DISCLAIMER:

This document does not meet the current format guidelines of the Graduate School at The University of Texas at Austin.

It has been published for informational use only.

The Dissertation Committee for Megan Alexandra Hyska
certifies that this is the approved version of the following dissertation:

This Machine Kills Fascists: Detecting Propaganda with Formal Models of Mass Discourse Structure (and other ideas)

Committee:

Joshua Dever, Supervisor

David Beaver

Lawrence Ray Buchanan

Johan Kamp

Craige Roberts

Jason Stanley

This Machine Kills Fascists: Detecting Propaganda with Formal Models of Mass Discourse Structure (and other ideas)

by

Megan Alexandra Hyska

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2018

Dedicated to my parents, Terry and Cathy, who were never once skeptical about their daughter going to school to study philosophy.

Acknowledgments

I would like to thank the members of my dissertation committee, and especially my supervisor, Josh Dever, for his unfailing intellectual generosity and personal kindness.

I would also like to express my gratitude for an indispensable, though non-academic, source of the calm and fortitude I needed in order to finish this thing off: my gym, Lions Krav Maga, and the many supportive training partners and coaches I've had the pleasure of working with there (in philosophy, as on the mat: drillers make killers).

Finally, I would like to thank the women of the UT philosophy department, whose friendship I have relied upon, and whose resilience and dedication to the discipline have given me heart.

This Machine Kills Fascists: Detecting Propaganda with Formal Models of Mass Discourse Structure (and other ideas)

Publication No. _____

Megan Alexandra Hyska, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Joshua Dever

This dissertation consists of three free-standing chapters, all of which deal with themes in the philosophy of language in general and in pragmatics in particular.

Chapter 1 augments an existing model of discourse-level information structure in order to articulate relevance conditions that make the right predictions about the class of utterances I call higher order utterances. The apparatus introduced turns out also to have some utility for the representation of a certain kind of discourse defectiveness.

Chapter 2 explores an analogy between communication and knowledge to ultimately advance the thesis that communication is not a uniquely valuable type of signaling event.

Chapter 3 draws connections between propaganda and polarization, and then between the measurement of polarization and the tools used in the

modeling of discourse structure. The upshot is that tools from formal pragmatics have an application in the characterization, and perhaps ultimately detection, of propaganda.

Table of Contents

List of Figures

Chapter 0

Introduction

The three chapters of this dissertation are free-standing contributions to literatures in and around the philosophy of language. However, they are each to some degree an examination of the phenomenon of *discourse-level information structure*. Informally, discourse-level information structure consists of the information shared in a discourse, along with the way this information is organized into topics. That the contents of iterated signaling transactions among human beings can be expected to be interrelated rather than independent— that conversations are, in this way, *organized*— seems so inevitable that it is difficult to imagine an agent who didn't already know this being able to engage in communication at all. But while the mere fact that conversation tends to involve the organized exchange of information may be self-evident, articulating the nature of this organization is a challenge. Meeting this challenge, however, pays considerable dividends (or stands to) in the study of language and, it is the suggestion of two of the dissertation's three chapters, in understanding the goals of collaborative deliberation. Whether in its capacity as a factor in linguistic inference, as a source of value for discourse, or as an observable proxy for the deliberative processes of the conversationalists, information structure figures in all three chapters.

The first chapter is an examination of how an existing approach to modeling information structure (Roberts (1996, 2012)) can be augmented to reflect the conversational dynamics of *higher order utterances* (HOUs). Higher order utterances are those that describe or comment upon the information structure of the discourse itself, and they are distinctive in that they are relevant at any point in discourse. Now, in question-under-discussion based models of information structure, an utterance's relevance has typically been accounted for in terms of its bearing the relation of answerhood, or else of subquestionhood, to the question presently at the top of the stack. But HOUs, characteristically, don't answer the question presently under discussion; they restate or describe the question, request clarification of it, or comment on its strategic merit or appropriateness. Question-under-discussion models have thus generated false predictions with respect to HOUs. I consider how to address this issue in the light of the insight that, at least for a subset of HOUs, their distinctive function involves questioning the relation of the question-under-discussion to a broader set of interlocutor goals. I argue that thinking of information structure as involving a *set* of stacks, one for each order of discourse, allows us to characterize the information structural effects distinctive of HOUs (i.e. that they add and answer questions on higher order stacks) and so articulate relevance conditions consistent with the data of speaker judgments. The positive proposal can be thought of as suggesting that, in interlocutors' attempts to answer the question "what is the world like?" the subquestion "what is this conversation like?" has a special status. A theory of context which carves out

a special way of representing information that answers this question offers not only a way to distinguish the effects of HOUs, but also a way of representing certain kinds of common discourse defectiveness.

The second chapter is an extended consideration of the analogy between communication and knowledge, which ultimately invokes information structure to account for the unique value of a kind of signaling event *other* than communication. The chapter starts from the observation that the justification, truth, and belief conditions of the classical analysis of knowledge each have an analog condition in standard analyses of communication; call the analogous conditions on communication A, B, and C. Next, it lays out the evidence that ABC are not jointly sufficient conditions on communication, in just the way that Gettier cases show JTB conditions on knowledge not to be. Moreover, it seems as though proposals for a fourth condition which, alongside ABC, would make for a set of necessary and jointly sufficient conditions on communication, are plagued by many of the same issues that have afflicted such proposals in the analysis of knowledge literature. My project in this chapter was not to propose a successful analysis of communication, but to depart from the conclusion that a Gettier-proof analysis will be hard to come by, and will no doubt involve a fairly tortured, gerrymandered fourth condition, in the way that Jonathan Kvanvig (2003) does from his analogous conclusion about knowledge. Kvanvig suggests that whatever this fourth condition on knowledge is, it will not be such as to render knowledge more valuable than justified true belief; knowledge, he therefore suggests, is not actually a uniquely valuable epistemic state.

What he goes on to argue, however, is that the reasons that suggest that justified true belief is more epistemically valuable than mere true belief likewise suggest that the state of *understanding* is more valuable than justified true belief or knowledge, where understanding consists in having (possibly Gettiered) justified true beliefs about a "chunk" of related information, and seeing how the individual bits of information in this chunk are related to one another. I explore the analogous argument that communication is not more valuable than signaling events which satisfy A, B and C; it follows that communication is not a uniquely valuable sort of signaling event. The question that then arises is whether understanding has an analog in the domain of signaling events. I suggest that it does; just as understanding consists in justified true beliefs about a chunk of related information and the ability to see the connections between these pieces of information, *comprehension* consists in ABC (the audience's entertaining the proposition intended by the speaker, and moreover doing so because it is manifest to her that the speaker wants her to do on the basis of the utterance), with respect to a *collection* of utterances, alongside an appreciation of the *connections* between utterances in this collection. A "collection of utterances" is naturally identified as a discourse, and grasping the connections between the discourse's constitutive utterances is naturally thought of as a matter of tracking the discourse's information structure. This suggestion elevates an audience's grip on discourse-level information structure as the distinguishing feature of a signaling event more intrinsically valuable than communication itself.

The final chapter of the dissertation suggests that the information structure of a conversation can be thought of as a proxy for the *deliberative strategy* undertaken in that conversation. Intuitively, agents may diverge in their consideration of an issue not merely insofar as they reach different conclusions; they may also regard different considerations as salient to their deliberation, and so diverge in the way they reason even prior to, or in the absence of, any discrepancy in conclusions. There are many compelling reasons one might want to investigate this sort of *deliberative strategy variation* in a population, but of course deliberation itself is not directly observable. However, if we regard discourse as a way of enacting a collaborative deliberative process, we might conclude that the information structure of the conversation is an attractive proxy for the phenomenon of deliberative strategy. QUD-based models, as opposed to topic-modeling techniques that construe topics as frequency distributions across a collection of terms, have particular potential as a tool for analyzing how deliberation proceeds in a given conversation because the sub-question/ super-question relations they include encode the instrumental relationships between different deliberative tactics. Where we are interested in comparing the ways that agents reason about an issue, we then do well to turn to the discourses that these agents participate in concerning this issue and compare the structured contents of their question stacks. More ambitiously, our information-structural tools suggest a way of zooming out to look at deliberative strategy variation around a particular issue across a whole polity. This possibility brings the theme of information structure into contact with issues

in democratic theory, specifically with ideals of public reason and justification as elaborated in the work of Rawls and Habermas. Whether we conceive of public reason as requiring actual inter-agent deliberation about key policy issues, or else as requiring their hypothetical capacity to offer one another reasons, looking at the distribution of deliberative strategy variation across a polity offers a powerful tool for quantifying the degree to which these ideals are satisfied. I argue furthermore that the type of polarization which this approach helps us detect, *deliberative* polarization, actually constitutes a kind of propagandistic presence in mass discourse. With this connection forged, our information structural tools then turn out also to be propaganda-detecting tools.

The implicit supposition which connects the chapters of this dissertation is that models of discourse-level information structure capture not only the contours of discourse, but of deliberation itself. With this in mind, chapter two's proposal that interlocutors' being on the same page about the information-structure of the conversation is uniquely valuable reads as the suggestion that *co-deliberation* is a central concern of signalers as such. And this suggestion seems like a scaled-down version of public reason theorists' contention that a sort of co-deliberation is a condition on the legitimacy of a policy regime, as discussed in chapter three. Taken as a whole, this dissertation's contribution is to make visible the connection between a subject matter of central concern in near-side pragmatics and issues in far-side pragmatics, epistemology, and political theory.

Chapter 1

Information Structure in Higher Order and Defective Discourse

1.1 Introduction

Discourses are collections of utterances. Scrutiny of these collections, moreover, reveals a variety of relations that obtain between these utterances: they are temporally ordered with respect to one another; some proffer content that is anaphorically tied to a part or the whole of the content of others; some are syntactically isomorphic to one another; some are made by the same conversational participant, and so on. All this is to say that discourses are *structured*, in many different ways. This paper contributes to the project of describing the level of discourse structure[1] at which *relevance* facts are visible. Specifically, I will be adding a refinement to accounts of discourse-level information structure which construe discourse as organized around questions under discussion that allows them to handle the conversational dynamics of what I call *higher order* utterances. I will be discussing specifically the imple-

[1]Roberts has since moved away from referring to the phenomenon at hand here as "information structure," partially in recognition of the fact that this same name is widely used for a different, utterance-level phenomenon(Roberts (2012b), p. 6, fn. 3). I speak here of "discourse-level information structure" in order to disambiguate while maintaining terminological continuity with Roberts' 1996/2012 essay.

mentation of the question-under-discussion approach found in Roberts (1996, 2012)[2].

In what follows, I outline the question-under-discussion approach to information structure and discuss relevance as a target phenomenon (§2). I then present a case of the sort that I argue poses a challenge to existing accounts (§3) and suggest a refinement of Roberts' model that would allow it to generate the right predictions about the relevance of higher order discourse (§4). I go on to discuss the ways in which the apparatus my solution relies upon is independently motivated by considerations around the representation of defective discourses (§5).

