditions, since they do not assume an existential reading. Both target pictures are only expected to be acceptable if a presuppositionless reading can be accessed. In contrast, if existential projection is possible, acceptance of (c) does not require access to the presuppositionless reading, while (b) does. Assuming the existential reading is more readily available than the presuppositionless one (given that local accommodation is commonly assumed to be dispreferred), we would expect different response patterns for (b) and (c).
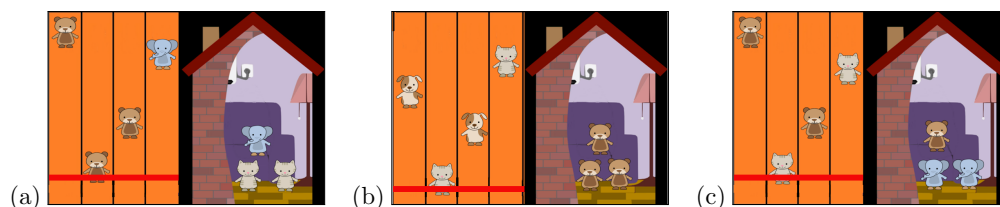

Figure 1: Sample target images accompanying (17). (a) depicts the context image; (b) depicts the *visible* picture on a NoRunner target; (c) depicts the *visible* picture on an OnlySome target. On a given trial, (b) or (c) was presented alongside a *covered* picture.

The two critical test conditions were presented in blocks with order counterbalanced across participants. All participants received four OnlySome targets and four NoRunner targets, as well as two clearly true and two clearly false controls in which the universal presupposition was satisfied, i.e. where all of the bears participated in the race, and one of them won (False control) or none of them did (True control). These control conditions provided a baseline for identifying presuppositionless and universal readings. The presuppositionless reading was compatible with the NoRunner targets but not with the False controls. Greater acceptance on NoRunner targets than on False controls was therefore indicative of presuppositionless readings. On the other hand, the universal projection reading was compatible with the True controls but not with the OnlySome targets (the other two readings were compatible with both). Greater rejection on OnlySome targets than on True controls was therefore indicative of universal projection readings. Finally, eight additional true/false controls were included in order to make sure that participants could correctly identify who had participated in a given race, and to ensure that participants could respond correctly to non-presuppositional sentences containing *none* and *participate*.

Response times (RTs) for response choices were also collected, in particular because previous experimental work on presupposition ([9, 19, 4], among others) suggests that presuppositionless responses are associated with delays in response times. This measure therefore has the potential to provide further information about the nature of the readings on which responses are based.
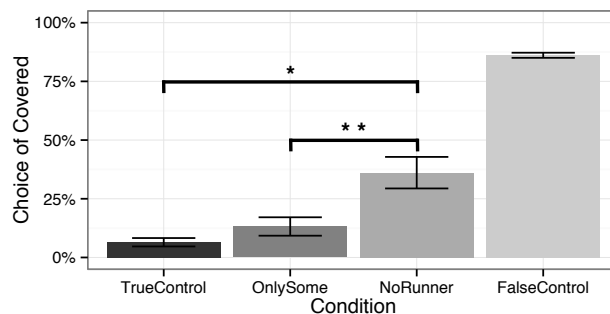
Figure 2: Percentage of covered picture selections in target and control conditions. Covered picture selections indicated access to universal projection readings in the ONLYSOME condition, and to either universal or existential projection readings in the NORUNNER condition.

2.2   Results & Discussion

2.2.1   Summary of results

Figure 2 presents the mean percentages of covered picture choices in the target and control conditions. A mixed-effect logistic regression using the maximal random effects structure that would converge ([1]), with random intercepts for participants and items, revealed more covered picture selections in the NORUNNER target condition than in the ONLYSOME target condition ($p < .01$) or the TRUE control condition ($p < .05$). The latter two conditions did not differ significantly from each other. The NORUNNER target condition also yielded significantly fewer covered picture selections than the FALSE control condition ($p < .01$).

For purposes of RT analyses, trials with RTs greater than two standard deviations above the mean were removed from the data (constituting 4.2% of the data, with equal distribution across NORUNNER and ONLYSOME conditions). A mixed-effect regression analysis, with random intercepts for participants and items, revealed that *visible* picture selections in the NORUNNER condition took significantly longer than in the ONLYSOME condition ($M = 4241ms$ vs. $M = 3855ms$; $\beta = -417.8$, $SE = 174.2$, $t = -2.398$).

2.2.2   Discussion

The greater rate of covered picture choices in the NORUNNER target condition, as compared to the ONLYSOME target condition, provides direct evidence for the existence of an existential projection reading. In contrast, we found no evidence that participants accessed a universal projection reading, with no significant differences observed between the ONLYSOME target and TRUE control conditions. The evidence for existential readings runs counter to the Universal-Only prediction in (16), and is therefore problematic for theories committed to this prediction ([15, 20, 21, 12, 10, 11], among others). Our results also provide evidence for the existence of a presuppositionless reading, as the visible picture was selected in the NORUNNER condition over 60% of the time. This interpretation can be accounted for in terms of local accommodation. The RT results further support the notion that acceptance in the NORUNNER condition requires local accommodation, whereas acceptance in the ONLYSOME condition does not; the relative RT delay for acceptances in the former condition is in line with previous findings of delays for local accommodation-based responses ([9, 19, 4]).

Our findings are in line with recent evidence for non-universal projection ([24, 14]). Unlike

the previous studies, however, we distinguish between genuine existential projection and presuppositionless readings, and find evidence for both interpretations. At this point, the relationship between our results and those of [7] deserves more discussion. While [7] found evidence for universal rather than existential projection readings, our results yielded the opposite pattern, with evidence for the latter but not for the former. Two points are worth highlighting here: first, these results are not necessarily incompatible, as each only provides positive evidence for one interpretation, and fails to provide direct evidence for the other (rather than providing direct evidence against it). Second, the two sets of studies utilized very different tasks, namely an inference-based task [7] and a picture selection task ([24, 14], and the present study), which could affect the outcomes. Moreover, Chemla's evidence for universal projection arose from the comparison between *none* and other quantifiers, and between presuppositions and scalar implicatures. Further investigation into potential effects of the different tasks, as well as the different quantifiers and environments, is therefore required in order to better understand the relationship between the two sets of results.

As things stand, both results need to be accounted for: (18) can be associated with the inference in (18-a) or the one in (18-b) (in addition to the presuppositionless reading).

(18)   None of the bears won.
    a.   ⤳ *At least one of the bears participated*
    b.   ⤳ *All of the bears participated*

This situation is, prima facie, equally incompatible with Universal-Only and Existential-Only theories. Both kinds of theories would have to be supplemented in such a way as to account for the respective missing readings. One way of accounting for all of the relevant results then is to try to spell out what these supplementary assumptions might be. For example, a Universal-Only theory could include a further mechanism that weakens the universal presupposition in a manner distinct from local accommodation, for example through domain restriction. Similarly, proponents of an Existential-Only theory could try to capture the apparent universal effect reported in [7] by appealing to additional reasoning beyond what presupposition projection yields. For instance, in an inference task, participants may find a uniform scenario where all individuals in the domain of the quantifier are homogeneous with regards to the presupposition more plausible on independent grounds. These possibilities need to be assessed in greater detail, particularly in relation to other aspects of the results, such as the reaction times and results for the other quantifiers.

Another way to deal with the emerging empirical picture is to build the option of both readings into the presupposition projection mechanisms themselves. As mentioned above, 'mixed' theories exist, according to which the force of the projection inference depends on the quantifier involved. Trivalent theories such as [12, 10, 11] fall into this category. However, while these theories make more nuanced predictions than 'pure' Universal-Only theories, by varying the force of projection with the quantifier, they make the same prediction for the present case (see [24] for discussion) and therefore do not fare better in explaining our results.[4]

Scalar implicature-based theories of presupposition, on the other hand, such as [8] and [18], do in principle predict both a universal and an existential reading for sentences like (18), and

---

[4]To be more precise, in addition to the universal and presuppositionless readings, these trivalent theories also predict the disjunctive presupposition reading in (i), which is weaker than the universal reading and stronger than the existential one. However, all of our targets were such that none of the bears won, so (i) in such contexts becomes equivalent to the universal reading, *All of the bears participated and none of them won*.

(i)   ⤳ *Either all of the bears participated and none of them won, or at least one bear participated and won*

are therefore more compatible with our results. However, by assimilating presuppositions to scalar implicatures (in particular presupposition triggers like *win* to scalar terms like *all*), such theories face challenges of their own. First, they need to explain [7]'s observed difference in the frequency of universal readings in the case of presuppositions and in the corresponding cases of scalar implicature, discussed in subsection 1.2.1. Second, data from different populations ([5, 17], among others) suggest differences between presuppositions and scalar implicatures that are unaccounted for by scalar implicature-based theories of presupposition.

3   Conclusion & Extensions

The goal of the present study was to assess the Universal-Only prediction in (16), by investigating the interpretation of sentences containing a presupposition trigger in the scope of the negative quantifier *none*. Our experiment revealed evidence for existential projection readings, and therefore provides evidence against the prediction in (16). Our results, taken together with those of [7], suggest that both universal and existential projection readings of such sentences need to be accounted for.

In comparing the different findings from the two studies, a further option to consider is that these are due to the fact that different presupposition triggers were used. Indeed, various authors have argued for differences between triggers with respect to the force of universal projection ([6, 11, 27]).[5] A promising next step then is to use both types of tasks with the relevant triggers, for example investigating *win* using an inferential task instead, and, conversely, investigating *stop*, one of the triggers used in [7]'s study, using a Covered Picture paradigm.

Finally, child language data may also be informative for our research questions. [5] report that children, unlike adults, tend not to access local accommodation readings under negation. Recall that in our paradigm, acceptance in the NoRunner condition requires local accommodation (of the existential presupposition), while acceptance in the OnlySome condition does not. [5]'s results would therefore lead us to expect that children tested on the current paradigm would select the visible image in the OnlySome condition but not in the NoRunner condition. Such a behavioral pattern in children would corroborate our interpretation of the present data.

References

[1] Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278, 2013.

[2] David Beaver. When variables don't vary enough. In M. Harvey and L. Santelmann, editors, *Proceedings of SALT 4*, pages 35–60. Cornell University, 1994.

[3] David Beaver. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications, Stanford University, 2001.

[4] Cory Bill, Jacopo Romoli, and Florian Schwarz. Are some presuppositions scalar implicatures? assessing evidence from reaction times. submitted, 2015.

[5] Cory Bill, Jacopo Romoli, Florian Schwarz, and Stephen Crain. Scalar implicatures versus presuppositions: The view from acquisition. *TOPOI*, pages 1–15, 2015.

[6] Simon Charlow. "Strong" predicative presuppositional objects. In *Proceedings of ESSLLI 2009*, Bordeaux, 2009.

---

[5][7] considers five different presupposition triggers, and reports an interaction between environments (i.e. presupposition vs. scalar implicature) and triggers in his first experiment, but no difference between triggers in a second experiment.

[7] Emmanuel Chemla. Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics*, 17(4):299–340, doi:10.1007/s11050-009-9043-9 2009.

[8] Emmanuel Chemla. Similarity: towards a unified account of scalar implicatures, free choice permission and presupposition projection. Unpublished manuscript, 2010.

[9] Emmanuel Chemla and Lewis Bott. Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes*, 38(3):241–260, 2013.

[10] Danny Fox. Two short notes on schlenker's theory of presupposition projection. *Theoretical Linguistics*, 34(3):237–252, 2008.

[11] Danny Fox. Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening. In Ivano Caponigro and Carlo Cecchetto, editors, *From grammar to meaning: the spontenous logicality of language*, pages 201–232. Cambridge University Press, 2012. MS the Hebrew University of Jerusalem.

[12] Benjamin George. Presupposition repairs: a static, trivalent approach to predicting projection. Master's thesis, UCLA, 2008.

[13] Bart Geurts. *Presuppositions and Pronouns*. Elsevier North-Holland, Inc., 1999.

[14] Bart Geurts and Bob van Tiel. When all the five circles are four: new exercises in domain restriction. *Topoi*, pages 1–14, 2015.

[15] Irene Heim. On the projection problem for presuppositions. In Daniel P. Flickinger, editor, *Proceedings of WCCFL 2*, pages 114–125, Stanford University, Stanford, California, 1983. CSLI Publications.

[16] Y. Huang, E. Spelke, and J. Snedeker. What exactly do number words mean? *Language Learning and Development*, 9(2):105–129, 2013.

[17] Linda Kennedy, Cory Bill, Florian Schwarz, Stephen Crain, Raffaella Folli, and Jacopo Romoli. Scalar implicatures vs presuppositions: The view from broca's aphasia. In *Proceedings of NELS 40*, Amherst, MA, 2015. GLSA.

[18] Jacopo Romoli. The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of semantics*, to appear.

[19] Jacopo Romoli and Florian Schwarz. An experimental comparison between presuppositions and indirect scalar implicatures. In Florian Schwarz, editor, *Experimental Perspectives on Presuppositions*, Studies in Theoretical Psycholinguistics. Springer, Dordrecht, 2015.

[20] P. Schlenker. Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics*, 34(3):157–212, 2008.

[21] Philippe Schlenker. Local contexts. *Semantics and Pragmatics*, 2(3):1–78, July 2009.

[22] Florian Schwarz. Introduction: Aspects of meaning in context - theoretical issues and experimental perspectives. In F. Schwarz, editor, *Experimental Perspectives on Presuppositions*. Springer, 2015.

[23] Florian Schwarz. Experimental work in presupposition and presupposition projection. *The Annual Review of Linguistics*, 2, to appear.

[24] Yasutada Sudo, Jacopo Romoli, Danny Fox, and Martin Hackl. Variation of presupposition projection in quantified sentences. In *Proceedings of the Amsterdam Colloquium 2011*, Amsterdam, The Netherlands, to appear.

[25] Sonja Tiemann. *The Processing of wieder ('again') and other Presupposition Triggers*. PhD thesis, Eberhard Karls Universität Tübingen, Tübingen, 2014.

[26] Rob van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377, 1992.

[27] Jérémy Zehr and Florian Schwarz. Experimental evidence for entailed vs non-entailed presuppositions. Talk presented at NELS 46, Concordia University, 2015.

NEG-raising does not involve
syntactic reconstruction

Hedde Zeijlstra

Georg-August-University Göttingen, Germany

`hzeijls@uni-goettingen.de`

Abstract

NEG-raising concerns the phenomenon by which certain negated predicates (e.g. *think*, *believe*, *hope*) can give rise to a reading where the negation seems to take scope from an embedded clause. The standard analysis in pragma-semantic terms goes back to Bartsch (1973) and is elaborated in Gajewski (2005, 2007), Romoli (2013) and others. Recently, this standard approach has been attacked by Collins and Postal (2014), who argue, by providing some novel arguments, that NEG-raising involves syntactic movement of the negation from the embedded clause into the matrix clause. The syntactic structure of 'I don't think you're right' would then be: I do[n't]$_i$ think you're t$_i$ right, and the NEG-raising reading would result from syntactic reconstruction of the negation.

In this talk I present four novel arguments against this account. First, following up work by Horn (2014), I show that Collins and Postal (2014), and their reply to Horn (Collins & Postal 2015), predict that every negated predicate that can license so-called Horn clauses (non-negative clauses containing NPIs and subject–auxiliary inversion) should receive a NEG-raising reading, contrary to fact. Second, Collins and Postal (2014) adopt phonological deletion of negative operators – a necessary ingredient for their account – but this cannot be independently motivated. Third, it turns out that for certain constructions, Collins and Postal (2014) must also allude to the original Bartschian approach. Finally, I demonstrate that the standard approach actually explains the grammaticality of Horn clauses better than the syntactic alternative presented by Collins and Postal (2014).

1   Introduction

NEG-raising concerns the phenomenon, illustrated in (1), by which certain negated predicates (e.g. *think*, *believe*, *hope*) can give rise to a reading where negation seems to take scope from the embedded clause: for instance, (1)a may have a reading (1)b. The standard analysis, which treats NEG-raising in pragma-semantic terms, goes back to Bartsch (1973) and is elaborated in Gajewski (2005, 2007), Romoli (2013), among many others. Under this approach, NEG-raising predicates such as *think* come along with an excluded middle or homogeneity presupposition. The predicate *think p* presupposes that either one thinks that p, or one thinks that not p. Applying this to (1), (1)a presupposes that the speaker either thinks you're right or thinks that you're not right. Together with this presupposition, (1)a entails (1)b.

(1)   a.   I don't think you're right
        b.   I think you're not right

Recently, the standard approach has been attacked by Collins & and Postal (2014, henceforward CP14), who argue that NEG-raising involves syntactic movement of the negation from the embedded clause into the matrix clause (a proposal tracing back to Fillmore 1963). The syntactic structure of (1) would then be as in (2), and the reading (1)b would follow from reconstruction of the negation (<NEG> indicating a lower copy/trace of NEG).

(2)   I do NEG think you're <NEG> right

CP14 provide a series of arguments in favour of the syntactic approach and against the standard pragma-semantic approach, which I will briefly outline in the next section. Then, in section 3, I will discuss several problems concerning the syntactic approach that would rather call for reinstalling the standard, pragma-semantic approach to NEG-raising. In section 4, I furthermore show that, upon closer inspection, the arguments presented in section 2 in favour of the syntactic approach actually involve facts that are better explained by the standard, pragma-semantic approach.

2   Arguments in favour of the syntactic approach

The two most important arguments by CP14 center around the licensing of embedded strict Negative Polarity Items (NPIs) by negated NEG-raising predicates and the possibility of negated NEG-raising predicates to embed so-called Horn clauses. Both arguments indicate that the negation present in a higher clause must have started out in a lower clause.

Strict NPIs, such as *until* or *breathe a word*, differ from other, non-strict NPIs (such as *any* or *ever*) in the sense that the licensing of the former (3)–(4) but not the latter (5) are subject to syntactic locality constraints, such as clause boundedness.

(3)   a.   Carolyn will *(not) breathe a word about it
        b.   *Stanley doesn't predict that Carolyn will breathe a word about it

(4)   a.   Calvin didn't / *moved in until June
        b.   *Calvin didn't claim that Mona moved in until June

(5)   a.   Stanley *(doesn't) predict that Carolyn will say anything about it
        b.   Calvin didn't claim / *claimed that Mona ever moved in

Strikingly, a negated NEG-raising predicate may license embedded strict NPIs, suggesting that the negation must have started out clause-internally to license the NPI before it raises into the matrix clause:

(6)   a.   Stanley doesn't believe that Carolyn will breathe a word about it
        b.   Calvin didn't think that Mona moved in until June

Further evidence for such a raising analysis comes from the fact that once the embedded clause forms a syntactic island, from where extraction is forbidden, this licensing is no longer possible, as is illustrated for topic islands in (7).

(7)   a.   *That Carolyn will breathe a word about it Stanley doesn't believe
        b.   *That Mona moved in until June Calvin didn't think

Perhaps even stronger evidence in favour of a raising analysis comes from Horn clauses. Horn clauses are clauses where an NPI in Spec,CP triggers subject–auxiliary inversion. Such clauses are fine under negated NEG-raising predicates.

(8)   a.   I *(don't) think that ever before have the media played such a major role in a kidnapping
       b.   She *(doesn't) suppose that under any circumstances would he help me

However, subject–auxiliary inversion is generally only allowed by (semi-)negative elements in Spec,CP. The availability of Horn clauses under negated NEG-raising predicates shows that the negation must have started out in the Horn clause itself. For CP14, those sentences have underlying structures with an abstract negation starting out in the embedded clause, with the abstract negation raising into the matrix clause, where it gets phonologically realized as *n't*. If the negation had not raised, it would have been incorporated into the NPI (with the realizations *never before* and *under no circumstances,* respectively).

For CP14, the licensing of strict NPIs and Horn clauses forms strong evidence for a syntactic approach to NEG-raising. And indeed, the existence of such examples has not been explained by any other account to NEG-raising. However, a major challenge for any syntactic approach to NEG-raising comes from NEG-raising readings triggered by negative indefinites. The example in (9) has a reading saying that everybody supposes that nuclear war is not winnable (until next year). The fact that *until next year* can be licensed by *nobody* reveals for CP14 that negation must have started out below as well.

(9)   Nobody supposes that nuclear war is winnable (until next year)

But if *nobody* is the realization of a negated indefinite (NEG ∃), the predicted reading would then be that somebody supposes nuclear war is not winnable (until next year), contrary to fact. Note that this is not a problem for the standard approach to NEG-raising; the excluded middle presupposition plus the assertion jointly entail the attested NEG-raising reading. CP14 acknowledge this fact and to solve this problem claim that constructions like (9) contain in total three negations:

(10) NEG_1 PERSON NEG_2 supposes that nuclear war is NEG_3 winnable (until next year)

Example (10) indeed has the attested NEG-raising reading. For CP14, the lower negation (NEG_3) then raises into the matrix clause NEG_3, using <…> to indicate lower copies of moved elements:

(11) NEG_1 PERSON NEG_2 NEG_3 supposes that nuclear war is <NEG_3> winnable (until next year)

Finally, CP14 postulate a mechanism by which in particular cases two negations in the same clause can be phonologically deleted under a downward entailing operator. Using ~~strikethrough~~ as an indicator for phonological deletion, (11) then becomes (12), which is phonologically realized as (9).

(12) NEG_1 PERSON ~~NEG_2 NEG_3~~ supposes that nuclear war is <NEG_3> winnable (until next year)

Naturally, the most stipulated step here is the presence of two negations that are phonologically zero. However, for CP14 this step can be motivated because (i) one negation must have started out below (given the licensing of Horn clauses and strict NPIs) and (ii), as they claim, there is

independent evidence for unpronounced negations.

3 Problems for CP14

CP14's proposal is an important contribution to the understanding of NEG-raising, but also faces several challenges. First, as pointed out by Horn (2014), it is not the case that only NEG-raising predicates can license Horn clauses; other negated predicates can do so as well, even though they do not trigger NEG-raising readings. In a reply, Collins and Postal (2015, CP15 henceforward) argue that these cases can be accounted for in a different way, but as I will show below (3.1), this alternative account suffers from the same problem as the original account. Second, it turns out that the proposed independent motivation for phonologically deleted negations is incorrect (3.2). Third, not every instance of NEG-raising can follow from the suggested reconstruction mechanism (3.3).

3.1 Horn clauses and Cloud-of-Unknowing predicates

Horn (2014) observed that not every negated predicate that licenses Horn clauses also triggers NEG-raising readings. He presents examples of non-NEG-raising predicates – dubbed Cloud-of-Unknowing predicates – that can also license Horn clauses, such as non-factive *know* and other predicates expressing particular subject or speaker knowledge. Horn's example is presented below in (13); (13) is another example.

(13)   a.   I *(don't) know that ever before had all three boys napped simultaneously
         b.   She's *(not) convinced that ever before had all three boys napped simultaneously

However, the examples in (13) clearly lack a NEG-raising reading. They are not equivalent to their counterparts in (14).

(14)   a.   I know that never before had all three boys napped simultaneously
         b.   She's convinced that never before had all three boys napped simultaneously

To solve these problems, CP15 reply to Horn (2014) by arguing that examples such as (13) again contain two phonologically unrealized negations: (13)a would underlyingly be like (15),a not (15)b. Cloud-of-Unknowing predicates then form another context under which phonological deletion of two negations may take place.

(15)   a.   [I do NEG_1 know ~~NEG_2~~ [<NEG_2> that ~~NEG_3~~ ever before had all three boys napped
                simultaneously]]
         b.   [I do NEG_1 know [that <NEG_1> ever before had all three boys napped
                simultaneously]]

In (15)a, NEG_2 starts out in the embedded clause, licensing subject–auxiliary inversion, and licensing phonological deletion of NEG_3. It then raises into the matrix clause to be phonologically deleted under NEG_1, just as was the case in the constructions involving negative indefinites. Evidence for this instance of raising comes again from island effects. If the Horn clause is an island, such sentences become ill formed.

(16)   a.   *That ever before had all three boys napped simultaneously, I don't know
         b.   *That ever before had all three boys napped simultaneously, she's not convinced

CP15 are correct that in their system (15)a is an alternative solution. However, what is problematic is that if raising a negation (NEG$_3$) out of a Horn clause into a matrix clause is possible, nothing rules out (15)b as an additional underlying structure. But since (15)b is the structure that gives rise to the NEG-raising reading, it is predicted that the sentences in (13) should exhibit the corresponding NEG-raising readings in (14) as well, contrary to fact. CP15 argue that one can rule this out by stipulating that if a negation raises into a clause containing a negated Cloud-of-Unknowing predicate, this predicate must be under the scope of a distinct negation. For them, this can be motivated on the basis of the fact that a negated Cloud-of-Unknowing predicate itself forms a semantic constituent. But that is circular: the problem is why it would not be possible to have a negation raised into such a main clause with which it does not form a semantic constituent (in CP14/15 terms). The *explanans* here is thus the *explanandum*. Hence, the solution CP15 provide suffers from the same problem (an unattested predicted NEG-raising reading) as their original proposal, unless such readings are ruled out by pure ill-motivated brute force.

3.2   Phonologically deleted negations

To defend their proposal that semantically present negations can be phonologically deleted in certain contexts, CP14 present several other cases of alleged phonological deletion of semantic negations so that the proposal can be independently motivated. The most important examples are negated modals in French and optionally negative minimizers in German.

As for the first, French has an expletive marker *ne* that in principle requires co-occurrence of an additional negation (usually *pas* or a negative indefinite):

(17)   Marie ne mange *(pas / rien)
         Marie neg eats neg / nothing
         'Marie doesn't eat / Marie doesn't eat anything'

However, when combined with a few particular modals, such as *pouvoir* 'must' or *savoir* 'know', or the verb *cesser* ('stop'), *ne* suffices to express negation, illustrated for *pouvoir* below.