1.2 Discourse-level Information Structure and Relevance

1.2.1 Roberts' Information Structure

The central project of near-side pragmatics is to give a theory of context; a theory of discourse-level information structure is an elaboration on such a theory. The question-under-discussion approach to discourse structure starts from some assumptions well-established in the literature: that the goal of discourse is information exchange (e.g. Stalnaker (1978)); that this information exchange is further regulated by a norm of cooperativity which either is identical to or else subsumes a norm of relevance (Grice (1967); cf. Sperber & Wilson (1986); and that which conversational moves count as licit at

[2]Though note that thee question-under-discussion approach was independently proposed and developed in Ginzburg (1994, 1995a,b, 1996, 2012)

a particular moment is a function of the state of discourse (or conversational 'score') just prior to that moment (e.g. Lewis (1979)).

The Roberts apparatus for modeling information structure builds off of the familiar Stalnaker (1978, 2002a) model of context. On this model, a conversation has an associated common ground, which contains the set of those propositions such that every party to discourse holds them to be true (or is committed to acting as though they do). The context set is the intersection of the propositions in the common ground. When one speaker makes an assertion and the other accepts its content as true (or commits to acting as though it is), that assertion functions to add a proposition to the common ground and so (typically) to shrink the context set. The semantic value of an interrogative is the set of propositions that would count as full answers to that question (Hamblin (1973a))[3]. In the model, questions can then be thought of as partitioning the context set into cells, each of which contains a proposition which constitutes some one complete answer to the question (Groenendijk & Stokhof (1984a)).

But the above doesn't yet provide the resources with which to make sense of information structure; as far as has been specified thus far, the func-

[3]A partial answer entails the truth or falsity of at least one the set of alternative propositions denoted by the question to which it is an answer. A full answer entails the truth or falsity of every member of that set (p. 11). An example: were the question "What did Donald Davidson like to eat?", a partial answer would be "Donald Davidson liked to eat Brussels sprouts," where this comes with no implication that Brussels sprouts were the only food he enjoyed. In contrast, a full answer would exhaustively list the food that he liked: "Donald Davidson liked to eat Brussels sprouts and baked yams and avocado sandwiches and...food$_n$".

tion an assertion or question has vis-à-vis the context set is insensitive to the relevance of the speech act.[4] Modeling information structure "requires a richer notion of context" (Roberts (2012), p. 59) that reflects the way in which participants' goals in a discourse are more specific than the exchange of any old bits of information about the world.

Roberts' enrichment strategy is to have certain intentions registered in the common ground along with the content of utterances. Specifically, when a question is raised and accepted by all parties, the intention to pursue it (until a full answer is reached or found to be presently unreachable) is registered in the common ground. That question becomes the "question under discussion" and constrains which utterances can count as relevant as follows:

RELEVANCE1: A move m is relevant iff, where q is the question under discussion,

1. m introduces (directly or by implicature) a partial answer to q OR
2. m is part of a strategy to answer q

One question, q_1, is part of a strategy to answer another, q_2, if any complete answer to q_2 entails a complete answer to q_1. Where this relationship obtains, Roberts (following Groenendijk & Stokhof (1984a)) says that q_2 *entails* q_1. For example, the question "Who went to the party" entails the

[4]Unless, of course, the irrelevance of the speech act is such as to make it uninterpretable by the audience, in which case the irrelevance might indeed prevent the context set's contracting to reflect the addition of any new information.

question "Did Mary go to the Party?" because a full answer to the former entails a full answer to the latter. Where this relation obtains between q_1 and q_2 , it can also be said that q_1 is a subquestion of q_2, and q_2 a superquestion of q_1. It should be noted that one question can also be said to contextually entail another when a full answer to the first question, in conjunction with some proposition(s) in the common ground, entail an answer to the second question. Some questions are then the sub- or super-questions of others only relative to the state of the common ground. It is frequently strategic to break unmanageably large questions down into more easily answered subquestions. Because it is often the case that there are, in this sense, multiple questions being pursued at once (i.e. a superquestion is being answered via answers to its subquestions) Roberts' enrichment of the common ground involves not merely the addition of a slot for the single question of immediate interest at any given time, but a stack of these, in which strategic subquestions are placed on top of their superquestions and then whisked off the top of the stack once they're answered or deemed unanswerable.

A natural question arises as to what "raising" a question amounts to. While it sometimes occurs that a speaker will explicitly utter a question in order to engage her interlocutor in answering it, there are clearly felicitous utterances that take place without a question's being explicitly uttered beforehand. In these cases, what is the question under discussion? An ongoing literature (Rooth (1985, 1992), Beaver & Clark (2008), Roberts (1996, 2012), Simons et al. (2017)) has developed the idea that intonational focus, among

other utterance features, associates the content of a declarative utterance with a set of other propositions. This set of propositions resembles the sort of object which the above analysis associates with interrogative utterances; in this way, declarative utterances too can be seen to introduce questions, as well as answering them.

On Roberts' account, the question with which an assertion is congruent (i.e. that which it answers) is a sort of presupposition of that utterance; the utterance presupposes that the context is such that the question it answers is the one under discussion. In cases where the question with which an utterance is congruent is not yet in the common ground, that question can sometimes be admitted by way of an accommodation process (Roberts (1996, 2012) pp. 31-34). This then allows questions to enter the common ground without being explicitly uttered by any party to the conversation.

However, there are some points in discourse that are particularly ripe for new questions to be introduced as those under discussion (i.e. the very beginning, and at any juncture where the prior question has been exhaustively answered or found to be unanswerable) and it is perhaps at these points that Roberts envisions new questions being introducible via this accommodation process. Moreover, it should already be clear that we can happily accept the legitimacy of adding a new question when that question is a subquestion of the prior question under discussion, as this amounts just to the initiation of a strategy to answer the prior question.

A formal characterization of the Roberts (2012) model is as follows:

$$\text{InfoStr}_D = \langle M, Q, A, <, Acc, CG, QUD \rangle$$

1. **M** is the *set* of moves in D
2. **Q** is the *set* of questions in D (Q $\subseteq$ M)
3. **A** is the *set* of answers in D (A $\subseteq$ M)
4. $<$ is the temporal precedence *relation*, a total order on M such that:
   (a) $m_i < m_k$ iff m_i is made before M_k in D
5. **Acc** is the *set* of accepted moves in the discourse (Acc $\subseteq$ M)
6. **CG** is a *function* from M to sets of propositions, yielding for each m $\in$ M the common ground of D just prior to the making of m.
7. **QUD** is a *function* from M to ordered subsets of Q $\cap$ Acc such that, for all m $\in$ M:
   (a) For all q $\in$ Q $\cap$ Acc, q $\in$ QUD(m) iff:
      i. q <m
      ii. CG(m) doesn't entail an answer to q and q has not been determined to be unanswerable.
   (b) QUD(m) is totally ordered by <.
   (c) For all q, q' $\in$ QUD(m), if q < q', then the complete answer to q' contextually entails a partial answer to q

1.2.2 Relevance?

Having presented Roberts' characterization of relevance (RELEVANCE1), I want to step back to consider a worry about the target phenomenon. Relevance seems clearly to admit of multiple conceptions. Various projects in which some notion of relevance has been prominent have foregrounded different aspects of our pre-theoretic notion(s). It is after all not clear that accounts which identify relevance with a maximization of the ratio of information to processing cost (e.f. Sperber & Wilson (1986)) and those which identify it with the quantity and improbability of an utterance's conveyed information (Bar-Hillel & Carnap (1953), van Rooij (2003)) are targeting the same phenomenon. And it's less plausible yet that the two foregoing types of theories are targeting exactly the same thing as either the Relevance logician's variable sharing principle (e.g. Anderson & Belnap (1975)) or the Gricean's Maxim of Relation (Grice (1967)). This multiplicity raises the specter of a catastrophic methodological confusion: we will here be in the business of considering intuitions about relevance as data to which a theory of relevance must be commensurate, but if the phenomenon tracked by these elicited intuitions is unstable, the data will not in fact present a target susceptible to one unified theoretical treatment. So much, goes the worry, for any objection to an existing theory of relevance based on such data, or for whatever theory we end up with on the basis of an adaptation to such an objection.

But that there is such a range of claims, both pre-theoretic and technical, made upon the nature of relevance (or at least upon the right to use the

term) is ultimately unproblematic for our project. What motivates Roberts' project is the promise that an account of relevance will allow for an explicit representation of the inferences behind, say, implicature derivation or ambiguity resolution. In the case of implicature derivation, relevance is presumed to be involved in the way suggested by Grice; that is, it's absence in an utterance triggers the search for an implication, and its presence is a necessary condition on a proposition's presence in the space within which the search for an implicatum takes place[5]. In the case of ambiguity resolution, relevance is a feature that disposes one parse (if syntactic ambiguity) or sense (if semantic) to be favored over another. These motivating uses of the term relevance provide a functional characterization of the phenomenon for which we're here interested in giving an account; that is, we are targeting specifically the feature that does this work in linguistic inference.

Moreover, the functional characterization of relevance provided by the Gricean story about implicature derivation suggests an oblique way of gathering intuitions about an utterance's status vis-à-vis the targeted property of relevance, one free of the pitfalls of a more direct elicitation method. Whereas proffering an example and asking respondents directly about whether it is relevant comes with the above worry about tracking different phenomena, the following three-step method provides a principled way of gathering data concerning specifically the notion of relevance in which we're interested.

[5]Except in the case of so-called "actual violations" of the maxim of relation. See fn. 6.

The first step is to create a coherent discourse segment with a final utterance, u. First ask the hypothetical respondent whether u is generally acceptable/ unmarked/ felicitous. A negative response here is evidence that u is not licit, though it doesn't isolate irrelevance as the reason for this. However, where we assume that relevance, in the targeted sense, is a necessary condition on general felicity, a positive verdict amounts *a fortiori* to a verdict of u's relevance.

Second, where u survives this first step, the respondent is then asked:

1. "What did the speaker mean to get across in making u?" OR
2. "Is it plausible that the speaker meant to get i' across in making u?"

Because a condition on a proposition's being arrived at as an implicatum of a felicitous utterance is that it satisfy the maxim of relation[6], if a felicitous utterance, u, made at a time, t, plausibly generates an implicature with content i, then some part of i must be relevant at t.