(18)   Je ne peut (pas)
         I neg can neg
         'I can't'

CP14 take this to be evidence for the deletion of a semantic negation. For them, the examples like (18) contain a semantic negation that has been deleted. However, as such examples are restricted to only a handful of modals, they have traditionally been analysed as remnants of previous stages of the languages that have fossilized into idiomatic expressions. As known at least since Jespersen (1917), Old French lacked the negative marker *pas* and only used the preverbal negative marker *ne* to express negation. Hence, it could very well be the case that expressions like (18) merely reflect Old French negation and should be thought of as idiomatic expressions (see Haegeman 1995, Zeijlstra 2004 for an overview and discussion of such facts). The existence of such an alternative analysis means that these examples do not form any hard evidence for the presence of phonologically deleted negations. At the same time, it must be acknowledged that the alternative analysis has also not received proper evidence; hence, it has not been shown that these French examples lack a phonologically deleted negation.

This is, however, different for the second kind of example that CP14 provide. Here it can actually be shown that they contain no phonologically deleted negation. The examples concern particular German pejorative NPIs. As Sailer (2006) observes, for many (though not all) German speakers, the

following sentences containing a minimizer (and similar pairs of sentences with other pejorative minimizers) have the same meaning (cf. Sailer 2006).

(19) a. Das interessiert mich einen Dreck
     That interests me a dirt
     'I'm not interested at all'
   b. Das interessiert mich keinen Dreck
     That interests me no dirt
     'I'm not interested at all'

For CP14, the fact that the sentences with and without negation have the same meaning is evidence that (19)a, which lacks an overt negation, must contain a covert negation. However, the semantic or pragmatic similarity of the two readings need not follow from the postulation of a covert negation in (19)a. It is possible that the two sentences have different readings whose usage conditions are more or less identical. If the reading of (19)a is that the degree of interest of the speaker is extremely low – even lower than some contextual threshold that indicates a minimal degree of interest – (19) expresses that the speaker's interest lies below this threshold. That means that (19)a expresses that the speaker has no contextually salient degree of interest.

However, if that is the case, (19)a can be uttered in exactly the same situations where the speaker expresses no degree of interest at all by uttering (19)b. Hence, the similarity of the readings in the minimal pair in (19) can be explained without postulating any covert negation. Interestingly, the two analyses make different predictions. For the alternative analysis, (19)a is a positive sentence and (19)b a negative sentence. Under CP14's proposal, both are negative sentences.

Sentential negation can be diagnosed in German by *auch (nicht)* ('also (not)') continuations. In German, positive clauses can be continued by *auch*; they cannot be continued by *auch nicht*. Negative clauses, on the other hand, trigger *auch nicht* continuations and only marginally allow *auch* continuations:

(20) a. Hans geht und Marie auch (*nicht)
     Hans goes and Marie also not
     'Hans goes and Marie does too'
   b. Hans geht nicht und Marie auch $^{??}$(nicht)
     Hans goes not and Marie also not
     'Hans doesn't' go and Marie doesn't either'

Exactly the same pattern can be observed for (19), as shown below. Hence, the test shows that (19)b carries a semantic negation but (19)a does not, disproving CP14's covert negation analysis. Independent evidence for the covert negations in the examples involving NEG-raising is thus absent as well.[1]

(21) a. Das interessiert mich einen Dreck, und ihn auch (*nicht)
     That interests me a dirt, and him also not
     'I'm not interested at all, and he is neither'
   b. Das interessiert mich keinen Dreck, und ihn auch $^{??}$(nicht)

---

[1] Note that this does not mean that phonologically covert negations cannot exist. For instance, in Ladusaw (1992), Zeijlstra (2008) and others, it is argued that negative indefinites in Negative Concord languages (so-called n-words) may enter a syntactic agree relation with a possibly covert negation. There, covert negations are syntactically licensed by an agreeing overt negative element. This licensing mechanism, however, does not extend to the kind of examples presented and discussed in CP14.

That interests me no dirt, and him also not
'I'm not interested at all, and he is neither'

3.3 Islands and NEG-raising

A third problem for CP14 concerns island effects. CP14 take NEG-raising to involve syntactic movement out of a lower clause into a higher clause. Evidence for that view comes from cases in which a Horn clause or a strict NPI is licensed by a clause-external, negated NEG-raising predicate. If such Horn clauses or strict NPIs are in an island (or form an island themselves), this movement is blocked and such licensing is no longer possible, as is exemplified for strict NPIs in (22) and for Horn clauses in (23):

(22)   *That Carolyn will breathe a word about it Stanley doesn't think

(23)   *That ever before had all three boys napped simultaneously, I don't believe

But if that is correct, and NEG-raising is indeed the result of syntactic movement, NEG-raising readings should not be allowed when the clause in which the negation appears to be interpreted forms an island. However, this prediction is not borne out. Examples (22) and (23) can easily give rise to a NEG-raising reading if the strict NPI is absent and subject–auxiliary inversion does not take place, as illustrated in (24)–(25):

(24)   That Carolyn will breathe Stanley doesn't think
        'Stanley doesn't think that Carolyn won't breathe'

(25)   That all three boys napped simultaneously I don't believe
        'I believe that all three boys didn't nap simultaneously'

But where does the NEG-raising reading come from? Clearly, it cannot be the case that the negation emerged in the embedded clause – otherwise the raising of a negation in (22) and (23) would not be problematic either. The only way to account for NEG-raising readings in (24) and (25) is by alluding to some pragma-semantic mechanism along the lines of Bartsch (1973) and her successors. However, the syntactic approach would then no longer be an alternative to the pragma-semantic approach, but rather an account that is at best co-existent with it.

4 Reinstalling the pragma-semantic approach

So, where do we stand? CP14's approach faces at least three serious problems: It predicts NEG-raising readings to be possible in cases where they are not attested, and it predicts NEG-raising readings to be impossible in cases where they are actually found. Moreover, the treatment of NEG-raising readings invoked by negative indefinites can only be maintained by making very specific assumptions, which on closer inspection turn out not to be independently motivated. However, CP14 can straightforwardly account for the fact that strict NPIs and Horn clauses can be licensed by higher negated NEG-raising predicates. Hence, to reinstall the pragma-semantic approach, it must be shown that this approach can also capture the observed facts concerning strict NPIs and Horn clauses. This may not be straightforwardly the case, as pragma-semantic approaches to NEG-raising do not take the negation to start out in the lower clause and instead have it reconstructed at a later stage.

That does not mean, however, that the pragma-semantic approach is incompatible with the negation starting out in a lower position. The central claim of the pragma-semantic approach is that in

NEG-raising readings the negation is interpreted in its surface position. CP14's central claim is that the negation starts out in a lower clause and is interpreted there. But that is actually a twofold claim: one claim saying that the negation starts out below, and one claim asserting that negation is interpreted in this lower position. However, CP14 only provide evidence for the first claim.

Hence, it is possible to reconcile CP14's observations with the pragma-semantic approach to NEG-raising, as it is a logical possibility that in particular cases negation starts out below, raises into the higher clause, and is interpreted there, with the excluded middle or homogeneity presupposition of the NEG-raising predicate triggering an additional inference that together with the assertion yields the NEG-raising reading. Under such an account, it is possible to derive the NEG-raising readings of sentences that contain lower strict NPIs or Horn clauses. The examples in (26) have syntactic structures as in (27), with <NEG> again indicating the basis position of NEG. The strict NPI and the subject–auxiliary inversion are licensed by the lower negation before it raises into the higher position, where it will be interpreted.

(26)   a.   Stanley doesn't believe that Carolyn will breathe a word about it
        b.   I don't think that ever before have the media played such a major role in a
              kidnapping

(27)   a.   Stanley does NEG believe that Carolyn will <NEG> breathe a word about it
        b.   I do NEG think that <NEG> ever before have the media played such a major role in
              a kidnapping

Such an analysis makes all the correct predictions discussed in this paper. First, it can account for the relevant aspects concerning the distribution of strict NPIs, including their island sensitivity: if the embedded clause is an island, the negation can never move into the matrix clause. Furthermore, the existence of NEG-raising readings involving island clauses (section 3.3) naturally follows. In (24), repeated below, there is no movement going on, but the assertion and the presupposition together still trigger the NEG-raising reading. Because NEG-raising does not involve any kind of syntactic reconstruction, movement of negation is not a prerequisite for NEG-raising readings.

(28)   That Carolyn will breathe Stanley doesn't think
        'Stanley thinks that Carolyn won't breathe'

In fact, following standard minimalist ideas on syntactic movement (cf. Chomsky 1995), movement takes place only when it is necessary. That is indeed the case in the examples in (26), but in other examples – for instance (1)a, repeated below – no negative movement has been going on. The surface position of negation is also its base position here.

(29)   I don't think you're right

Adopting this version of the pragma-semantic approach to NEG-raising also avoids alluding to deleted double negations. In examples such as ((30), the universal NEG-raising follows immediately. Negation is simply interpreted in its surface position (which is also its base position), and the excluded middle or homogeneity presupposition does the rest. Note also that in (31)a, where the negation must have started out below to license *until next year*, it raises into the position where it is interpreted together with the existential realized as *nobody*. Consequently, no phonologically deleted double negations are needed.

(30)   Nobody supposes that nuclear war is winnable

(31) a.   Nobody supposes that nuclear war is winnable until next year
      b.   [[NEG ∃] supposes [that nuclear war is <NEG> winnable until next year]]

Finally, the facts concerning Cloud-of-Unknowing predicates follow. What NEG-raising predicates – at least the ones discussed thus far – and Cloud-of-Unknowing predicates share is that they do not impose strict locality conditions on their embedded clauses; other predicates, such as *say* or *claim*, do. Therefore, NEG-raising predicates and Cloud-of-Unknowing predicates can license embedded strict NPIs and Horn clauses. As the other predicates impose stronger locality conditions, negation cannot move out of them and therefore they also cannot license embedded strict NPIs and Horn clauses. However, Cloud-of-Unknowing and NEG-raising predicates differ with respect to the excluded middle or homogeneity presupposition: NEG-raising predicates have it; Cloud-of-Unknowing predicates do not. Hence, the latter class of predicates does not trigger NEG-raising readings.

One may wonder, then, why NEG-raising predicates have two distinguishing properties: weak locality conditions imposed on their complement clauses and the excluded middle or homogeneity presupposition. Ideally, NEG-raising predicates should follow from one distinguishing property only. There is no reason why predicates with an excluded middle or homogeneity presupposition should also impose weak locality conditions on their complement clauses to yield a NEG-raising reading.

Indeed, there are predicates that only have this excluded middle or homogeneity presupposition. *To be of the opinion* is a good example. Example (32)a clearly has a reading (32)b. However, it cannot license strict NPIs or Horn clauses.

(32) a.   I am not of the opinion you're right
      b.   I am of the opinion you're not right

(33) a.   *I am not of the opinion that Carolyn will breathe a word about it
      b.   *I am not of the opinion that ever before have the media played such a major role in a kidnapping.

Hence, whether predicates impose weak and strong locality conditions or their complements, and whether predicates come with an excluded middle or homogeneity presupposition, are independent parameters, as demonstrated in the following table:

(34) Four types of predicates

	Weak locality constraints	Strong locality constraints
Excluded middle or homogeneity presupposition present	*think, believe, hope*	*to be of the opinion*
Excluded middle or homogeneity presupposition absent	*know, to be convinced (Cloud-of-Unknowing predicates)*	*say, claim, predict*

It may be striking, though, that the large majority of predicates that come with the excluded middle or homogeneity presupposition also impose weak locality constraints on their complements. However, this may very well be because most predicates in this class are also non-factive, and it has been claimed in syntactic theory that non-factive predicates often impose weaker constraints on extraction from their complements than factive predicates do (cf. Giorgi 2004 a.o.) and both NEG-raising and Cloud-of-Unknowing predicates are generally non-factive.

Hence, the proposed alternative in this paper indeed reinstalls the pragma-semantic approach to

NEG-raising while still being able to account for the relevant distribution of strict NPIs and Horn clauses. However, a possible objection may arise in the sense that in examples such as (26), the negation lacks any phonological or semantic effect in its base position; the lowest trace or copy appears to be semantically vacuous. There are three reasons why this objection does not hold. First, there is nothing in the theory maintaining that lowest traces/copies cannot be semantically vacuous. Elements are interpreted at their position at LF, not in their base position. Hence, nothing in the theory forbids negation to start out low and be interpreted high. But even if one were to maintain this position, note that the lowest trace/copy does have some grammatical effect: it is responsible for the licensing of the NPI. Hence, movement of negation is not a grammatically redundant operation. And if ungrammaticality is determined at the interfaces only (cf. Chomsky 1995), these lowest traces/copies of the negation indeed have a phonological or semantic effect. Finally, it should be noted that CP14 must also assume that sometimes the role of the lowest negation is just to license the NPI and nothing more. After all, every negated NEG-raising predicate also allows for a non-NEG-raising reading (admittedly, something not very well understood in the pragma-semantic approach). This is also the case in (35).

(35)   Mary didn't believe that Bill stayed until June; she simply doesn't have any beliefs about Bill.

Example (35) lacks the NEG-raising reading. Under CP14, the negation must still have started out below (in order to license *until June*), but the negation is only interpreted in its highest position.

References

Bartsch, R. (1973). ''Negative transportation' gibt es nicht.' In *Linguistische Berichte 27*: 1-7.
Chomsky, N. (1995). The Minimalist Program. Cambridge, MA: The MIT Press.
Collins, C. & P. Postal (2014). *Classical NEG Raising*. Cambridge, MA: The MIT Press.
Collins & Postal (2015). 'Dispelling he Cloud of Unknowing.' Ms., NYU. LingBuzz/002269.
Fillmore, C. (1963). 'The position of embedding transformations in a grammar.' In *Word 19*: 208-231.
Gajewski, J. (2005). *Neg-raising: polarity and presupposition*. PhD Dissertation, MIT.
Gajewski, J. (2007). 'Neg-raising and polarity.' In *Linguistics and Philosophy 30*: 298-328.
Romoli, J. (2013). A scalar implicature-based approach to neg-raising. In *Linguistics and Philosophy 36*: 291-353.
Giorgi, A. (2004). 'From temporal anchoring to long distance anaphor binding.' Paper presented at the 23rd West Coast Conference in Formal Linguistics. April 23-25, University of California at Davis, Davis CA.
Ladusaw 1992. 'Expressing negation.' In C. Barker & D. Dowty (eds.), *Proceedings of SALT 2*. Columbus: The Ohio State University. 237-259.
Jespersen, O. (1917). *Negation in English*. Copenhagen: A.F. Høst.
Haegeman, L. (1995) *The syntax of negation*. Cambridge: Cambridge University Press.
Horn, L. (2014). 'The cloud of unknowing.' In J. Hoeksema & D. Gilbers (eds*.), Black Book: a festschrift in honor of Frans Zwarts*. Groningen: University of Groningen. 178-96.
Sailer, M. (2006). 'Don't believe in underspecified semantics: Neg raising in lexical resource semantics.' In *Empirical issues in syntax and semantics 6:* 375-403.
Zeijlstra, H. (2004). *Sentential negation and Negative Concord*. PhD Dissertation, University of Amsterdam.
Zeijlstra, H. (2008). 'On the syntactic flexibility of formal features.' In: Th. Biberauer (ed.) *The limits of syntactic variation*. Amsterdam: John Benjamins. 143-173.

Toward Probabilistic Natural Logic for Syllogistic Reasoning

Fangzhou Zhai, Jakub Szymanik and Ivan Titov

Institute for Logic, Language and Computation, University of Amsterdam

Abstract

Natural language contains an abundance of reasoning patterns. Historically, there have been many attempts to capture their rational usage in normative systems of logical rules. However, empirical studies have repeatedly shown that human inference differs from what is characterized by logical validity. In order to better characterize the patterns of human reasoning, psychologists have proposed a number of theories of reasoning. In this paper, we combine logical and psychological perspectives on human reasoning. We develop a framework integrating Natural Logic and Mental Logic traditions. We model inference as a stochastic process where the reasoner arrives at a conclusion following a sequence of applications of inference steps (both logical rules and heuristic guesses). We estimate our model (i.e. assign weights to all possible inference rules) on a dataset of human syllogistic inference while treating the derivations as latent variables in our model. The computational model is accurate in predicting human conclusions on unseen test data (95% correct predictions) and outperforms other previous theories. We further discuss the psychological plausibility of the model and the possibilities of extending the model to cover larger fragments of natural language.

1 Introduction

1.1 Syllogistic Reasoning

The psychology of reasoning tries to answer one fundamental question: how do people reason? This question is also central for many other scientific disciplines, from linguistics and economics to cognitive science and artificial intelligence[1]. Logic was first to study reasoning systematically and Aristotle proposed the syllogistic theory as an attempt to normatively characterize rationality. And even though modern logic has developed intricate theories of many fragment of natural language, the syllogistic fragment continuously receives attention from researchers (see [11] for a review of the theories of syllogisms). The sentences of syllogisms are of four different sentence types (or 'moods'), namely:

$$
\begin{aligned}
&\text{All } A \text{ are } B : \text{universal affirmative (A)} \\
&\text{Some } A \text{ are } B : \text{particular affirmative (I)} \\
&\text{No } A \text{ are } B : \text{universal negative (E)} \\
&\text{Some } A \text{ are not } B : \text{particular negative(O)}
\end{aligned}
$$

Each syllogism has two sentences as the premises, and one as the conclusion. Traditionally, according to the arrangements of the terms in the premises, syllogisms are classified in to four categories, or 'figures':

---

[1]See, e.g., [8] for a survey of logic and cognitive science.

	Figure 1	Figure 2	Figure 3	Figure 4
	BC	CB	BC	CB
	AB	AB	BA	BA
	AC	AC	AC	AC

Syllogisms are customarily identified by their sentence types and figures. For example, 'AI3E' refers to the syllogism whose premises are of sentence types A and I, and whose terms are arranged according to figure 3, and whose conclusion is of type E. Therefore, altogether, 'AI3E' refers to the following syllogism:

$$\text{All } B \text{ are } C$$
$$\text{Some } B \text{ are } A$$

$$\text{No } A \text{ are } C$$

As there are four different sentence types and four different figures, there are 256 equivalent syllogisms in total. These syllogisms are also referred to as the ones that follows the scholastic order. Of those 24 are valid according to the semantics of traditional syllogistic logic, and 15 of these 24 are valid according to the semantics of modern predicate logic.

Psychologists have developed a battery of experimental tests to study human syllogistic reasoning. In one typical experimental design, the reasoners are presented with the premises and asked 'What follows necessarily from the premises?'. Chater and Oaksford [3] have compared five experimental studies of this sort and computed the weighted average of the data, that is the percentage that each conclusion was drawn. The data is shown in Table 1.

One important observation made by Chater and Oaksford [3] is that logical validity seems to be a crucial factor for an explanation of the participants' performance. Firstly, the average percentage of reasoners arriving at a valid conclusion is 51%, while that of arriving at an invalid conclusion is 11%: it seems that participants indeed made an effort along the path of validity. Secondly, reasoners tends to mistakenly arrive at invalid syllogisms that are different from valid ones just by their figures. For example, the AO2O syllogism is the only valid one among the four AOO syllogisms, however, reasoners endorse the other three AOO syllogisms (namely AO1O, AO3O and AO4O) with fairly high probability. This might be a sign that people are actually not that bad at logic (see, e.g., [6]): even if an error is made, the most probable wrongly endorsed syllogism is quite similar to a valid one, which differs only in the figure. Thirdly, the mean entropy of the syllogistic premises that yields at least one valid conclusion, according to the table above, is 0.729, however, that of the ones that yield no valid syllogisms is 0.921. The difference indicates that the psychological procedures triggered by the two groups of premises are likely to be different.

1.2 Mental Logic

Rips [13] has proposed a theory of quantified reasoning based on formal inference rules. The underlying psychological assumption is that logical formulas can be used as the mental representations of reasoning steps and that the inference rules are the basic reasoning operations of the mind. Rips has argued that, deductive reasoning, as a psychological procedure, is the generation of 'a set of sentences linking the premises to the conclusion', and 'each link is the embodiment of an inference rule that reasoners consider intuitively sound'. He has formulated a set of rules that includes both sentential connectives and quantifiers and implemented such system as a computational mode PSYCOP.

Syllogism	Conclusion					Syllogism	Conclusion				
	A	I	E	O	NVC		A	I	E	O	NVC
AA1	**90**	**5**	0	0	5	AO1	1	6	1	57	35
AA2	58	8	1	1	32	AO2	0	6	3	**67**	24
AA3	57	**29**	0	0	14	AO3	0	10	0	66	24
AA4	75	**16**	1	1	7	AO4	0	5	3	72	20
AI1	0	92	3	3	2	OA1	0	3	3	68	26
AI2	0	57	3	11	29	OA2	0	11	5	56	28
AI3	1	**89**	1	3	7	OA3	0	15	3	**69**	13
AI4	0	71	0	1	28	OA4	1	3	6	27	63
IA1	0	72	0	6	22	II1	0	41	3	4	52
IA2	13	49	3	12	23	II2	1	42	3	3	51
IA3	2	**85**	1	4	8	II3	0	24	3	1	72
IA4	0	**91**	1	1	7	II4	0	42	0	1	57
AE1	0	3	59	6	32	IE1	1	1	22	16	60
AE2	0	0	88	**1**	11	IE2	0	0	39	30	31
AE3	0	1	61	13	25	IE3	0	1	30	33	36
AE4	0	3	**87**	**2**	8	IE4	0	42	0	1	57
EA1	0	1	**87**	**3**	9	EI1	0	5	15	**66**	14
EA2	0	0	**89**	**3**	8	EI2	1	1	21	**52**	25
EA3	0	0	64	**22**	14	EI3	0	6	15	**48**	31
EA4	1	3	61	**8**	28	EI4	0	2	32	**27**	39
OE1	1	0	14	5	80	OO1	1	8	1	12	78
OE2	0	8	11	16	65	OO2	0	16	5	10	69
OE3	0	5	12	18	65	OO3	1	6	0	15	78
OE4	0	19	9	14	58	OO4	1	4	1	25	69
IO1	3	4	1	30	62	OI1	4	6	0	35	55
IO2	1	5	4	37	53	OI2	0	8	3	35	54
IO3	0	9	1	29	61	OI3	1	9	1	31	58
IO4	0	5	1	44	50	OI4	3	8	2	29	58
EE1	0	1	34	1	64	EO1	1	8	8	23	60
EE2	3	3	14	3	77	EO2	0	13	7	11	69
EE3	0	0	18	3	78	EO3	0	0	9	28	63
EE4	0	3	31	1	65	EO4	0	5	8	12	75

Table 1: Percentage of times each syllogistic conclusions was endorsed. The data is from a meta-analysis in [3]. 'NVC' stands for 'No Valid Conclusion', all numbers have been rounded to the closest integer. A bold number indicates that the corresponding conclusion is logically valid.

The reasoner modeled by the theory derives only but not all logically valid conclusions (i.e., it is logically sound but not complete). It puts constraints on the application of the inference rules to deal away with logical omniscience: certain logical truths are not derivable in Mental Logic theory. Instead of accepting standard proof-theoretical system, Rips has selected the inference rules that seem psychologically 'primitive', even if derivable from other rules. Nevertheless, the model still uses arbitrary abstract rules and formal representations (roughly corresponding to the natural deduction system for first-order logic). Moreover, the model, by its mere design, cannot explain reasoning mistakes (see also [9]).

1.3 Natural Logic

Due to psychological, computational and linguistic influences, some of the normative inference rules have been adapted to natural language as a part of so-called Natural Logic Program [15, 2]. Contrary to the Mental Logic of Rips, the Natural Logics identify valid inferences by their lexical and syntactic features, without requiring a full semantic interpretation. For example, some natural language quantifiers are upward monotone in their first argument, like the quantifier 'some'. It means that the inference from 'Some pines are green' to 'Some plants are green' is valid since all pines are plants. The 'pines' can be actually replaced by any object that contains all pines. People can reason based on monotonicity even when the underlying meaning of terms is unclear for them. For example, from 'Every Dachong has nine beautiful tails' people would infer 'Every Dachong has nine tails', without knowing the meaning of 'Dachong' (which simple means 'tiger' in Chinese). In a way, monotonicity operates on the surface of natural language.

Using ideas from Natural Logic, Geurts [6] has designed a proof system for syllogistic reasoning that pivots on the notion of monotonicity. Geurts' proof system for syllogistic reasoning consists of the following set of rules R:

- *All-Some*: 'All A are B' implies 'Some A are B'.

- *No-Some_not*: 'No A are B' implies 'Some A are not B'.

- *Conversion*: 'Some A are B' implies 'Some B are A'; 'No A are B' implies 'No B are A'.

- *Monotonicity*: If A entails B, then the A in any upward entailing position can be substituted by a B, and the B in any downward entailing position can be substituted by an A.

Geurts has further enriched the proof system with difficulty weights assigned to each inference rules to evaluate the difficulty of valid syllogistic reasoning. Geurts assumed that different rules cost different amount of cognitive resources. He gives each reasoner an initial budget of 100 units; each use of the monotonicity rule costs 20 units; a proof containing a 'Some Not' proposition costs an additional 10 units. Taking the remaining budget as an evaluation of the difficulty of each syllogism, the evaluation system fits the experimental data from [3] well. However, the system cannot make any evaluation on most invalid syllogisms, hence cannot explain why reasoners can possibly arrive at invalid conclusions.[2]

2 Data-driven Probabilistic Natural Logic for Syllogistics

2.1 Approach

In this paper we design and estimate a computational model for syllogistic reasoning based on a probabilistic natural logic.[3] This can be treated as a first step to integrate the Mental Logic approach and the Natural Logic approach. It improves upon Mental Logic approach by substituting formal abstract inference rules with Natural Logic operating on the surface structure of Natural Language. That means, the mental representations are given directly as

---

[2]This is not a criticism of [6]. According to Geurts, the system was never intended to give a 'full-blown account of syllogistic reasoning' in the first place, see also [11].