The above qualification about *part* of i being relevant anticipates the necessity of the third step of this elicitation method. It is widely appreciated that a single utterance can communicate more than one proposition and that these propositions may"differ in their primacy"(Tonhauser (2012), p. 1), with some being "at-issue" and others "not-at-issue". For instance, non-restrictive

[6]At the end of Section 4 I'll cop to the fact that this isn't strictly true on the Gricean account. At that time, however, I'll also provide a reason to think that the simplification at this stage was harmless

relative clauses, epithets and conventional implicatures (see e.g. Potts (2005)) generally encode secondary information. Not-at-issue content, it turns out, typically isn't relevant. In order for our method to isolate that content which a respondent can be taken to have deemed relevant, it will be necessary to strip not-at-issue out from the yield of the previous step. This can be done through the application of diagnostics like projection (Simons et al. (2010, 2013)) and assent/dissent tests (Tonhauser (2012)).

We should regard a bit of content's surviving these three steps as evidence of its relevance, and this content's status as relevant as data to which an account must be commensurate.

In the interests of sharply delineating the boundary between data and theory, we note the following: the worry that started this section suggested that a principled way of isolating relevance data would have to have some basic characterization of relevance built in. The three step elicitation method just given amounts to a stipulation that relevant content is that which is felicitous, speaker-meant, and at-issue. An account of discourse-level information structure like Roberts' is a *substantive* account of *why* some content is this way.

One worry about maintaining the independence of data from theory here is that at-issueness is invoked in our stipulative characterization of relevance but is elsewhere stipulatively connected with the QUD-framework (i.e. Simons et al. (2010)); assuming the transitivity of this connection, this forges a stipulative connection between our target notion of relevance and the QUD-

based explanation of this phenomenon, whereas this relationship is supposed to be substantive. I think this apparent triviality is remediable if we here characterize at-issueness in the employed in, say Tonhauser (2012). That is, we can *define* at-issueness here informally as simply the state of being primary or "the point" of the utterance, or else operationally as that which survives projection and assent/ dissent diagnostics, thereby severing the stipulative connection and leaving room for a substantive one.

1.3 Problem Cases

Consider the following case:

COOKIES: Alexia and Bridget are out running errands before going to join other friends. Bridget happens to know that Alexia is in love with one of the friends they will be seeing, Cara. However, both Bridget and Alexia know that Cara is not romantically available. Bridget and Alexia have not discussed Cara in any connection thus far today. While at the bakery, Alexia says:

A: What kind of cookies do you think Cara likes?

B1: Oh, maybe oatmeal?

A2: No, too healthy.

B2: Oh, well in that case they have those double fudge ones?

A3: Too sticky!

Now imagine the following possible responses by Bridget:

B3i: Alexia, this conversation is annoying.

B3ii: Alexia, Cara has a girlfriend.

B3iii: Honestly, wouldn't roses be better?

To my ear, B3i-iii are felicitous, if perhaps unwelcome by Alexia. Moreover, note that B3i-iii come after Bridget has, over the course of several utterances, acquiesced to Alexia's line of questioning; "what kind of cookies do you think Cara likes?" was accepted as the question under discussion. But note that B3i-iii do not clearly answer this question.

That these are obviously answers to the question wouldn't be a counterexample to the above account if either of them implicated something that did. Grice characterized merely apparent violations of relevance as cases where an utterance would be irrelevant if a speaker didn't have some background belief (which they intended to be inferrable) in light of which the content of the utterance was relevant to the matter at hand, thereby implicating the content of that background belief (p. 51)[7]. Roberts is likewise happy to accept an utterance as not actually violating relevance so long as it provides or implicates an answer to a question contextually entailed by that under discussion

[7]In contrast, Grice's famous example of an actual violation of relevance is one in which a party guest says something rude and her interlocutor immediately begins to talk about something else, implicating that she disapproves of the first speaker's utterance. For most discussion, see §4 of this paper

(Roberts (1996, 2012), p. 12). Below I paraphrase Grice's familiar example of a merely apparent violation of relevance, and directly beneath it I offer a reconstruction of how we might explain the derivation of the implicatum Grice had in mind using the QUD account of relevance:

D: Where can I get petrol?[8]

E: There is a garage around the corner.

Grice's original gloss on this case is as follows:

> [The utterer of E] would be infringing the maxim 'Be relevant' unless he thinks, or thinks it possible, that the garage is open, and has petrol to sell; so he implicates that the garage is, or at least may be open, etc (Grice (1967), p. 311)

And here is a gloss which respects Grice's basic interpretive moves but makes explicit a question-under-discussion account of relevance:

1. E does not answer the question raised by D (NB: it would have been a direct answer if it had read "At the gas station around the corner").
2. If E is nonetheless assumed to be cooperative and to be consistent with having accepted D as the question under discussion, it must be that

[8]Grice's original has the initial utterance as "I'm out of petrol." I've altered it so as to make the Roberts-style derivation briefer, but I expect that, with a certain pattern of intonational focus assumed and a longer derivation, the machinery at hand could nonetheless do the job.

there is a piece of information in light of which E can be seen to further a strategy to answer D.

3. In light of the fact that petrol is sold at garages that are open and haven't run out of stock (which is a proposition in the common ground), one strategy for answering D would be to ask, "Where is there a garage that is open and not out of stock?"

4. "There is a garage around the corner" (the proposition directly said in E) still doesn't quite provide an answer to this question, but the closely related proposition "there is a garage around the corner that is open and isn't out of stock" does.

5. The supposition that "there is a garage around the corner that is open and isn't out of stock" is a proposition meant by E would render E relevant.

6. E implicates that the garage around the corner is open and has petrol in stock.

Note that step 3 relies on an accommodation of an elaboration on the QUD stack, and that this step is critical to the ultimate derivation of the implicature; this is a step that Grice's original gloss doesn't explicitly include. What it suggests is that the derivation of conversational implicatures on the basis of the Maxim of Relation's being flouted requires not just the search for an answer to the question under discussion different from what is said, but a search for a new (if strategically related) question under discussion.

Returning to my putative problem cases then, the fact that B3i-iii aren't obviously answers to the question under discussion doesn't by itself mean that they fail to satisfy Roberts' conditions on relevance. If a derivation like the one above can show that they constitute a partial answer to A, or else that they implicate one, these cases would be no counterexample at all. But this is not possible. Specifically, there's no reasoning akin to that performed in step 3 of the sample derivation above which could be applied in the problem case: provided that Alexia and Bridget don't have some very unusual mutual beliefs, there is no fact in the common ground (or retrievable for accommodation into the common ground) which, taken in conjunction with A, would contextually entail a question to which B3i-iii were answers.

B3i (depending on intonational focus, *inter alia*) can be seen as an answer to questions like "What is this conversation like?" or "What is annoying?". It is not that it would be impossible for these questions to be contextually entailed by A; if a proposition in the common ground was that Bridget and Alexia having an annoying conversation entailed that Cara's favorite cookies were chocolate chip, then this fact in conjunction with A would entail a question to which B was an answer. But among the problems with this interpretation is the fact that we can judge B relevant even where it is stipulated that there is no such fact in the common ground. B itself just isn't even a partial answer to A, and nor does it clearly implicate something that is.

B3ii and B3iii, while similar to B3i in their capacities as counterexam-

ples, raise some additional complexities it will be worth our while to attend to. Like B3i, B3ii answers neither A nor any question that we need suppose A to contextually entail in order to find B3ii felicitous. It plausibly answers questions like "Does Cara have a girlfriend?" and "What does Cara have?" and, absent a common ground which includes the proposition that people with significant others prefer snickerdoodles, neither of these questions is contextually entailed by A. B3ii might strike us intuitively as implicating something like "Wooing Cara with cookies is pointless/ inappropriate/ not likely to succeed" or else perhaps that "*Talking* about wooing Cara with cookies is pointless/ inappropriate/ not likely to succeed." But these questions, certainly, are not contextually entailed by A. B3iii likewise doesn't answer the question at hand, and like B3ii seems to dismiss A as not worth discussing; however, whereas B3ii rejected any line of conversation that involved scheming about how to woo Cara, B3iii is a less profound critique; it merely suggests that cookies are the wrong tactic (in contrast, at least, with flowers)when it comes to Cara , and therefore that *talking* about which cookies to get Cara is misguided.

The key observation we should draw from this examination of B3i-iii is that comments concerning the merit of the strategy embodied by the question stack are generally not answers to the questions on that stack, or answers to strategic subquestions which might be added to the stack.

B3i-iii are instances of what we'll call a higher order utterance (HOU). Higher order talk concerns the character, strategy, or desirability of the discourse structure itself, and a distinctive feature of such utterances seems to

be that they can be relevantly inserted at virtually any point in the discourse whose merit or organization they concern. B3i-iii are higher order utterances because they either state or generate an implicature to the effect that the matter being discussed is not one worth talking about, and this is a statement about the desirability of the current conversation.

The category I want to circumscribe with the phrase "higher order utterance" is related to other categories of speech which have accrued associated literatures, but is not, I don't think, precisely coextensive with any of them. For instance, higher order discourse is related to what, in discourse analysis and language education literatures, is called *metadiscourse*, which includes all the heterogenous phenomena that might be thought of as "'discourse about discourse' or 'talk about talk'" (Hyland (2005), p. 16). These literatures have interpreted this to include, for instance, any use of epistemic modals (e.g. "It must have rained") or propositional attitudes attributions (e.g. "I believe that he will arrive soon") as these function to organize or comment on other contents of a text or spoken discourse. But the category of HOUs doesn't include mere commentary on other expressed content, at least not as such; the term "higher order utterance" is being used here to refer specifically to utterances about the discourse's information structure (i.e about which pieces of information are in the common ground, and what is on the question stack), not just about any bits of content that may have been proffered, with assertoric force or otherwise, in the course of that discourse. I therefore take HOUs to be a proper subcategory of metadiscourse.

The category of HOUs is likewise related to that of *metacommunicative interaction* (MCI), where

> By metacommunicative interaction one means any interaction that comments about the communicative process underlying an utterance. More specifically, the commonest MCI utterances are: acknowledgements that an utterance has been understood, clarification requests (CRs) in which an unclear aspect of the utterance is queried, and corrections, where indications are provided of erroneous assumptions concerning naming, concepts associated with predicates etc. (Ginzburg & Macura (2007), p. 288)

I take the categories of MCI and HOUs to overlap, but they differ in two respects. First, MCI involves interlocutors working to get on the same page about the *description* of the context, while higher order utterances can be descriptive but can also be normative, i.e. can concern how the discourse *should or should not* be. Second, MCI includes attempts to verify, acknowledge and correct the *language* that was used in an utterance, i.e. "Did you say 'Bo'?" "Did you mean to say 'Bro'?" (Ginzburg (2012) p. 149, Purver et al. (2001)). In contrast, we are restricting the category of higher order discourse to those utterances which concern the information structure of the discourse, and so excluding utterances which concern features of the discourse upstream from the expression or comprehension of content.