[3]Compare with [5], where the authors designed probabilistic semantic automata for quantifiers whose parameters are also determined by the experimental data.
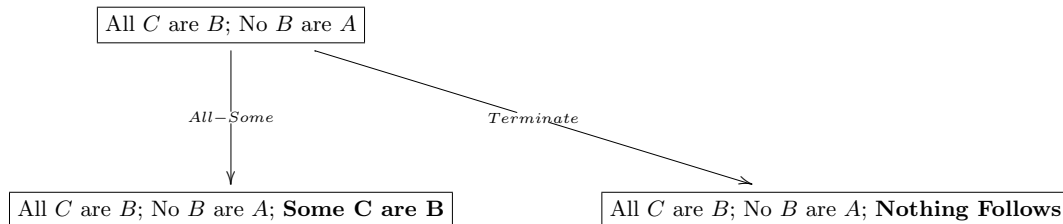
Figure 1: The Mental Representations

natural language sentences, without an intermediate layer of an abstract formal language. Our starting point is the logic developed by Geurts in [6] (see Section 1.3).

We assume that the procedure of reasoning consists of two types of mental events: the inferences made by the reasoners, which are deliberate and precise, and the guesses, which could be less reliable but fast. Accordingly, the model consists of two parts: the inference part, which takes the form of a probabilistic natural logic (i.e., the inference rules are weighted with probabilities) and the guessing part, which leads the reasoner to a possible conclusion in one step depending on a few heuristics. We implemented the model, and estimated it on the experimental data. The model is accurate at predicting human conclusions on unseen syllogisms (including mistakes) and the results yield interesting psychological implications.

2.2 Mental Representation

Similar to Rips' [13] proposal, we take the set of syllogistic sentences as the mental representation of reasoning. Namely, the reasoner maintains a set of sentences in the working memory to represent the state of reasoning, or more specifically, the reasoner keeps a record of the sentences that he considers true at the moment. We will refer to each representation as a **state**. Reasoning operations change the mental states. When performing reasoning, the reasoner generates a sequence of states in the working memory, where the initial state is the set of premises, and the final state contains the conclusion. These states are linked by the **reasoning events**, which can be a specific adoption of an inference rule. For example, given the 'AE4' premises, if the reasoner adopt the 'All - Some' rule (i.e., 'All A are B' implies 'Some A are B') on the premise 'All C are B', a 'Some C are B' will be obtained, possibly as a conclusion. The reasoner may also terminate the reasoning and decide that 'nothing follows', see Figure 1.

We would like to point out here that mental states may not be logically consistent. There are many reasons for this assumption. For example, people tend to adopt illicit conversions which often lead to the inconsistency. After all, people do often make mistakes resulting in conclusions that are inconsistent with assumptions, even while reasoning in a conscious, deliberate way (see, e.g., [10]).

2.3 Statistical Model of Reasoning Procedure

We formulate a generative probabilistic model of reasoning. First, reasoners conduct formal inferences, adopting possible logical rules with different probabilities (related to the cognitive

difficulty of the rule or some sort of reasoning preference). Each inference rule, $r \in R$ is adopted with a different probability specified by the associated weight w_r (*a tendency parameter*) which is estimated from the data. Formally, a probability of transitioning from state S to state S_r using a specific application of rule r is given by:

$$p(r|S, \mathbf{w}) = \frac{w_r}{w_G + \sum_{r' \in R} c_{r'} \cdot w_{r'}},$$

where $c_{r'}$ is the number of different possibilities how the rule r' can be adopted in the given state S and $\mathbf{w}$ is the vector of all tendency parameters. The parameter w_G reserves probability mass for 'terminating' the inference at state S and making a heuristic guess.

The reason to turn to the guessing scenario may have to do with the complexity of inference or the reasoner doubting the conclusion that was already obtained. When the reasoner enters the guessing scenario, the probability that the reasoner guesses 'nothing follows' is negatively correlated with the **informativeness** level (see [4]) of the premises, i.e., the amount of information that the premises carries: the more informative the premise, the less faith the reasoner have for a 'nothing follows' conclusion. The reasoner chooses the remaining options with probabilities determined according to the atmosphere hypothesis. This hypothesis proposes that a conclusion should fit the premises' 'atmosphere', namely, the sentence types of the premises [1]. In particular, whenever at least one premise is negative, the most likely conclusion should be negative; whenever at least one premise contains 'some', the most likely conclusion should contain 'some' as well; otherwise the conclusion are likely to be affirmative and universal. Formally, the probability that the reasoner will switch to the guessing model is given by:

$$\frac{w_G}{w_G + \sum_{r' \in R} c_{r'} \cdot w_{r'}}$$

There are five possible outcomes of the guessing scenario: the subject could guess any conclusion, or could decide that nothing follows from the premises. The probability of nothing follows, given that the guessing scenario is chosen on the previous step, is computed as

$$\frac{v_{dl}}{3 \cdot v_{nd} + v_{as} + v_{dl}},$$

where $v_{dl} = \frac{1}{u_{t_1} + u_{t_2}}$ quantifies doubts of the reasoner that any valid conclusion can be derived from the premises. The quantity v_{dl} is computed relying on the amount of informativeness of both premise sentences (see [4]), the informativeness parameters u_t are estimated from the data and depend on the type t of a sentence (A, I, E or O). In the above expression, t_1 and t_2 refer to the types of sentences in the premises. The probability of guessing the conclusion predicted by the atmosphere hypothesis is:

$$\frac{v_{as}}{3 \cdot v_{nd} + v_{as} + v_{dl}},$$

where v_{as} is the weight assigned to the atmospheric hypothesis (also estimated from the data). Finally, the probability of guessing any of the remaining three options is

$$\frac{v_{nd}}{3 \cdot v_{nd} + v_{as} + v_{dl}},$$

where v_{nd} is a model parameter.[4]

---

[4]Without loss of generality, we set it to 1 as the model is over-parameterized.

Predictions \ Exp. Data	< 30%	≥ 30%
< 30%	Correct Rejection	Miss
≥ 30%	False Alarm	Hit

Table 2: Break-down of Predictions

The probability that a subject could arrive at a particular syllogistic conclusion is estimated from the tree by summing over all the leaf nodes containing the conclusion. Consequently, we can obtain posterior distribution of conclusions given the premises. These posterior distributions (for each premises) can be treated as model predictions, and we evaluate them (on unseen test set) against the distribution of human conclusions.

2.4 Estimation

We use the data from the meta-analysis by Chater and Oaksford [3], as is shown in Table 1. We denote the dataset as $\{X_i, y_i\}_{i \leq n}$, where X_i stands for the pair of premises and y_i stands for the conclusion. We randomly select 50% of the premises (i.e., half of the dataset) and use the corresponding examples as the training data. The rest of the data is used for evaluation. We use maximum likelihood estimation to obtain the parameter values. As our derivations are latent, there is no closed form solution for the optimization problem. Instead we use a variant of the Expectation Maximization (EM) algorithm which starts with a randomly initialized model and alternates between predicting derivations according to the current model (E-step) and updating model parameters based on these predictions (M-step, maximization of the expected likelihood). In our approach, the set of potentially applicable rules is determined by the reasoner state and, consequently, this set is not constant across the states (as discussed above, c_r was dependent on the state S). This implies that, unlike standard applications of EM, there is no closed-form solution for the M-step of the algorithm. Instead we use so-called generalized EM: instead of finding a maximum of the expected likelihood at M-step, we perform just one step of stochastic gradient ascent.

3 Results and Discussion

3.1 Evaluation

We use a mixed means of evaluation. We mainly use the evaluation method proposed in [11], which is based on the signal detection theory. The authors assume that the conclusions of the participants are noisy, that is unsystematic errors occur frequently. Hence, they classify the experimental data into two categories: those conclusions that appear reliably more often than chance level, which a theory of the syllogisms should predict to occur; and those that do not occur reliably more than chance level, which a theory should predict will not occur. In our context, there are five possible conclusions that can be drawn by subject. The chance level is thus 20%. In the following, we count a conclusion as reliable if it is drawn significantly often, i.e., in at least 30% of the trials.[5] As far as a theory predicts what will be concluded from each pair of premises, the method can be applied to evaluate the theory. According to the type of fitting, the predictions of a model are classified into four categories, see Table 2.

---

[5]This is slightly different from what used by [11] since they also included the non-scholastic order syllogisms, hence there are nine possible conclusions in their experiments, while we have five.

Data Set	Correct Prediction		Size	Mean Entropy	
	Count	Percentage		Predictions	Data
Test Set	153	95.6%	160	0.901	0.875
Training Set	151	94.4%	160	0.830	0.852
Complete Set	304	95.0%	320	0.870	0.864
NVC Premises*	212	94.2%	225	0.939	0.921
Valid Syl. Premises*	92	96.8%	95	0.706	0.729
Valid Syllogisms	23	95.8%	24	N/ A	N/ A

Table 3: Predictions evaluated according to the [11] method.
* The NVC premises are those from which no valid conclusion follows; the valid syl. premises are those from which at least one valid conclusion follows.

Sentence Types	A	I	E	O
Informativeness parameters	1.11	0.33	0.19	-0.78

Table 4: Values of the Informativeness Parameters.

3.2 Results

Table 3 shows the results. We see that the model is doing a good job, its proportion of correct predictions approximating a 95%.

3.3 Discussion

3.3.1 The Informativeness Parameters

The values of the informativeness parameters, as shown in Table 4, allow to make an interesting observation. Recall that we assumed that informativeness determines the confidence the reasoner has in the premises and, hence, the probability with which he concludes 'nothing follows'. We made no assumptions on 'which type of sentences' are more informative. The training results show that the amount of informativeness follow the order:

$$A(1.11) > E(0.33) > I(0.19) > O(-0.78),$$

which completely coincides with the proposal by Chater and Oaksford [3]. Besides, we see that sentence type 'O' is exceptionally uninformative, which also agrees with the authors' suggestion. The values of the informativenesses were learnt by the model. The result supports then the theory of Chater and Oaksford that the probabilistic validity plays an important role in human reasoning.

3.3.2 Parallel Comparison to Other Theories of the Syllogisms

We examined the predictions of a number of existing theories of the syllogistic reasoning. We were able to obtain the predictions of the PSYCOP model from Rips. The rest of the predictions were obtained from Table 7 in [11][6]. The results of the comparison are summarized is shown in Table 5. As far as we can see from the presented data our model outperforms other models.

---

[6]The table provided the predictions of the syllogistic theories on both the syllogisms that follow the scholastic order and the ones that do not. Our data are restricted to the scholastic order. The restriction has no influence

Theory	Hit	Miss	False Alarm	Correct Rejection	Correct Predictions
Atmosphere	44	41	20	215	259 /80.9%
Matching	41	44	55	180	221 /69.1%
Conversion	52	33	12	223	275 /85.9%
PHM*	40	45	63	172	212 /66.3%
PSYCOP	45	40	26	209	254 /79.4%
Verbal Models*	54	31	29	206	260 /81.2%
Mental Models*	85	0	55	180	265 /82.8%
Ver. 1 Test Data	26	15	12	107	133/83.1%
Ver. 2 Test Data	33	8	3	116	149/93.1%
Ver. 3 Complete Data**	70	14	5	231	301/94.1%
Ver. 3 Test Data	37	4	3	116	153/95.6%

Table 5: Predictions of the Theories of Syllogisms: A Summary.
*: Due to the limitations of the data we were able to obtain, the corresponding theory is likely to perform better than what is shown in the table.
**: The data in this line result from a cross-test: we take the predictions on the test data, then switched the test data and the training data and train the model again to get the predictions on the other half of the data.

4 Conclusion and Future Work

We have developed a preliminary framework of combining Natural Logic and data-driven inference weights and applied it to model syllogistic reasoning. The computational model learns from the experimental data, and as a result it may represent individual differences and explains subjects' systematic mistakes. This is achieved by assigning weights to all possible inference rules using machine-learning techniques and available data. The system is based on a Natural Logic proof system by Geurts [6], but it is less arbitrary, since it is empirically informed. In our approach we specify a tendency parameter for each inference rule. The agent begins with a pair of syllogistic premises and adopts each possible inference rule with a certain probability. As a result the longer the proof the less likely it is that an agent will find it. This simple setting solves the logical omniscience problem: not all derivations are available. Moreover, the approach takes into account various cognitive factors. For instance, the model enables the agents to adopt illicit conversions (e.g., yielding 'All A are B' from 'All B are A') in order to explain some systematic errors. Other version includes heuristic guesses based on two psychologically grounded principles. Firstly, the probability of drawing certain conclusions depends on the informativeness of the premises. Secondly, the model relies on the atmosphere hypothesis, e.g., when there is a negation in the premises, the agent is likely to draw a negative conclusion. We implemented and trained the models using the methodology outlined above and the empirical data from Chater and Oaksford [3]. We used a generalized EM algorithm to estimate the model and used it to compute the most probable syllogistic conclusions. The model was evaluated using the detection theory methods proposed in [11] to assess the performance of the theories of syllogistic reasoning. The complete version of the model makes 95% correct predictions, and therefore, outperforms all other known theories of syllogistic reasoning. In conclusion, the proposed combination of ideas gives rise to new, improved models of reasoning, where Natural Logic has replaced abstract rules, and the probabilistic parameters were derived from the data.