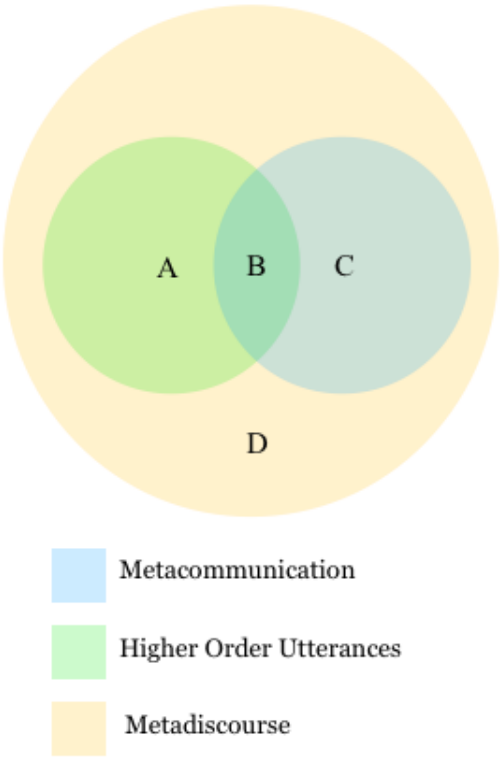


Figure 1.1: Higher Order Discourse and Related Phenomena

In the above figure, Region A includes what I've suggested we think of as normative HOUs; these are questions and assertions about how the question under discussion stack should be (e.g. "Should we be talking about this?"). Region B is populated by utterances that count as both metacommunicative interaction and HOU; these are *descriptive* HOUs which are clarification requests, acknowledgements, and corrections *about the QUD stack* (e.g. "What were we talking about?", "How is this related that last thing?") . Region C contains metacommunicative utterances that comment on elements of the dis-

course upstream from content (e.g. "Did you mean to say..."). Finally, region D includes metadiscourse which comments on the speaker's relationship to other content expressed in the discourse without commenting on the discourse *per se* (e.g. "I think I'll go to the park", "The Cubs might win").

This terminological issue out of the way, let's be clear on the nature of the foregoing relevance conditions' inadequacy when it comes to HOUs. The issue is that the model doesn't have the resources to characterize higher order utterances, and so doesn't allow for the articulation of relevance conditions that respect their special character. That there are constraints on moving between first order questions is intuitive and is, indeed, a basic supposition of Roberts' project. The higher order cases which I have discussed demonstrate that movement up to second order discourse is relatively unrestricted. To motivate a third general principle of discourse movement, consider the following discourse amongst interlocutors C, D, and E:

GOSSIP

F1: I heard that Mary's been looking for jobs elsewhere.

G1: I heard the same from Lisa in accounting.

H1: We shouldn't be talking about this.

F2: Do you really think it's inappropriate to engage in a bit of harmless gossip?

G2: Her supervisor's also been looking worried.

The intuition here is not that G2 is a totally bizarre or implausible utterance, but that it *is* irrelevant *unless* it is given an implicative reading, on which its surface irrelevance triggers a search for an implicatum. Indeed, the case strongly resembles Grice's example of an implicature-generating "actual" violation of the maxim of relation; speaker G, one might infer, is uninterested in the second order question which $H1$ and $F2$ are attempts to answer (i.e. "Is it appropriate to talk about whether Mary is looking for jobs elsewhere?"). G2's reversion to the first order question "Is Mary looking for work elsewhere?" generates this conclusion as an implicature[9]. All of this suggests that moving down through orders of discourse, when a higher order question under discussion has not yet been answered or deemed to be unanswerable, is something that our model should predict to be irrelevant.

In summary, I think that relevance intuitions suggest as coarse desiderata on a set of relevance conditions that they preserve the following three principles:

Restricted Lateral Movement (RLM): there are constraints on moving between questions of the same order.

Unrestricted Upward Movement (UUM): there are no constraints on moving to a question one order up from that addressed by the immediately prior move.

[9]Notably, G2's implicatum is actually third order, which the relevance conditions I offer below predict to make the utterance relevant. For a fuller gloss on G2's dynamics, see fn. 14

Restricted Downward Movement (RDM): there are constraints on moving to a question of a lower order than that that addressed by the immediately prior move.

This is clearly only very schematic. However, it will be serviceable to have introduced this distinction among desiderata. Specifically what it allows us to point out is that the original model is designed to satisfy RLM, but that it has no resources for respecting UUM or RDM. The next section explores a way of enriching the model in a way that allows us to articulate conditions that do.

1.4 Solutions

Normative HOUs like B3i-iii are naturally understood as suggesting that the question under discussion stack is inconsistent with the broader aims or values of at least some of the conversationalists. This insight seems to suggest that, in order to capture what is distinctive of HOUs, a model will have to include a richer picture of interlocutor goals. Indeed, in Roberts (2012b) and Roberts (2004), we see an acknowledgement of the way in which the "discourse goals" represented by elements of the stack are situated within a larger set of all goals had by any participant in the discourse. She points out that there is a notion of relevance more general than the one targeted by her earlier account, on which "A move m is **Relevant** at a given point in a collaborative, task-oriented interaction if and only if it promotes the achievement of an accepted goal of the interlocutors." (Roberts (2012b), p. 9, bolding Roberts').

The contrast between this more general notion of relevance and that aimed at by the Gricean maxim of relation is evident: on the more general notion, an utterance might count as relevant even though it does nothing to address the question currently under discussion. That this broader notion of relevance has some currency in discourse is evident. Consider:

> WASP: We're in the middle of a discussion of the differences between early and late Wittgenstein. Suddenly you say "Oh, there's a wasp on your arm."

Knowing that a wasp has landed on my arm does nothing to answer our erstwhile question under discussion of course, but this would hardly matter in such a case; the goal of my not being stung by a wasp, which you presume me to hold and cooperatively adopt as well, is mutually understood to take temporary priority over the goal of reaching some insight into the *Philosophical Investigations*.

But higher order utterances are not like the one made in WASP; it's not that they are relevant in some extended sense of the term, but in the very Gricean sense targeted by Roberts' account. Their acceptability is not a consequence of some shift in our environment, say, inducing a previously non-salient shared goal to leap to the foreground. HOUs are relevant because of the goals already salient in the conversation, not because of some other domain goals.

Nonetheless, stepping back and looking at the question stack as constituting a special subset of the broader set of interlocutor goals is necessary in order to appreciate how the model needs to be augmented in order to reflect the relevance of HOUs. An utterance can be related to a question as an answer (full or partial) to either it or one of its subquestions. Questions can be related to one another via sub-/super-question relations. But neither answerhood nor sub-/super-question relations take non-question goals as a relatum. HOUs, meanwhile, sometimes have as their upshot that a bit of presumed instrumental reasoning which the current information structure is predicated on is faulty. Here is a sketch of a solution that is responsive to this view of the problem; while it won't be the solution I ultimately adopt, considering it will be instructive.

In line with the suggestion in Roberts (2012b) and Roberts (2004), add to *InfoStr* a set of all the discourse participants' shared goals[10] and an instrumentality relation on this set of goals. Add also a function that helps characterize the way that the goals on the question under discussion stack figure in the discourse participants' overall structured set of goals; we might think of this function as yielding the "G-profile" of a question. On this suggestion, InfoStr_D becomes $\langle M, Q, A, <, Acc, CG, QUD, G, R, S\rangle$ with new elements characterized asfollows:

8. **G** is the *set* of all goals shared by discourse participants.

[10]Roberts' actual suggestion involves adding a set of all goals, shared or not. This is a complexity I abstract away from in this sketch.

9. **R** is a binary *relation* on G such that xRy just in case satisfying x is (perceived as) instrumental[11] in satisfying y.

10. **S** is a *function* from M to subsets of the transitive closure of R on G such that the first relatum in each tuple is a question on the QUD stack at m, i.e. to $\{\langle x, y\rangle : \langle x, y\rangle \in$ transitive closure of R on G and $x \in QUD(m)\}$

Note that Q is a subset of G and that the subquestion relation on Q will be a subset of R on G; that is, where one question, q_1 helps answer another, q_2, satisfying the goal of answering q_1 is instrumental in satisfying the goal of answering q_2. Having added these elements to the model, we might revise the original relevance conditions as follows:

> RELEVANCE2: A move m is relevant iff m either introduces a partial answer to the question under discussion, q, (m is an assertion) or is part of a strategy to answer q (m is a question) or m alters the G-profile of a question currently under discussion (i.e. $S(m_n) \neq S(m_{n-1})$).

While no doubt amplified by the terseness of the sketch, this approach's flaws are, I think likely to endure the sketch's refinement. First, this approach won't address what I've called descriptive HOUs; an utterance which serves merely to signpost or acknowledge an already accepted discourse structure

[11] I am giving no very detailed account of instrumentality here; I do assume that x's being instrumental in achieving y does not entail that achieving x is a necessary or sufficient condition on y

(e.g. "To review, we've been debating whether to approve the school board's budget and will next discuss the results of the by-election") doesn't alter the way that the goals on the question under discussion stack figure in the discourse participants' overall structured set of goals. Second, this approach doesn't by itself address how HOUs should be dealt with by the QUD function. While the approach can respect RLM and UUM, it doesn't reflect the way in which the move to higher order discourse constrains what comes thereafter; that is, it doesn't respect RDM. Finally, it introduces extensive additional apparatus. This line of criticism is perhaps not too pressing; the addition of G, at least, needn't seem ad hoc if we appreciate the way it offers, *inter alia*, a way of reflecting the effects of requests(Roberts (2004), p. 215). But the solution I will propose makes G unnecessary for at least the purposes of representing the information-structural dynamics of HOUs.

My augmentation of the model removes the QUD function and adds an infinite set of QUD-like functions to the tuple that defines the information structure of a given discourse:

$\text{InfoStr}_D = \langle M, Q, A, <, Acc, CG, QSet \rangle$.

The functions in QSet (QUD_1, QUD_2... QUD_n) admit of general definition:

7*. **QSet** is the *set* of all *functions*, QUD_n, $n > 0$ and $n \in \mathbb{N}$, such that:

(a) QUD_n is a function from M to ordered subsets of $Q \cap Acc$ such that, for all $m \in M$:

i. For all q $\in$ Q $\cap$ Acc, q $\in$ QUD_n(m) iff:

I. q <m

II. CG(m) doesn't entail an answer to q and q has not been determined to be unanswerable.

III. For all QUD_n such that n > 1:

A. all p such that p $\in$ q contextually entail a full or partial answer to the 'big question,' or BQ_n: "What is QUD_{n-1}(m) like?"

(b) QUD_n(m) is totally ordered by <.

(c) For all q, q' $\in$ QUD_n(m), if q < q', then the complete answer to q' contextually entails a partial answer to q.

Naturally, the subscript on a QUD function corresponds to the level or 'order' of discourse with which the stack is associated. That is, 2nd order utterances will answer questions generated by QUD_2 and so on. The picture here is one on which, even prior to the initiation of any higher order discourse, there are multiple (infinite!) single-question QUD stacks always present, with their initial question, as it were, auto-generated in a way that's responsive to the elements of lower order stacks at that point in discourse.

It may seem odd that this model countenances many more orders of discourse than any normal discourse will involve; my expectation is that advancing as far as fourth-order discourse is rare and going beyond that practically unheard of. I regard this as analogous to the way that a grammar may

countenance, say, center-embedding to an infinite depth, even though depths past three are virtually nonexistent in corpora. In that case, as in the information structural project here, the expectation is that some distributional data will be explained by limits on processing capacity or some other constraint on natural language use, rather than by the theory presently being offered. When it comes to explaining the dearth of utterances of orders greater than 3, one feels inclined to advert not only to speakers' limited processing capacity, but also to the limited interest that such discourse will hold for almost anybody.

A second worry might be that there's no reason to think that the initial questions of HO stacks are elements of Q ∩ Acc. If this worry is derived from a concern that such questions just aren't "moves" in the discourse at all, and so not elements of M or, *a fortiori*, of Q, it should be susceptible to easy dismissal. The success of the original proposal requires that we allow a question to be an element of Q as a result of some accommodation process, as well as by explicit utterance. The same thing must apply to higher order questions. However, one might worry that the initial questions on HO QUD stacks are still importantly dissimilar from questions that enter the first order QUD stack via accommodation. Specifically, the initial questions on HO stacks have never even been even passively "accepted," and so might be regarded as failing to be elements of Q ∩ Acc because of their not even being elements of Acc. That an initial question might exist as the first constituent of a QUD stack before the first move of its proprietary discourse level was made, and so prior to any sort of acceptance of it seems possible is, however, entirely consistent with the

original picture. On this picture, the initial question of QUD_1, "What is the world like?", is a sort of boilerplate feature of contexts, existing before anything is said at all. Whether Roberts (or, indeed, Stalnaker) would reconcile this aspect of the view with 7 by arguing for a notion of acceptance on which such boilerplate elements of discourse are tacitly accepted, or by adding an exception to 7, I assume that I would be safe in making exactly the same move when it comes to the initial questions on HO stacks.

With the above alterations made to the modeling apparatus, it will perhaps be evident how I want to proceed to alter RELEVANCE1 in a way that respects UUM and RDM. Recall that the original conditions on relevance were:

RELEVANCE1 A move m is relevant iff, where q is the last element of (QUD(m)),

1. m introduces (directly or by implicature) a partial answer to q OR
2. m is part of a strategy to answer q

I suggest the following revision:

RELEVANCE3: A move m_x is relevant iff, where m_{x-1} was of order n and $\text{n} \in \mathbb{N}$,

1. m_x introduces a partial answer to the last element in $(\text{QUD}_n(\text{m}_x))$ OR

2. m_x is part of a strategy to answer the last element in($QUD_n(m_x)$)

   OR

3. m_x introduces a partial answer to the last element in ($QUD_{n+1}(m_x)$)

   OR

4. m_x is part of a strategy to answer the last element in ($QUD_{n+1}(m_x)$)

   OR

5. if m_{x-1} completely answered (or established as unanswerable) the last element in ($QUD_n(m_{x-1})$) and n > 1,

   (a) m_x introduces a partial answer to the last element in ($QUD_{n-1}(m_x)$)

   OR

   (b) m_x is part of a strategy to answer the last element in ($QUD_{n-1}(m_x)$)

1-4 reflect ways in which an utterance can be relevant by answering or strategizing to answer questions on the same stack as the prior utterance in the discourse, or on the stack above this one. In short, 1-4 are meant to respect RLM and UUM. 5 reflects the ways in, and circumstances under, which an utterance can be relevant by answering or strategizing to answer a question on a lower order stack than the one answered by the prior utterance. This allows the model to respect RDM. Cases like B3i-iii then plausibly satisfy 3 or 4, either of which our new model predicts to be sufficient for relevance. This is a correct prediction.

Note that, unlike RELEVANCE2, this approach 1) deals appropriately with both descriptive and normative HOUs, 2) addresses how QUD deals with HOUs, and 3) uses a version of the same resources needed in order to appropriately handle first-order discourse to deal with higher order discourse.

A potential worry concerns how these conditions connect with the exact specification of the "big-question" which we're considering as a boilerplate feature of higher order question stacks. First, it should be clarified that "QUD_{n-1} is to be read de dicto; were it read de re, descriptive HOUs would be trivial. Next, it might be worried that this formulation of the question excludes as answers those utterances which clarify or acknowledge the content of the common ground (e.g. "so, you're saying Ben *did* attend?"), which I had suggested do number among HOUs. Whether the above formulation of the "big question" does so depends on a detail about the nature of the elements of Q that went undiscussed earlier. Where an uttered question is identified with a set of propositions (and, equivalently, with the way this set of propositions partitions a set of possible worlds), is the matter of *which set of propositions that is* sensitive to the content of the common ground? That is, are the elements of Q partitions i) on the total set of possible worlds, or ii) on the context set? There are considerations that point in both directions, and settling this detail is not generally necessary for this essay's purposes. However, strictly speaking the above candidate "big question" is not neutral with respect to i and ii. In order for the proposed big question to be answerable by utterances that describe the common ground, we need to suppose that the elements of

Q (and so the output of QUD) are in accordance with ii. This is because, where the elements of QUD(m) are partitions on CG(m), queries about what QUD(m) is like will be partially answered by information about what CG(m) is like, thereby bringing the otherwise excluded descriptive HOUs back into the fold. We can however maintain neutrality between i and ii without loss of our basic proposal, by offering an alternative to the above proposal which implementations committed to i can adopt. On this alternative, the big question is "What is QUD'$_{n-1}$(m) like?" where we simply define QUD' functions so that they map elements of the QUD function of the same order (which will be partitions on the entire set of worlds) onto partitions of the common ground at m.

How does the QSet approach relate to the insight we started this section with, that HOUs (of the normative variety anyway) characteristically comment on the relationship of the first order stack to the broader aims or values of at least some the conversationalists? Note that, where answering a question, q, is instrumental in satisfying some further goal, g, then when q is on the first order stack, the question "Is g appropriate/ worthwhile?" is a subquestion of "Is q appropriate/ worthwhile?". That is, a subquestion of BQ_2 is "is QUD_1 appropriate/ worthwhile?" and where it a proposition in the common ground is that the question(s) on QUD_1 was motivated by g, and that a goal is worthwhile only insofar as the further goal that motivates it is worthwhile, the question "Is g appropriate/ worthwhile?" is contextually entailed as well. In this way, there is space for questions' connections to other goals to be

reflected in our handling of HOUs without embellishing *InfoStr* to include elements specifically for the representation of these other goals– so long as we have resources for representing higher order questions.

To see this approach to HOUs in action, let's walk through the way that our revised model and relevance conditions handle B3i-iii[12]. Prior to any of B3i-iii, the state of the common ground and QSet are as follows:

- Common Ground: CG(A3)
- QUD_1: $\{BQ_1$, "What kind of cookies does Cara like?$\}$
- QUD_2: $\{BQ_2\}$
- QUD_3: $\{BQ_3\}$
- ...

B3i-iii update these elements of *InfoStr* as follows:

B3i: Alexia, this conversation is annoying.

Step 1: Add "Is (QUD_1 such that) this conversation (is) annoying?" to the top of QUD_2.

- Common Ground: CG(A3)[13]

[12]I am assuming here that a single utterance effects a multi-step update (see Murray (2014)) and that each step must satisfy RELEVANCE3.

[13]I am holding the common ground fixed through this step though, in fact, at some point in the update effected by this utterance, the proposition that B3i was uttered will be entered into the common ground as well

- QUD$_1$: {BQ_1, "What kind of cookies does Cara like?}
- QUD$_2$: {BQ_2, "Is (QUD$_1$ such that) this conversation (is) annoying?"}

Step 2: Add to the common ground the proposition that the conversation is annoying and remove "Is (QUD$_1$ such that) this conversation (is) annoying?" from QUD$_2$ to reflect its having been answered.

- Common Ground: CG(A3)$\cup${"This conversation is annoying"}
- QUD$_1$: {BQ_1, "What kind of cookies does Cara like?}
- QUD$_2$: {BQ_2, "Is this conversation annoying?"}

B3ii: Alexia, Cara has a girlfriend.

Step 1: Add to QUD$_2$ the question "Is QUD$_1$ appropriate/ good/ useful?"

- Common Ground: CG(A3)
- QUD$_1$: {BQ_1, "What kind of cookies does Cara like?}
- QUD$_2$: {BQ_2, "Is QUD$_1$ appropriate/ good/ useful?"}

Step 2: Add to the common ground the propositions that the reason Alexia is interested in buying Cara cookies is an interest in winning her favor romantically and that Cara's having a girlfriend speaks against this effort's appropriateness or likelihood of success.

- Common Ground: CG(A3)∪{"Alexia is interested in buying Cara cookies because of an interest in winning her favor romantically", "Cara's having a girlfriend speaks against this effort's appropriateness or likelihood of success."}
- QUD_1: {BQ_1, "What kind of cookies does Cara like?}
- QUD_2: {BQ_2, "Is QUD_1 appropriate/ good/ useful?"}

Step 3: Add to QUD_2 the question "Does Cara have a girlfriend?"

- Common Ground: CG(A3)∪{"Alexia is interested in buying Cara cookies because of an interest in winning her favor romantically", "Cara's having a girlfriend speaks against this effort's appropriateness or likelihood of success."}
- QUD_1: {BQ_1, "What kind of cookies does Cara like?}
- QUD_2: {BQ_2, "Is QUD_1 appropriate/ good/ useful?", "Does Cara have a girlfriend?"}

Step 2: Add the proposition "Cara has a girlfriend" to the common ground and remove "Does Cara have a girlfriend?" from QUD_2 to reflect its having been answered.

- Common Ground: CG(A3)∪{"Alexia is interested in buying Cara cookies because of an interest in winning her favor romantically", "Cara's having a girlfriend speaks against this effort's appropriateness or likelihood of success."}

- QUD$_1$: {BQ_1, "What kind of cookies does Cara like?}
- QUD$_2$: {BQ_2, "Is QUD$_1$ appropriate/ good/ useful?", "Does Cara have a girlfriend?"}

B3iii: Honestly, wouldn't roses be better?

Step 1: Add to QUD$_2$ the question "Is QUD$_1$ appropriate/ good/ useful?"

- Common Ground: CG(A3)
- QUD$_1$: {BQ_1, "What kind of cookies does Cara like?}
- QUD$_2$: {BQ_2, "Is QUD$_1$ appropriate/ good/ useful?"}

Step 2: Add to the common ground the propositions that Alexia is interested in buying Cara cookies is an interest in winning her favor romantically.