---

on the predictions of the atmosphere, matching, and conversion theories. However, for the PHM, the verbal model theory, and the mental model theory, we are unsure about the consequences.

The syllogistic fragment is an informative yet small arena for theories of reasoning. A natural next step would be to extend the model to cover a broader fragment of natural language by exploring existing Natural Logics [7] and designing new logics. We should then study formal (e.g., computational complexity) and psychological (e.g., cognitive resources) properties of the obtained models to draw new psychological conclusions and test the models against the data. The Natural Logics are usually computationally very cheap [12]. This guarantees that our models will easily scale-up to natural language reasoning. The computational complexity analysis will allow assessing the resources and strategies required to perform the reasoning tasks, cf. [14]. This in turn should open new ways of comparing our approach with other frameworks in psychology of reasoning.

References

[1] Ian Begg and Peter Denny. Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology*, 81(2):351, 1969.

[2] Johan van Benthem. *Language in Action: Categories, Lambdas and Dynamic Logic*. North-Holland, Amsterdam & MIT Press, Cambridge, 1991.

[3] Nick Chater and Mike Oaksford. The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 38(2):191–258, 1999.

[4] Nick Chater and Mike Oaksford. *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. Oxford University Press, 2008.

[5] Jakub Dotlačil, Jakub Szymanik, and Marcin Zajenkowski. Probabilistic semantic automata in the verification of quantified statements. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pages 1778–1783, 2014.

[6] Bart Geurts. Reasoning with quantifiers. *Cognition*, 86(3):223–251, 2003.

[7] Thomas Icard III and Lawrence Moss. Recent progress in monotonicity. *Linguistic Issues in Language Technology*, 9, 2014.

[8] Alistair Isaac, Jakub Szymanik, and Rineke Verbrugge. Logic and complexity in cognitive science. In Alexandru Baltag and Sonja Smets, editors, *Johan van Benthem on Logic and Information Dynamics*, volume 5 of *Outstanding Contributions to Logic*, pages 787–824. Springer International Publishing, 2014.

[9] Philip Johnson-Laird. An end to the controversy? A reply to Rips. *Minds and Machines*, 7(3):425–432, 1997.

[10] Philip Johnson-Laird and Ruth Byrne. *Deduction.* Lawrence Erlbaum Associates, Inc, 1991.

[11] Sangeet Khemlani and Philip Johnson-Laird. Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, 138(3):427, 2012.

[12] Ian Pratt-Hartmann. Fragments of language. *Journal of Logic, Language and Information*, 13(2):207–223, 2004.

[13] Lance Rips. *The psychology of proof: Deductive reasoning in human thinking.* MIT Press, 1994.

[14] Jakub Szymanik. *Quantifiers and Cognition. Logical and Computational Perspectives*. Studies in Linguistics and Philosophy. Springer, forthcoming, 2016.

[15] Víctor Manuel Sánchez Valencia. *Studies on Natural Logic and Categorial Grammar*. PhD thesis, University of Amsterdam, 1991.

Comparatives Revisited: Downward-Entailing Differentials Do Not Threaten Encapsulation Theories

Linmin Zhang[1] and Jia Ling[2]*

[1] Department of Linguistics, New York University
`linmin.zhang@nyu.edu`
[2] Department of Biology, New York University
`jl4327@nyu.edu`

Abstract

We analyze comparative morphemes (e.g., *-er*, *more*) as intervals of type $\langle dt \rangle$ that serve as differentials in comparatives. We propose that comparatives are about the distance between two intervals on a scale: the differential, which is an interval, is the result of subtracting the interval representing the position of the comparative standard on a scale from the interval representing the position of the comparative subject.

We show that our analysis has at least two advantages. First, it accounts for the semantics of comparatives with downward-entailing or non-monotone differentials in a very natural way, without relying on any strategy that essentially makes quantifiers inside of the *than*-clause take scope over the matrix clause. Second, it opens up new possibilities to give a unified account for various uses of comparative morphemes (e.g., *the more*, comparative correlatives, etc). We mainly focus on the first advantage in this paper.

1 Introduction

A large body of recent literature on comparatives has been focusing on comparatives that contain quantifiers inside the *than*-clause (see [21, 15, 17, 9, 7, 18, 20, 2, 1, 3, 5] among many others). These data raise a crucial question: whether *than*-clause-internal quantifiers take scope over the matrix clause.

As (1) illustrates, there are two ways to analyze the meaning of this sentence: (i) the **endpoint-based analysis** (see (1a)), according to which John's height is compared with the height of the tallest girl, and (ii) the **distribution-based analysis** (see (1b)), according to which John's height is compared with the height of each girl. In terms of truth condition, these two analyses are equivalent here: if John's height exceeds the height of the tallest girl, it follows necessarily that John's height exceeds the height of each girl, and vice versa.

(1)   John is taller than <u>every girl</u> is.

     a.   height(John) > height(the tallest girl)   **The endpoint-based analysis**

     b. $\forall x[\text{girl}(x) \rightarrow \text{height}(\text{John}) > \text{height}(x)]$ **The distribution-based analysis**

Evidently, the endpoint-based analysis does not involve distribution of *than*-clause-internal quantifiers. Consequently, theories that adopt this analysis (called **encapsulation theories** in [6]) do not require *than*-clause-internal quantifiers take scope over the matrix clause. In contrast, the distribution of *than*-clause-internal quantifiers is a necessary ingredient in the distribution-based analysis, and consequently, theories that adopt this analysis (called **entanglement theories** in [6]) necessarily require *than*-clause-internal quantifiers take scope over the matrix clause ([9, 3, 5]) or at least part of the matrix clause ([17]).

---

*We thank Chris Barker, Anna Szabolcsi and the anonymous reviewers for feedback.

While it is still debatable whether and how *than*-clause-internal quantifiers *can* take scope over the matrix clause in a syntactically plausible way (i.e., how they scope out of a syntactic island),[1] [6] suggests that somehow *than*-clause-internal quantifiers *must* take scope. As (2) and (3) show, [6] argues that only entanglement theories (e.g., [17, 9, 3, 5]), but not encapsulation theories (e.g., [2, 1]), can account for the semantics of the *than*-clause in a unified way, no matter whether there are non-monotone (see (2b) and (3b)) or downward-entailing (DE) differentials (see (2c) and (3c)). Thus, [6] concludes that only entanglement theories, i.e., theories that essentially require *than*-clause-internal quantifiers take scope, are empirically adequate.

(2) Entanglement theories: [[than every girl is (tall)]] $\approx \forall x[\text{girl}(x) \rightarrow \text{height}(x)...]$
    a. John is taller than every girl is.
    b. John is **exactly 4 inches** taller than every girl is
    c. John is **less than 4 inches** taller than every girl is.

(3) Encapsulation theories: how to interpret [[than every girl is (tall)]] in a unified way?
    a. John is taller than every girl is.   MAX reading
       height(John) > height(the tallest girl)
    b. John is **exactly 4 inches** taller than every girl is.   MAX=MIN reading
       height(John) > height(the tallest/shortest girl) $\rightsquigarrow$ Girls are of the same height.
    c. John is **less than 4 inches** taller than every girl is.   MAX-&-MIN reading
       height(the shortest girl) $+4'' >$ height(John) > height(the tallest girl)

In this paper, we show that DE or non-monotone differentials do not necessarily threaten encapsulation theories, and thus *than*-clause-internal quantifiers do not *have to* take scope.

Following [17, 14], we cast our endpoint-based analysis of comparatives not in terms of degrees, but in terms of intervals (i.e., convex sets of degrees), with the differential and the comparative standard analyzed as two intervals. In a nutshell, we claim that:

(4) a. In comparatives, *more/-er* refer to intervals that play the role of differentials.
    b. The differential (i.e., result of interval subtraction) is the distance between intervals.

Based on these new claims, we provide a simple and unified mechanism showing how to compositionally derive the truth conditions in (3) and explaining why the interpretation of *than*-clauses seems to vary with differentials and give rise to MAX/MAX=MIN/MAX-&-MIN readings.

§2 presents empirical motivation for our claims. §3 introduces the definition of interval subtraction. §4 shows how basic data of comparatives are analyzed with our proposal, and based on this, §5 shows how various kinds of differentials contribute to the computation of the semantics of comparatives. §6 compares the current analysis with [17]. §7 further shows that the current analysis of comparative morphemes opens up new possibilities to give a unified account for their various uses. §8 concludes this paper.

2 Empirical Motivation: New Observations

2.1 Comparatives Express the Distance Between Two Positions

Comparatives are a most interesting type of degree constructions. **Degrees**, which are points, are elements of **scales** (i.e., totally ordered sets); convex subsets of scales are often called **intervals**. Here we show that in analyzing comparatives, it is useful to (i) distinguish **interval**

---

[1]See, e.g., [15, 17], for arguments against the view that usual quantifier raising strategies can work.

scales from **ratio scales**, and (ii) consider the essential meaning of comparatives as a relation among **two intervals on an interval scale** and **one interval on a ratio scale**.

Interval scales and ratio scales are subtly different in whether they contain a meaningful, non-arbitrary and unique zero point: interval scales do not necessarily contain one, while ratio scales necessarily contain one. The distinction between **interval scales** and **ratio scales** as well as the use of both of them in comparatives are clearly shown in (5) and (6). Evidently, the scales of time and ranking are interval scales (e.g., *Rank 0* makes no sense; *8 o'clock* does not mean twice of *4 o'clock*). In contrast, the relevant scales that measure differentials are ratio scales: they have a meaningful, non-arbitrary and unique zero point – zero means no difference.

(5)   We arrived 2 hours earlier than the check-in time.

   a.   On an interval scale: (i) our arrival time; (ii) the check-in time.

   b.   On a ratio scale: the differential *2 hours*.

(6)   FSU ranked 3 spots higher than UNC.

   a.   On an interval scale: (i) the position of FSU; (ii) the position of UNC.

   b.   On a ratio scale: the differential *3 spots*.

This distinction between interval scales and ratio scales explains why comparatives cannot express the absolute position of the comparative subject or comparative standard on a scale: the absolute position depends on the choice of the origin (i.e., zero point), and this choice can be arbitrary on an interval scale. Instead, comparatives express the absolute distance between the positions standing for the comparative subject and the comparative standard: once these two positions on an interval scale are settled, the distance between them remains constant, no matter how the zero point is chosen and how the absolute positions are defined accordingly.

Thus, we consider the essential meaning of comparatives as a relation among three things: **two positions on an interval scale** (i.e., the one representing the comparative standard, e.g., the check-in time in (5), and the one representing the comparative subject, e.g., our arrival time in (5)) and **the distance between them**.[2] Based on this, we follow [17, 14] and use intervals (i.e., convex sets of points), instead of degrees (i.e., points), to represent positions. An interval represents a value as **a range of possibilities**, and thus intuitively, they can be seen as larger and more generalized markers of positions on a scale. Intervals not only mark positions, but also have size (consider error bars) and carry endpoint information (e.g., boundedness, closeness).[3]

(7)   Interval notation: Type of **degree**: d; type of **interval**: $\langle dt \rangle$
        An interval $\lambda \delta_d.\{\delta | D_{\min} \leq \delta \leq D_{\max}\}$ can be written as $[D_{\min}, D_{\max}]$.[4]

---

[2]One of the anonymous reviewers pointed out that our view is inconsistent with traditional assumptions in studying comparatives (e.g., [4, 21, 8, 18]). According to traditional assumptions, comparatives express relations between (positive) thresholds: thresholds can be ordered in such a way that (i) if something meets or exceeds one of them, it meets or exceeds all **lower thresholds**, and (ii) if the highest threshold A meets exceeds the highest threshold B meets, A meets or exceeds **more thresholds** than B does. Crucially, this threshold-based view implicitly assumes that the highest thresholds A and B meet respectively are not infinitely far away (compared to the distance between them) from a certain reference point, i.e., the highest thresholds A and B meet cannot be situated at a $+\infty$ position. Otherwise, even though the highest threshold A meets exceeds the highest threshold B meets, A does not exceed *more* thresholds than B does (e.g., $+\infty + 5$ is not larger than $+\infty$). This underlying assumption is certainly not guaranteed in the worst cases: in (5), when the temporal scale extends to an infinite future, even though our arrival time exceeds the check-in time in being early, they are equally far away from the zero point, which is in infinite future. Since the scale on which the comparative subject and the comparative standard are situated can have an arbitrary zero point, considering lower thresholds in analyzing comparatives can potentially ruin the analysis and thus makes no sense. After all, only the distance between the positions standing for the comparative subject and the comparative standard matters.