- Common Ground: CG(A3)$\cup${"Alexia is interested in buying Cara cookies because of an interest in winning her favor romantically"}
- QUD$_1$: {BQ_1, "What kind of cookies does Cara like?}
- QUD$_2$: {BQ_2, "Is QUD$_1$ appropriate/ good/ useful?"}

Step 3: Add to QUD$_2$ the question "would roses be better (for the purposes of winning Cara's favor)?"

- Common Ground: CG(A3)∪{"Alexia is interested in buying Cara cookies because of an interest in winning her favor romantically"}
- QUD_1: {BQ_1, "What kind of cookies does Cara like?}
- QUD_2: {BQ_2, "Is QUD_1 appropriate/ good/ useful?", "Would roses be better (for the purposes of winning Cara's favor)?")}

In all three cases, the new model is able to reflect the way in which the target utterance is intuitively shaping the information structure, and RELEVANCE3 generates the correct verdict: all three utterances are relevant.

Having now laid my positive proposal on the table, I will confess to a simplification that I made all the way back in §2. I represented it as an article of Gricean doctrine that:

> If an utterance, u, made at time, t, plausibly generates an implicature with content i, then i must be relevant at t.

But this is not quite right– Grice thought this was untrue in cases of "actual" violations of the maxim of relation. I've distorted the Gricean record somewhat because I think that Grice's distinction between "actual" and "merely apparent" violations of relation is unmotivated, and in fact damaging to his own theory. Note that in all cases of implicature except for those of so-called "actual" violations of Relation, the Gricean says 1) that a broken maxim triggers the search for an implicatum, and 2) that the search for this

implicatum is then guided by the presumption that *it* will be a proposition that is consistent with the apparently broken maxim. In the case of actual violations of Relation, however, the breach allegedly has the first effect without the second. What then constrains the search space in a way that would make it remotely possible to reason one's way to the implicatum (or even a modest range of plausible ones)? As always, some requirement that the implicatum be such that the implicative utterance itself is plausibly helpful in reaching it may constrain the search space to some extent. But without reference to information structural features of broader context that make some proposition more likely than others (doing which would smuggle 2 back in), it's very difficult to see how implicatures are supposed to be derivable in cases of actual violations of relation. This makes for an exacerbated case of Griceanism's habitual troubles with incalculability (see e.g. Davis 1998).

Given that distinguishing a class of implicature for which 2 is dropped incurs this theoretical cost, why does the Gricean do it? Perhaps to appear attentive to a disparity in the relevance (pre-theoretically conceived of) of the implicata in the "merely apparent" and "actual" cases. But I confess that I don't myself see the implicatum in, say, Grice's paradigm case of an actual violation of Relation (see fn 5) as intuitively "irrelevant". So this rationale for the distinction seems likewise uncompelling. To the charge that I have failed to target a notion of relevance exactly like what Grice had in mind (and so have targeted something different from what Roberts probably had in mind) when he composed his maxim of relation, I suggest then that I've

adhered to the spirit of his project, rather than the letter. Where the maxims were supposed to have genuine explanatory power with respect to implicature derivation, the distinction between merely apparent and actual violations of Relation is senseless. The account I've given allows a distinction to be drawn without incurring the theoretical hardships that Grice's did; what Grice called actual violations of Relation are assimilable to the category of merely apparent violations, and unique only insofar as their implicata are higher order[14]

GOSSIP

F1: I heard that Mary's been looking for jobs elsewhere.

G1: I heard the same from Lisa in accounting.

H1: We shouldn't be talking about this.

F2: Do you really think it's inappropriate to engage in a bit of harmless gossip?

G2: Her supervisor's also been looking worried.

[14]Incidentally, this results in an interesting gloss on GOSSIP, which we noted earlier seemed to involve something like what Grice called an actual violation of relation. It was apparent that F1 and G1 were part of first order discourse, and H1 and F2 elements of second order discourse. G2, which on its surface reverts to first order discourse, seems to carry the implication that the foregoing discourse wasn't worth having. This suggests that G2 effects the following multi-step update of the discourse's structure: 1) it adds to the top of QUD_3 the question "Is the line of questioning on QUD_2 worthwhile?", 2) it answers this third order question in the negative, 3) this third-order question being deemed answered, it strips QUD_3 back down to BQ_3, 3) it strips all questions beyond BQ_2 off of QUD_2 to reflect the third order commentary just issued, and finally, 4) with all higher order discourse reduced to its boilerplate state it can address the question currently on the top of QUD_1 without actually violating RDM.

1.5 Connections with Defective Discourse

I have introduced a set of functions, QSet, into the apparatus used to model information structure and articulate relevance conditions. I want to demonstrate that having multiple orders of QUD stack is independently helpful for representing interlocutors' information states in a particular kind of *defective* discourse. A defective discourse is one marked by any deviation from PERFECT COMMUNICATION, where a discourse satisfies PERFECT COMMUNICATION iff, for each move in the discourse, the audience perfectly understands the speaker's specifically intended discourse update. A discourse might be defective because an audience is confident about the speaker's intended update, but they're wrong–call this MISALIGNMENT)–or because the audience just wasn't sure about what the speaker's intended update was–call this UNCERTAINTY). Additionally, a discourse might be defective because the speaker just doesn't have a highly specific update in mind–call this NONSPECIFICITY. Each of these three kinds of defectiveness can be manifest at both the level of propositional and structural updates[15]. Of note is that this third type falls short of PERFECT COMMUNICATION while nonetheless being pervasive in cases that we are typically happy to describe as *successful* communication. My primary interest here is in drawing attention to an information structural manifestation of this third kind of defectiveness and making some preliminary comments about how we should model it which suggest the use of the

[15]NB: As some of the foregoing discussion will make clear, all utterances effect both propositional and structural updates.

apparatus I've introduced above.

1.5.1 Information-structural Nonspecificity

Consider:

> Real estate: Kahlil and Larry stand on the sidewalk in front of their newly purchased fixer-upper property. Kahlil glances from the broken windows to the collapsing doorframe to the massively overgrown lawn. Larry says "Well?" to which Kahlil responds by gesturing broadly around them and saying "A mess!"

What has been observed about cases like Kahlil's utterance is that it's not clear what proposition an audience ought to update with on the basis of his non-sentential utterance. What is reasonably clear is that candidates include the following:

i. This house is a mess.

ii. The grounds are a mess.

iii. This whole property is a mess.

iv. Our situation is a mess.

However, it's not clear that any of these is *the* right interpretation of Kahlil's utterance; it's quite plausible in fact that Kahlil didn't mean any particular one of them, while finding all of them more or less satisfactory. That

non-sentential assertions often have this feature is an established observation (see Buchanan (2010), Fernández & Ginzburg (2002), **?**); non-sentential assertions often admit of many plausible completions which the speaker would find equally acceptable. But nor is non-sentential assertion the only possible locus for this kind of nonspecificity; the domain of quantificational expressions is often specifiable in multiple ways that yield the same extension at this world and so are equally acceptable for most practical purposes, or else yield different extensions equally consistent with the speaker's intentions. And depending on the extent of one's contextualist leanings, very many other expressions (e.g. comparative adjectives, counterfactual conditionals, propositional attitudes) may likewise yield many equally satisfactory interpretations in a given utterance.

What I am most interested in, however, is the way that NONSPECIFICITY manifests at the level of information structure:

> FLOWERS: Frantically rushing around on the morning of his wedding, Nelson asks Mariko "Have the flowers been delivered?"
>
> PARTY: Orville and Priya have both been invited to an upcoming party. They have previously discussed their concern with whether they will know anyone there. They also know the guest list is quite exclusive, and only those with invitations will be admitted. Finally, they mutually know that Quinn is among their mutual friends. Today, Orville approaches Priya and says, downcast, "Quinn didn't get an invitation to the party."

I offer two examples, because there are two distinct phenomena here. FLOWERS bears an obvious relation to REAL ESTATE; just as "the flowers" would admit of several different completions were it to occur in an assertion (the flowers that we ordered, the flowers that will be part of the wedding etc.) so too does this expression's presence in a question give rise to a multiplicity of plausible, likely-to-be-satisfactory interpretations. And because this utterance adds a question-under-discussion to the discourse, this multiplicity registers at the level of information structure. That is, Mariko would do equally well to update her information structural representation with the question "Have the flowers we ordered been delivered?" *or* with "Have the flowers that will be part of the wedding been delivered?" and Nelson might furthermore be thought not to have intended one rather than the other, and to see both as equally consonant with his intentions.

In PARTY, the issue may be less apparent; it's true that we again have a definite description ("the party"), but the multiplicity of plausible completions of this definite description is not the issue here; this just shows that the already described kind of NONSPECIFICITY is pervasive, and that it can combine with the other. What I in fact mean to draw attention to is what we might call *deep* information-structural desires which are attributable to Orville, where an information structural desire is *deep* just in case it has as its object the addition of more than one question to a single QUD stack. It may be obvious that speakers have such deep desires in the sense that they sometimes have a plan for the trajectory of the conversation; they intend to raise question x,

then question y, and so on. But making sense of cases like PARTY, as well as utterances like B3ii and B3iii, requires acknowledging that speaker intentions sometimes have as their object the simultaneous[16], rather than consecutive, addition of multiple questions. I'll spell this out in terms of PARTY and then argue that this is the right interpretation of such cases.

"Quinn didn't get an invitation to the party" implicitly adds a question to the QUD stack. In order to be principled about the way in which we infer implicit questions from assertions, we need to take any assertion that p, to at least proffer (and, lacking objections from the audience, add) the polar (yes/no) question of the form "Is it the case that p?" (cf. Simons, Beaver, Roberts, & Tonhauser (2017)). It follows from this that one question which Orville can certainly expect to see added on the basis of his utterance is q_1:

q_1: Did Quinn get an invitation to the party?

Now, note that where it is common knowledge that only those who received an invitation may go to the party, as was stipulated, q_1 is a subquestion of q_2:

q_2: Will Quinn go to the party?

And where it is common knowledge that Quinn is friends with both Orville and Priya, as stipulated, q_2 is in turn a subquestion of q_3:

[16]I take this simultaneity to consist in having multiple questions added by a single utterance, but not to speak against these questions being added in separate update steps

q_3: Which of my friends will be at the party?

Which is in turn a subquestion of q_4:

q_4: Who will go to the party?

Now suppose that Orville's ultimate goal is to discuss the answer to q_3. Moreover, his utterance was made with this goal in mind. That is, he would ideally like Priya to have responded to his utterance by updating her QUD stack with $\{q_1, q_2, q_3\}$. He would at least have *liked* to add more than one question to the stack at once.

However, it is clear that Priya might fail to have understood Orville's desired update based on his actual utterance, and this may be reasonably clear to Orville himself. This raises two questions: First, why think that the way to understand the information structural desires of such a speaker is as I've described? Second, what does this mean for the propositional attitudes of such a speaker? And finally, how do we model what goes on in this kind of case?

One might charge that, rather than understanding a speaker like Orville to be attempting to add multiple elements to the QUD stack at once, they are simply adding one (q_1) now, while maintaining an intention to add q_2 and q_3 in turn once q_1 is successfully answered and so taken off the top of the stack.

The problem with this approach is that it represents the progression through q_1,q_2, and q_3 in a way equally appropriate to three entirely unrelated questions (i.e. questions between which no subquestion/ superquestion

relations held). This fails to capture the logical relationship between these questions, and the model fails to reflect the structure in virtue of which interlocutors are often able to anticipate the ends (i.e. superquestions) for which speakers may be asking their subquestions. It also ends up generating some false negatives with respect to relevance. Consider the following variation:

> PARTY2: Priya asks, "Did Quinn get an invitation to the party?" (q_1) and Orville responds, "No, none of our friends are going."

That "none of their friends is going" does entail an answer to q_1, but also provides additional information which does not help answer q_1. Where q_1 but no other questions to which this additional information *was* an answer was on the stack, this information would be irrelevant. Only by taking q_3 to be on the stack as well do we get the relevance of this information [17].

An alternative relevance-saving account of the above discourse would suggest that Priya indeed succeeded only in adding q_1 to the stack, but Orville managed to answer q_1, implicitly add q_3, and answer it in one fell swoop. This saves Orville's utterance from seeming infelicitous. What this account still

[17] This sounds like, if anything, a classic example of a violation of the maxim of quantity, so it may sound odd that I'm characterizing it as an apparent violation of relation. Which it would count as depends on whether parts of an utterance (even of a single uttered sentence) are assessed individually or jointly for relevance. The first part ("No") of Orville's utterance does answer the question under discussion (q_1), but the latter part does not. If these parts are to be assessed for relevance separately, the second part is a violation of relation. If they are assessed jointly (i.e. the relevance of the first part is enough to render the entire utterance relevant) then the issue is that the latter part of the utterance seems superfluous for the purposes at hand, and so like a violation of quantity.

fails to reflect is that Orville's addition of q_3 is (I'll make this stipulative here, but not without thinking that this also describes a frequent reality) motivated by the inference that Priya wanted to see q_3 answered, and moreover that she likely wanted such a desire to be recognized on the basis of her utterance of q_1. In short, it seems to get something wrong to attribute the addition of q_3 entirely to Orville rather than Priya.

I'll now address the second question mentioned above. I have above been careful to speak of Priya as having *wanted*, rather than *intended* to add q_2 and q_3 to the stack. This is because it is typically thought that a constraint on intending to ϕ by ψ-ing is that one have a credence above a certain threshold that one will succeed in ϕ-ing by ψ-ing, and in the sort of case I am discussing it is an open question to Priya herself how confident she should be that she will succeed in having her audience recognize her deep simultaneous information structural desires on the basis of her utterance of q_1. To be clear, there is perhaps some fact of the matter concerning Priya's credence, and so some fact of the matter about whether she succeeded in intending. But I suggest that it is fully coherent to suppose that speakers in Priya's position don't *know* whether their hope was rational, and so don't know whether they succeeded in intending to have communicated their deep structural preferences. In fact, there's very little reason for a speaker to scrutinize the rationality of the hope that their deep information structural desires will come across; should Orville fail to grasp any more than q_1, Priya will be able to add q_2 and q_3 thereafter, and still achieve her discourse goals. This will not require making any moves

that Priya would not have had to make in the first place, were she to have explicitly unfolded her deep preferences up front. The fact that an imagined scenario in which Priya did announce a number of nested questions up front sounds ludicrous actually suggests that, insofar as you are convinced that speakers ever try to simultaneously add multiple questions, they typically do so in the "top-down" method whose use we've attributed to Priya (starting with the smallest subquestion and anticipating inferences concerning underlying super questions), rather than "bottom-up" (starting with the largest super-question and proceeding to the smallest sub-question). Making utterances with a questionably rational hope that they will achieve a simultaneous deep update is a sort of very low-stakes hail-Mary; you might as well give it a shot, since if it doesn't work you won't be any worse off than you were to begin with, and if it does you've gotten your desires across more efficiently.

1.5.2 Multiplication and HO QUDs

The above was a long digression to establish the existence of the kind of cases I think higher order QUDs will help us model. To review, in the above case, Priya isn't sure what inferences Orville might have made about her information structural desires beyond her intention to see q_1 answered. She may be unsure whether Orville has the background beliefs in light of which q_1 stands in the sub-question relationship to these further questions, or she may think that he has these beliefs but may still fail to infer her desires (after all, we fail to make inferences on the basis of information we do have all the time).

Moreover, Priya is aware that Orville will be performing a related evaluation of her information structural preferences, and, even if he is an avid inferrer of her possible deeper interests, may be unsure which of these she in fact wanted him to understand her to have on the basis of her uttering q_1. Both speakers have as their goal to represent all and only those information structures represented by the other, and so Priya will take the sort of uncertainty that Orville is bound to undergo under advisement. All of this suggests that Priya won't be in a position to know whether Orville has updated the QUD stack with $\{q_1, q_2, q_4\}$, $\{q_1, q_2\}$, $\{q_1, q_3\}$, $\{q_1, q_4\}$, or $\{q_1\}$ [18], *or* whether Orville might himself feel unable to determine what it is most likely that Priya will think he has updated with, and so similarly represent a plurality of options to himself.

The first response to this apparent plurality of information states for each discourse participant is to suggest a proliferation of different QUD stacks (of the first order alone) in Priya's context representation, and indeed I think this is on the right track[19]. I will call this move MULTIPLICATION. In the most general terms, MULTIPLICATION is the representing of a given agent's information state using more than one region of modal space subject to independent partition. But the exact procedure of MULTIPLICATION raises several questions. Below I represent a naive view of the MULTIPLICATION process. I then suggest how we might fix it.

[18] I am assuming that subquestionhood is transitive.

[19] As soon as we introduce defectiveness into the picture, individuating interlocutors' information states rather than using a shared context becomes inevitable. The discussion here is not implementation specific, but for a model that does this see the "dialogue game board" approach in Ginzburg (2012)

On this first version of the type of move I'm calling MULTIPLICATION, one first duplicates the context set so that an agent's individual information state features two copies of the context set (CS) (cf. von Fintel & Gillies (2011)). The resulting collection of two CSs we might call the second order context set. One then takes the first possible (set of) question(s) and partitions the first copy of the CS according to it, before partitioning the second copy according to the second possible (set of) question(s).

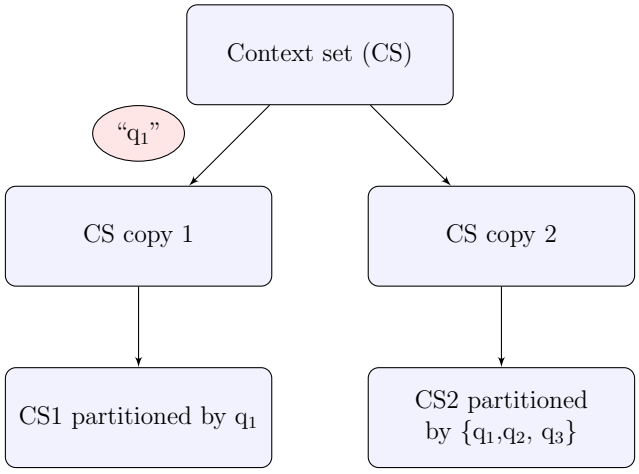


Figure 1.2: MULTIPLICATION, Version 1

But this can't quite be right. Suppose that it becomes apparent that the originally desired information structural update was $\{q_1, q_2, q_3\}$ (as in a case of UNCERTAINTY) or that one speaker addresses q_3, making it evident that they are functioning as though they take this to be the case (as in cases of NONSPECIFICITY); this should presumably trigger a contraction of CS_2, so that the CS copy partitioned by $\{q_1\}$ (alone) is excluded and only the copy

partitioned according to $\{q_1, q_2, q_3\}$ remains– so far, so good. But now recall what the context set was supposed to do: it was supposed to keep track of what, for all the conversational participants knew (or committed to acting as though they knew, or believed, or believed to be believed etc.) might be the world they were in. While we primarily attend to the way that the context set contracts in response to speaker-meant information, it can contract in response to other information as well. It is typically assumed that some worlds are excluded even at the beginning of the discourse because of the assumptions one shares with one's interlocutor. Moreover, if there's suddenly a loud bang outside the window while we have our conversation, that bit of information rules out certain states of the world and so will contract the context set as well. Here, then, is a further state of the world that we should expect the context set to contract in order to reflect: that the question under discussion in our conversation is x. According to the above version of MULTIPLICATION, one might contract CS_2 without contracting the surviving CS to reflect the fact that the it was now known (or decided) that the information structure of the conversation was as implied by the partition on that CS. So, while information structural updates impose a partition on the context set, they must also be expected to effect a propositional update on the context set. MULTIPLICATION must then actually look more like the following:

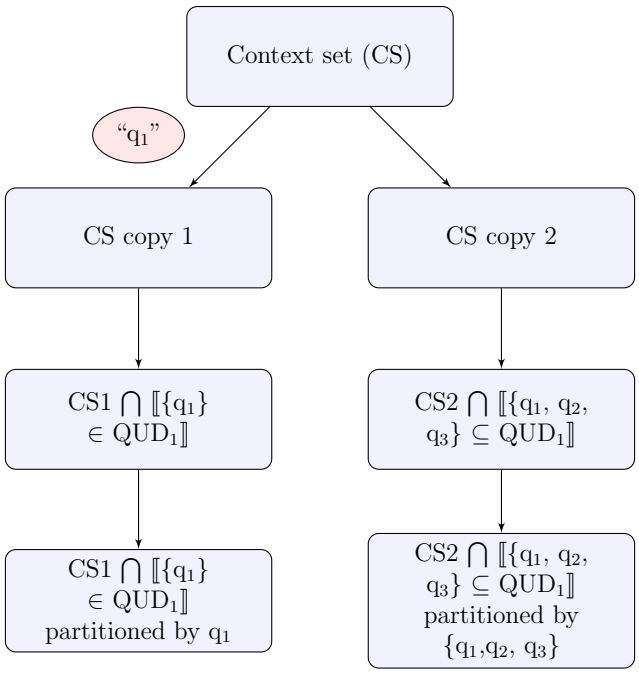


Figure 1.3: MULTIPLICATION, Version 2

The above is a basically adequate picture. I want to suggest though that we can more parsimoniously represent these sorts of updates using the alternative, independently motivated modeling apparatus developed earlier with an eye to capturing the relevance conditions of higher order utterances. When I am uncertain of the state of QUD_1, a way of representing this uncertainty is to partition the CS according to a second order question.