[3]See [17] for additional arguments for using intervals, instead of degrees, as position markers on a scale.

[4]When $D_{\min} = D_{\max}$, it is a singleton set, i.e., it contains a single point.

2.2 Comparative Morphemes Represent Differentials

The semantic contribution of comparative morphemes is a fundamental issue in studying comparatives. Here we show that comparative morphemes play the role of differentials.

The crucial empirical motivation is shown in (8) and (9). The most natural interpretation for the use of *more* in (8) is not that the amount he then drank is (a bit) larger than the amount he had drunk previously, but just an amount (a bit) over zero. In other words, *more* is related to the part that is added onto some *augend* (i.e., thing to be increased). Similarly, in (9), *more* signals a second event (i.e., a bringing-chaos event) being added onto the event already existing in the context (i.e., the bringing-depression event), i.e., *more* corresponds to the differential part between the first event e_1 and the sum of the two events $[e_1 + e_2]$. Thus *more* is reminiscent of additive words (e.g., *other, also, too*) in (i) expressing an additive meaning and (ii) being anaphoric: it is felicitous only when there is already an augend in the context.

(8)   He drank till he blacked out. Then he drank (a bit) **more**.

(9)   War brings depression; **what's more**, it brings chaos.

Then how to account for the use of *more* in (8) and (9) and its use in comparatives in a unified way? If we start from comparatives and analyze the fundamental contribution of *more* as relating two degree (or interval) expressions (i.e., *more* is of type $\langle d, \langle dt \rangle \rangle$ or $\langle dt, \langle dt, t \rangle \rangle$), the use of *more* in (8) and (9) remains a puzzle.

However, if we start from (8) and (9) and analyze *more* as an addend (or differential), i.e., the difference between a sum and an augend, this analysis can be immediately extended to cover comparative data. In comparatives, the augend, the addend (or differential) and the sum are all in the same sentence: (i) the comparative standard plays the role of augend, (ii) the comparative subject the role of sum, and (iii) comparative morphemes the role of differential.

In §2.1, we have proposed to use intervals of type $\langle dt \rangle$ to represent positions on a scale. When an interval (e.g., the interval marking the position of the comparative subject) minus another interval (e.g., the interval marking the position of the standard), the result, i.e., the distance between two positions, is also an interval. Thus, if *more/-er* are analyzed as differentials, then in comparatives, they should be intervals of type $\langle dt \rangle$. We propose that they are intervals in the domain $\lambda D.[D \subseteq (0, +\infty)]$. When a comparative sentence contains a more specific differential, e.g., *2 hours* in (5) and *a bit* in (8), this specific differential further restricts the value of *more/-er*. (10) shows how intervals of type $\langle dt \rangle$ can be compared to individuals of type e:

(10)

x_e:	someone	other	the other	another	John	John, a linguist
$D_{\langle dt \rangle}$:	some	more	the more	one more	$[3'', +\infty]$	3 feet more/-er

3 Interval Subtraction

In §2, we have motivated our analysis of using three intervals (i.e., two representing positions on a scale and one representing the differential/distance between them) to characterize the semantics of comparatives. Here we introduce the definition of interval operations:

(11)   Interval operations: $[x_1, x_2]\langle op \rangle [y_1, y_2] = [\text{MIN}(x_1 \langle op \rangle y_1, x_1 \langle op \rangle y_2, x_2 \langle op \rangle y_1, x_2 \langle op \rangle y_2),$
        $\text{MAX}(x_1 \langle op \rangle y_1, x_1 \langle op \rangle y_2, x_2 \langle op \rangle y_1, x_2 \langle op \rangle y_2)]$ (see [16])

(12)   Interval subtraction: $[x_1, x_2] - [y_1, y_2] = [x_1 - y_2, x_2 - y_1]$

(13)   a. $[5, 8] - [1, 2] = [3, 7]$
        b. $[5, 8] - [3, 7] = [-2, 5]$

Since an interval represents a value as a range of possibilities, as (11) shows, interval operations result in the largest possible range. Thus, we can simply write interval subtraction as shown in (12). (13) shows two examples. Notice that interval subtraction is different from subtraction defined in number arithmetic: when X, Y and Z represent numbers, if $X - Y = Z$, it follows necessarily that $X - Z = Y$; however, when they represent intervals, as (13a) and (13b) illustrate, if $X - Y = Z$, generally speaking, it is not the case that $X - Z = Y$.

A consequence is that in interval arithmetic, given $X - Y = Z$ and given the values of Y and Z, to compute the value of X, we cannot perform interval addition on Y and Z (see (14)).

(14)   a.   If $X - [a, b] = [c, d]$, then generally speaking, it is not the case that $X = [a+c, b+d]$.
              b.   If $X - [a, b] = [c, d]$, X is undefined when $b + c > a + d$ (i.e., when the lower bound of X is larger than the upper bound of X); when defined, $X = [b + c, a + d]$.

4   Accounting for Basic Data

4.1   The Semantics of Scalar Adjectives

We follow standard treatments of scalar adjectives (see [4, 21, 8, 13, 10, 12, 11] among others): scalar adjectives relate individuals with abstract representations of measurement on a scale. Since we use intervals of type $\langle dt \rangle$ to represent positions on a scale, scalar adjectives are of type $\langle dt, et \rangle$ in our analysis, as shown in (15). (16) shows the semantics of the positive form. In (16), D_c is is definite. It is shorthand for 'the contextually salient interval such that it is from the lower bound to the upper bound of being tall for a relevant comparison class'. Its semantic contribution is somehow similar to that of, e.g., [11]'s *pos* operator. In (17), *exactly 6 feet* is interpreted as an interval, which is a singleton set of degrees, i.e., $[6', 6']$.

(15) $[\![\text{tall}]\!]_{\langle dt, et \rangle} \stackrel{\text{def}}{=} \lambda D_{\langle dt \rangle}.\lambda x_e.[\text{height}_{\langle e, dt \rangle}(x) \subseteq D]$
          i.e., the height of the individual x is in the interval D.

(16) $[\![\text{John is } D_c \text{ tall}]\!] \Leftrightarrow \text{height}(\text{John}) \subseteq D_c$
          i.e., the height of John is in the contextually salient interval of being tall.

(17) $[\![\text{John is exactly 6 feet tall}]\!] \Leftrightarrow \text{height}(\text{John}) \subseteq [6', 6']$
          i.e., the height of John is at the position '6 feet' on the height scale.

4.2   The Semantics of Comparatives

As we have proposed in §2.2, (18) shows that comparative morphemes denote an interval. As shown in (19), we propose that $[\![\text{than}]\!]$ takes two interval arguments – D_{standard} (i.e., the interval standing for the comparative standard) and $D_{\text{differential}}$ (i.e., the differential) – and returns the unique interval that is $D_{\text{differential}}$ away from D_{standard}.[5]

(18) $[\![\text{more/-er}]\!]_{\langle dt \rangle} \stackrel{\text{def}}{=} D$ such that $D \subseteq (0, +\infty)$
          (Presupposition requirement: there is an augend in the context.)

(19) $[\![\text{than}]\!]_{\langle dt, \langle dt, dt \rangle \rangle} \stackrel{\text{def}}{=} \lambda D_{\text{standard}}.\lambda D_{\text{differential}}.\iota D[D - D_{\text{standard}} = D_{\text{differential}}]$

Evidently, based on the definition of interval subtraction (12), the operation of $[\![\text{than}]\!]$ is well defined if and only if the sum of the lower bound of $D_{\text{differential}}$ and the upper pound of D_{standard} is not larger than the sum of the upper bound of $D_{\text{differential}}$ and the lower bound of D_{standard}.

---

[5]The semantic operation we propose in (19) might be carried out by a silent item. We stay ignorant on this.

Based on (18) and (19), we show in (20) details of a compositional derivation for the truth condition of a comparative sentence containing a specific differential.

(20)   Computing the truth condition of ⟦John is 5 inches taller than Mary is (tall)⟧:

   a.   ⟦Mary is D (tall)⟧ ⇔ height(Mary) ⊆ D i.e., Mary is D tall.
   b.   Following, e.g., [14, 2, 1, 3], we assume that there is a lambda abstraction. We also assume a silent operator ⟦THE⟧ here (defined as $\lambda P_{\langle \alpha t \rangle}.\iota x[P(x)]$) (see [9]), which turns $[\lambda D.[\text{height(Mary)} \subseteq D]]$ into a contextually unique interval (that allows some vagueness), i.e., the definite interval standing for Mary's height. ⟦THE⟧$[\lambda D.[\text{height(Mary)} \subseteq D]]$ can be written as $[D_{\text{Lower-Mary}}, D_{\text{Upper-Mary}}]$, i.e., the interval from the lower bound to the upper bound of Mary's height.
   c.   ⟦5 inches ... -er⟧ ⇔ $[5'', 5''] \cap (0, +\infty)$ ⇔ $[5'', 5'']$
   d.   ⟦5 inches ... -er than Mary is⟧ ⇔ ⟦than⟧$([D_{\text{Lower-Mary}}, D_{\text{Upper-Mary}}])([5'', 5''])$
        ⇔ $\iota D.[D - [D_{\text{Lower-Mary}}, D_{\text{Upper-Mary}}] = [5'', 5'']]$
   e.   ⟦John is 5 inches taller than Mary is (tall)⟧
        ⇔ ⟦tall⟧⟦5 inches ... -er than Mary is⟧(John)
        ⇔ height(John) ⊆ $\iota D.[D - [D_{\text{Lower-Mary}}, D_{\text{Upper-Mary}}] = [5'', 5'']]$
        i.e., on the height scale, John's height is at such a position that it is $[5'', 5'']$ away from the interval $[D_{\text{Lower-Mary}}, D_{\text{Upper-Mary}}]$.
   f.   After simplification: height(John) ⊆ $[D_{\text{Upper-Mary}} + 5'', D_{\text{Lower-Mary}} + 5'']$.
        i.e., on the height scale, John's height is at the position represented by the interval $[D_{\text{Upper-Mary}} + 5'', D_{\text{Lower-Mary}} + 5'']$.
        This interval is defined only when $D_{\text{Upper-Mary}} + 5'' \leq D_{\text{Lower-Mary}} + 5''$, i.e., $D_{\text{Upper-Mary}} = D_{\text{Lower-Mary}}$. In other words, the position that stands for Mary's height has to be a single point, and John's height is a point $5''$ farther away from the point representing Mary's height on the scale.

5   The Interplay between Differentials and the Interpretation of *Than*-Clause-Internal Quantifiers

Here we analyze comparatives containing various kinds of differentials. To begin with, we first show in (21) the interpretation of a comparative standard that contains a universal quantifier.

(21)   a.   ⟦every girl is D (tall)⟧ ⇔ $\forall x.[\text{girl}(x) \rightarrow \text{height(x)} \subseteq D]$
             i.e., for each girl x, x's height is situated in the interval D on the height scale.
        b.   After a lambda abstraction and the application of a silent ⟦THE⟧, it becomes
             ⟦THE⟧$[\lambda D.[\forall x.[\text{girl}(x) \rightarrow \text{height(x)} \subseteq D]]]$
             i.e., the contextually unique interval in which every girl's height is situated. In the following, we write this as $[D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}]$, i.e., the interval from the lower bound to the upper bound of girls' height.

(22) shows the derivation of the so-called MAX reading. In fact, when the differential is upward-entailing, the upper bound of $D_{\text{differential}}$ is unbounded, i.e., $+\infty$, and the sum of the lower bound of the comparative standard and $+\infty$ is still $+\infty$, which is necessarily larger than the sum of the lower bound of the differential and the upper bound of the comparative standard. This has two consequences: (i) there is no extra requirement to make the interval representing John's height well defined; (ii) only the upper bound (but not the lower bound) of the comparative standard shows up in the truth condition after simplification (see (22c)).

(22)   John is taller than every girl is.

    a. $D_{\text{differential}} = [\![\text{-er}]\!] = (0, +\infty)$

    b. $[\![\text{John is taller than every girl is (tall)}]\!]$
        $\Leftrightarrow [\![\text{tall}]\!][\![\text{... -er than every girl is (tall)}]\!](\text{John})$
        $\Leftrightarrow \text{height}(\text{John}) \subseteq \iota D[D - [D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}] = (0, +\infty)]$

    c.   After simplification: $\text{height}(\text{John}) \subseteq (D_{\text{Upper-Girls}}, +\infty)$

(23) and (24) show the derivation of the MAX-&-MIN reading for two sentences containing DE differentials. It is evident that the MAX-&-MIN reading is due to the fact that when the overt differential is downward-entailing, $D_{\text{differential}}$ is bounded at both endpoints. Moreover, to make the interval representing John's height well defined here, it has to be the case that $D_{\text{Upper-Girls}} < D_{\text{Lower-Girls}} + 4''$, i.e., the length of the interval containing girls' height is less than 4 inches. Also notice that *less than X* and *at most X* differ in that the upper bound of *less than X* is open while the upper bound of *at most X* is close: the openness of the upper bound of the DE differential also determines whether the interval representing the position of the comparative subject has an open or close upper bound.

(23)   John is **less than 4 inches** taller than every girl is.

    a. $D_{\text{differential}} = [\![\text{less than 4 inches ... -er}]\!] = (0, +\infty) \cap (-\infty, 4'') = (0, 4'')$

    b. $[\![\text{John is less than 4 inches taller than every girl is (tall)}]\!]$
        $\Leftrightarrow \text{height}(\text{John}) \subseteq \iota D[D - [D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}] = (0, 4'')]$

    c.   After simplification: $\text{height}(\text{John}) \subseteq (D_{\text{Upper-Girls}}, D_{\text{Lower-Girls}} + 4'')$

(24)   John is **at most 4 inches** taller than every girl is.

    a. $D_{\text{differential}} = [\![\text{at most 4 inches ... -er}]\!] = (0, +\infty) \cap (-\infty, 4''] = (0, 4'']$

    b. $[\![\text{John is at most 4 inches taller than every girl is (tall)}]\!]$
        $\Leftrightarrow \text{height}(\text{John}) \subseteq \iota D[D - [D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}] = (0, 4'']]$

    c.   After simplification: $[\text{height}(\text{John}) \subseteq (D_{\text{Upper-Girls}}, D_{\text{Lower-Girls}} + 4'']$

Finally, (25) and (26) illustrate the meaning derivation of comparatives containing non-monotone differentials. To make the interval representing John's height well defined, in (25), it has to be the case that all the girls have the same height (i.e., $D_{\text{Upper-Girls}} = D_{\text{Lower-Girls}}$), and thus the sentence has the so-called MAX-&-MIN reading. Similarly, in (26), the length of the interval containing girls' height cannot be larger than 2 inches.

(25)   John is **exactly 2 inches** taller than every girl is.

    a. $D_{\text{differential}} = [\![\text{exactly 2 inches ... -er}]\!] = (0, +\infty) \cap [2'', 2''] = [2'', 2''].$

    b. $[\![\text{John is exactly 2 inches taller than every girl is (tall)}]\!]$
        $\Leftrightarrow \text{height}(\text{John}) \subseteq \iota D[D - [D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}] = [2'', 2'']]$

    c.   After simplification: $\text{height}(\text{John}) \subseteq [D_{\text{Upper-Girls}} + 2'', D_{\text{Lower-Girls}} + 2'']$

(26)   John is **between 2 and 4 inches** taller than every girl is.

    a. $D_{\text{differential}} = [\![\text{between 2 and 4 inches ... -er}]\!] = (0, +\infty) \cap [2'', 4''] = [2'', 4''].$

    b. $[\![\text{John is between 2 and 4 inches taller than every girl is (tall)}]\!]$
        $\Leftrightarrow \text{height}(\text{John}) \subseteq \iota D[D - [D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}] = [2'', 4'']]$

    c.   After simplification: $\text{height}(\text{John}) \subseteq [D_{\text{Upper-Girls}} + 2'', D_{\text{Lower-Girls}} + 4'']$

In sum, we have shown how to compositionally derive the correct truth condition of comparatives containing various kinds of differentials in an effortless and precise way: no distributive operation is needed, and no *ad hoc* tweak is employed. In our account, since we do not need

to have access to each individual's height, it follows naturally that we do not need to make *than*-clause-internal quantifiers take scope over the matrix clause. The whole mechanism only requires that we have access to the lower and upper bounds of the girls' height. We assume a silent $[\![\text{THE}]\!]$ to achieve this in our account, i.e., we interpret the part following *than* as a definite interval (see also [9]). Other existing encapsulation theories (e.g., [2, 1]) have proposed their own mechanisms to derive the semantics of the endpoints of the comparative standard. A detailed comparison among these mechanisms is left for future work.

5.1 Extension: Accounting for *Fewer Than*

Here we extend our account to comparative data using *fewer/less than*. We propose that the semantics of *less/fewer* includes two parts: (i) the comparative morpheme $[\![\text{-er}]\!]$ and (ii) an operator that changes the direction of comparison (see (27)). How to connect the analysis in (27) with other syntactic/semantic behaviors of *few/less* is left for future research.

(27) $[\![\text{few- than}]\!]_{\langle dt, \langle dt, dt \rangle \rangle} \stackrel{\text{def}}{=} \lambda D_{\text{standard}}.\lambda D_{\text{differential}}.\iota D[D_{\text{standard}} - D = D_{\text{differential}}]$

(28)   If $[a, b] - X = [c, d]$, X is undefined when $b + c > a + d$; when defined, $X = [b - d, a - c]$.

(29)   John is **more than 4 inches** less tall than every girl is.

    a. $D_{\text{differential}} = [\![\text{more than 4 inches ... -er}]\!] = (0, +\infty) \cap (4'', +\infty) = (4'', +\infty)$.

    b. $[\![\text{John is more than 4 inches less tall than every girl is (tall)}]\!]$
        $\Leftrightarrow \text{height(John)} \subseteq \iota D[[D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}] - D = (4'', +\infty)]$

    c.   After simplification: $\text{height(John)} \subseteq (-\infty, D_{\text{Lower-Girls}} - 4'')$

(30)   John is **at most 4 inches** less tall than every girl is.

    a. $D_{\text{differential}} = [\![\text{at most 4 inches ... -er}]\!] = (0, +\infty) \cap (-\infty, 4''] = (0, 4'']$.

    b. $[\![\text{John is at most 4 inches taller than every girl is (tall)}]\!]$
        $\Leftrightarrow \text{height(John)} \subseteq \iota D[[D_{\text{Lower-Girls}}, D_{\text{Upper-Girls}}] - D = (0, 4'']]$

    c.   After simplification: $\text{height(John)} \subseteq [D_{\text{Upper-Girls}} - 4'', D_{\text{Lower-Girls}})$

6   Comparison with [17]

[17] also uses intervals, instead of degrees, to implement the semantics of comparatives. A crucial difference between our analysis and [17]'s consists in the definition of interval subtraction, and along with it, the definition of differential. Our account follows the standard definition developed in interval arithmetic (see [16]). (31) shows [17]'s definition of subtraction. (32) illustrates how the definitions (31) and (12) differ: the contrast between (32a) and (32b) clearly shows that the analysis of [17] is problematic.

(31)   Assuming I is above K, we want $[I - K]$ to pick out the part of the scale that is below I and above K. The differential is considered as the length of $[I - K]$.
    For intervals I, K:
    If $K \prec I$, then: $\forall J : (J \prec I \,\&\, K \prec J) \leftrightarrow J \sqsubseteq [I - K])$
    Otherwise: $[I - K] = 0$ (56) in [17]

(32)   Suppose the height of each boy is somewhere between $5'8''$ and $5'11''$, and suppose the height of each girl is somewhere between $5'3''$ and $5'7''$.

    a.   According to (12), the result of $[5'8'', 5'11''] - [5'3'', 5'7'']$ is $[1'', 8'']$, and in our account this result is the differential. To describe the situation, we would say:
        Every boy is <u>between 1'' to 8''</u> taller than every girl.   **True** in the scenario

b. According to (31), the result of $[5'8'', 5'11''] - [5'3'', 5'7'']$ is $(5'7'', 5'8'')$, and in [17], the differential in comparatives is understood as the length of this subtraction result: in this case, it is less than $1''$. To describe the situation, we would say:

Every boy is <u>less than $1''$</u> taller than every girl.   **False** in the scenario

7   Discussion: other uses of *more*

As we have shown in §2.2, *more* essentially refers to a differential. We have also suggested in (10) that *more* should behave quite similarly to indefinites in many cases. The crucial difference between *more* and usual indefinite expressions is that *more* brings a presuppositional requirement: there has to be an augend in the context. Our analysis of *more* in comparatives opens up new possibilities to relate various data of *more*, comparatives and superlatives.

    The more.   [19] questions how *more* is related to *the more*, and points out that while *more* can take a *than*-clause, *the more* cannot. Under our analysis, the meaning of *taller than Bill is* (in (33a)) is totally parallel to the meaning of D_c *tall* (in (34a)). Thus, it is unsurprising that if *the* cannot compose with *tall* to form a grammatical construction (in (34b)), *the* cannot compose with *taller than Bill is* to form a grammatical construction either (in (33b)). Then it should be due to the same reason that (35b) is ungrammatical.

    Now when we look back at our lexical entry for *than* in (19), evidently, the result of performing $[\![than]\!]$ on $D_{standard}$ and $D_{differential}$ is already a definite interval. Thus, our analysis explains why *than*-clause is no longer compatible with an overt *the*.

(33)   a.   John is taller than Bill is. $[\![(33a)]\!] \Leftrightarrow [\![tall]\!][\![\text{ ... -er than Bill is}]\!](\text{John})$
       b.   *John is the taller than Bill is.

(34)   a.   John is D_c tall. $[\![(34a)]\!] \Leftrightarrow [\![tall]\!][\![D_c]\!](\text{John})$
       b.   *John is the tall.

(35)   a.   John earns more money than Bill does.
       b.   *John earns the more money than Bill does.

    Comparative correlatives.   Interestingly, *the more* seems to be a cross-linguistically very prevailing pattern in expressing correlations. By analyzing *more/-er* as differential, a unified account for comparatives and comparative correlatives should be readily available.

    One intriguing question here is in comparative correlatives, as illustrated in (36), whether the correlation is between two sums, or just between two differentials. Given the previous discussion, it seems that in comparative correlatives, if the correlation is established between two sums, there cannot be an overt *the* in the part *the more*. Thus, *the more* should refer to only the differential part (consider also *the other* (see (10))), and the correlation should be established between two differentials. A further question is whether there is a binding relation between the two uses of *the more* in comparative correlatives. This is left for future research.

(36)   The more I read, the more I understand.

8   Conclusion

In this paper, we provide a new implementation of the endpoint-based analysis to account for comparative data. Our new implementation is based on two claims: (i) comparatives express

the relation among three intervals, among which one represents the differential between the other two, and (ii) comparative morphemes should be analyzed as differentials. Technically, our implementation is based on interval arithmetic. With this new implementation, we account for comparative data containing various kinds of differentials in an easy and unified way. More particularly, for comparatives containing *than*-clause-internal quantifiers, no scope taking is needed in our account. Hopefully, our analysis will shed light on more issues concerning comparatives, *more* and uses of intervals in natural language.

References

[1] Peter Alrenga and Christopher Kennedy. *No more* shall we part: Quantifiers in English comparatives. *Natural Language Semantics*, 22(1):1–53, 2014.

[2] Sigrid Beck. Quantifiers in *than*-clauses. *Semantics and Pragmatics*, 3(1):1–72, 2010.

[3] Sigrid Beck. Plural predication and quantified 'than'-clauses. In *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 1, pages 91–115, 2014.

[4] Max Cresswell. The semantics of degree. In Barbara Partee, editor, *Montague grammar*, pages 261–92. New york: Academic Press, 1976.

[5] Jakub Dotlačil and Rick Nouwen. The comparative and degree pluralities, 2015. Manuscript.

[6] Nicholas Fleisher. Comparing theories of quantifiers in *than* clauses: Lessons from downward-entailing differentials. *Semantics and Pragmatics*, 2015 (to appear).

[7] Jon Gajewski. More on quantifiers in comparative clauses. In *Semantics and Linguistic Theory (SALT) 18*, pages 340–357, 2008.

[8] Irene Heim. Notes on comparatives and related matters, 1985. Manuscript, University of Texas, Austin.

[9] Irene Heim. Remarks on comparative clauses as generalized quantifiers, 2006. Manuscript, MIT.

[10] Christopher Kennedy. *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge, 1999.

[11] Christopher Kennedy. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45, 2007.

[12] Christopher Kennedy and Louise McNally. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81(2):345–381, 2005.

[13] Ewan Klein. Comparatives. In A. von Stechow and D. Wunderlich, editors, *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, pages 673–691, 1991.

[14] Fred Landman. Internal and interval semantics for CP-comparatives. In *Logic, Language and Meaning: Proceedings of the 17th Amsterdam Colloquium*, pages 133–142. Springer, 2010.

[15] Richard Larson. Scope and comparatives. *Linguistics and Philosophy*, 11(1):1–26, 1988.

[16] Ramon E. Moore. *Methods and Applications of Interval Analysis*. SIAM, 1979.

[17] Roger Schwarzchild and Karina Wilkinson. Quantifiers in comparatives: A semantics of degree based on intervals. *Natural Language Semantics*, 10(1):1–41, 2002.

[18] Roger Schwarzschild. The semantics of comparatives and other degree constructions. *Language and Linguistics Compass*, 2(2):308–331, 2008.

[19] Anna Szabolcsi. Compositionality without word boundaries: *(the) more* and *(the) most*. In *Semantics and Linguistic Theory (22)*, pages 1–25, 2013.

[20] Robert van Rooij. Comparatives and quantifiers. *Empirical issues in Syntax and Semantics*, 7:1–22, 2008.

[21] Arnim von Stechow. Comparing semantic theories of comparison. *Journal of Semantics*, 3(1):1–77, 1984.