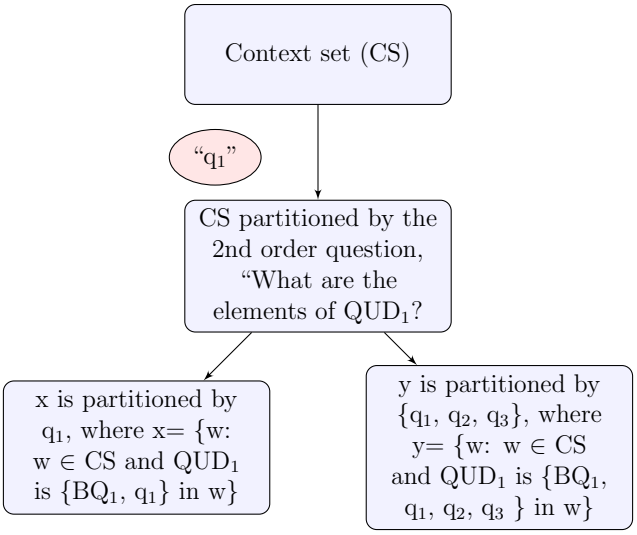


Figure 1.4: MULTIPLICATION, Version 3

This version of the move may seem to have lost the feature in virtue of which the previous two were intuitively characterize able as “MULTIPLICATION”. That is, it doesn't involve the duplication of the context set; higher order QUDs effectively take over the role of higher order CSs. It does however still establish two regions subject to independent partition, whereas there was previously only one; hence, still a sort of MULTIPLICATION. An appealing part of this picture is that the job of representing salient alternatives is managed entirely by QSet, rather than being oddly and somewhat arbitrarily split between QUD functions and various orders of context sets.

1.6 Conclusion

This essay has characterized the class of higher order discourse and drawn attention to the unique freedom that conversationalists have to make this variety of utterance. The attempt has been to put our finger on the way in which an influential implementation of the question-under-discussion approach to information structure lacked the resources to articulate relevance conditions that made the right predictions about higher order utterances, and to offer a solution. The proposed solution provides the additional benefit of helping deal with an issue around the representation of defective discourses. In the argument for this proposed solution, we considered an alternative account on which an utterance's higher order status was captured by its effect on G, the total set of interlocutor goals. I concluded that this approach was non-ideal for the purposes of characterizing higher order utterances so as to articulate relevance conditions that were appropriately permissive toward them. However, there is much to be explored concerning the other ways that a model enriched along these lines can help represent the dynamics of HOUs. For instance, although HOUs might be adequately characterizable in terms of their answering higher order questions, representing the inferential procedures associated with comprehending statements as higher order may turn out to require a special field for the representation of goals, along the lines of G. This line of inquiry also promises to show points of connection between HOUs and other linguistic phenomena, like requests, pragmatic accounts of which make use of something like a goal or preference ranking.

Chapter 2

Luck and the Value of Communication

2.1 Introduction

This paper explores the analogy between communication and knowledge. More precisely, it explores the analogy between the state characteristic of the audience in a successful communication event and that characteristic of the knower in a successful knowledge acquisition event. §2 and §3 examine analogies in the analysis of communication and knowledge. We then apply Jonathan Kvanvig's (2003) argument about the *value* of knowledge (detailed in §4) in §5's discussion of the value of communication. An interesting outcome of this process is the advancement of the thesis that communication is not a uniquely valuable type of signaling event.

2.2 The Analogy

When an instance of communication takes place, one thing that happens is that the audience understands the speaker's utterance. A very schematic and preliminary characterization of what goes on with the audience, in successfully communicative signaling events, is as follows:

A. she entertains that p

B. **i**. she entertains that p because of the signal, and

ii. it is manifest to her that the speaker intended the signal to cause her to entertain that p.

C. p is (consistent with) the content the speaker intended to get across with the signal.

C notes what is uncontroversial, which is that communication involves some kind of conformity of what the speaker entertains to what the speaker wanted to express in the first place. I have included in C a parenthetical hedge against some non-classical accounts of communication, but if this causes any consternation, we can suppose that the sort of conformity relation between the propositions considered by speaker and audience is just that of identity.

Something like B rules out cases where a speaker issues a signal intending to cause the audience to entertain that p, only for the audience to fail to notice the signal at all but coincidentally entertain p just at that moment. It also rules out cases where the signal causes the audience to entertain p without also coming to see that this was the speaker's intention. B doesn't, however, contain what would be needed in order to restrict cases of successful communication to those of successful *speaker meaning*. This is intentional. With Sperber & Wilson (1986, 2015), I think that, even in cases involving two human beings[1], our normal usage of the expression "communicates that

[1]That is, not just cases where we say things like "Those tree rings communicate that

p" encompasses cases that involve showing evidence that p, as well as those where someone meant that p, and that many cases of communication are on a continuum between meaning and showing, as in the following examples:

> When asked who is the tallest pupil in the class, the teacher points to an individual who at first sight is the tallest in the class (although some pupils might be absent) and says, "He is." She both means that the pupil she is pointing at is the tallest (since some of the evidence for the intended conclusion comes from her intentions), and displays direct evidence that he is the tallest. (Sperber & Wilson (2015), p. 124)
>
> Displaying a bandaged leg (in response to a squash invitation) (Grice (1969), p. 170)

So whereas in the subset of cases of communication that involve successful speaker meaning, the audience's entertaining the target proposition has to bear a more specific relationship to the signal and the speaker's intentions (see Grice (1957, 1969)), I think our normal evaluations about whether cases

---

the tree is x years old." The above characterization of communication clearly fails to square with these uses anyway, since the audience's and speaker's both being the kinds of entities (e.g. humans, sophisticated non-human animals) who can have attitudes toward propositions (i.e. who can intend and entertain them) is built into it. I lean toward thinking that communication can be naturalized to a greater extent than I do here; attitude-involving cases of communication are probably construable as special cases of a phenomenon aptly called communication in which at least the entity doing the communicating (the "speaker") needn't be capable of propositional attitudes. I do suspect that many of the points I make in this paper would be hard to make sense of if communication were not thought to essentially involve an *audience* capable of propositional attitudes though. In any case, the naturalization project is not one I pursue here.

constitute communication suggest that communication per se doesn't require this. Sperber and Wilson think that "in all cases of communication, wherever they fall on the meaning/showing continuum...the intended import is achieved in the same way: by making mutually manifest one's intention to make an array of propositions manifest or more manifest to the audience." (Sperber & Wilson (2015), p. 139, Sperber & Wilson (2015), p. 50-54)[2].

The above presentation of the conditions on communication is meant to suggest an analogy between the audience's half of a communicative transaction, and knowledge. This analogy is natural because communication, like knowledge, is a normative notion. It's somewhat awkward to draw analogies between knowledge and communication per se, because communication involves more than one agent, and so more than one set of attitudes which one might fix on as the analog of belief; but if we center our analysis of communication on the audience, the analogy becomes clear and fitting. After all, knowledge too had (prior to 1963) classically been taken to be correctly analyzed in terms of three jointly-sufficient conditions which bear (more-or-less) neat analogy to those listed above: JUSTIFICATION, BELIEF, and TRUTH. Above, A is analogous to BELIEF, in that it specifies the bearing of some propositional attitude to a content. C is analogous to TRUTH; it specifies a directional fit relation between the audience's entertained content and that intended by the speaker just as the TRUTH condition does between the content of the knower's belief

[2]For discussion of how this varies from the Gricean account of speaker meaning, see Sperber & Wilson (2015), especially pp. 142-147

and the world.

Lastly, B is the analog of JUSTIFICATION. This point is a little more complicated because exactly what JUSTIFICATION comes to is of course a matter of longstanding dispute in the form of the intractable internalism/ externalism debate. On a rough characterization, sure to have run rough-shod over distinctions important to some, internalism is the view that the features in virtue of which a belief is justified (or not) are mind-internal, like the evidence that one possesses. In contrast, externalism says that (at least some of) these features are (at least partially) mind external, like the use of a particular belief-forming process, such as perception or memory, which is reliably correlated with the reaching of true beliefs about the world. But, whether specified in an internalist manner or an externalist one, what JUSTIFICATION does is put a constraint on how the belief is reached[3]. Like JUSTIFICATION, B (specifically Bi) places a constraint on how the content-attitude complex specified in A (entertaining P) comes about.

It might be suggested that the audience's side of a communicative transaction isn't *like* knowledge– it just *is* an instance of knowledge. Certainly a person can come to know things in the course of a communicative transaction. It may well be not only common but necessary that they come to know what the signal was like in order to come to entertain p because of it (as specified in Bi). But this knowledge is instrumental in achieving communication, not

[3]To be clear, this is not posited as a sufficient condition on an account's being one of justification.

constitutive of it. It's also true that, under the right circumstances, the audience can, as a result of the transaction, come to know that p; but this is a highly contingent consequence of communication, not something constitutive of it. On the above characterization, the audience must also come to know that the speaker intended them to entertain that p, that the speaker issued the signal with the desire that they, the audience come to entertain p, and so on. The key observation though is that the attitude that the audience has toward p which is constitutive of their having successfully received communication of p, , is not that of knowledge. The relationship between that attitude and knowledge is an analogical one.

Above, I endorsed a characterization of communication based on Sperber and Wilson's account. Even with the caveat that this characterization is meant to capture communication in general, and not just cases of speaker meaning, I think it is evident that counterexamples to the sufficiency of these conditions arise. The next section discusses these and makes some suggestions about their significance.

2.3 Communication and Luck

Historically, much philosophical concern with communication has been a result of the fact that data about what constitutes communication seem indirectly to place desiderata on a theory of meaning. Three closely related schematic descriptions of this relationship between communication and meaning that one sees at least tacitly endorsed in the literature are as follows:

1. If A's entertaining p in response to S's utterance, u, is intuitively not an instance of **communication**, then u didn't (just) **mean** that p.
2. If A's entertaining p in response to S's utterance, u, intuitively is an instance of **communication**, then u **meant** that p.
3. If A's entertaining p in response to S's utterance, u, intuitively is an instance of **communication**, and A's grasping p' as a result of S's utterance, u, intuitively is not an instance of communication, and p' differs from p only in that p includes further information, f, (i.e. p and p' are a minimal pair) then f must be a part of the **meaning** of u.

For an example of this reasoning, consider the following case from Brian Loar:

4. Smith and Jones are unaware that the man being interviewed on television is someone they see on the train evert morning and about whom, in that latter role, they have just been talking. Smith says 'He is a stockbroker', intending to refer to the man on television; Jones takes Smith to be referring to the man on the train. (Loar (1976), p. 357)

Prior to a recap of Loar's handling of this case, let's pause to observe the pre-theoretic reasons one might provide for why this doesn't seem like a case of successful communication: Jones arrived at what Smith was getting at,

but it was pure *luck*[4]. Luckyness, it seems, undermines communication just as it does knowledge. We'll return to this theme below.

Having adduced this datum (a case paired with a negative judgment of its constituting communication), Loar then brings this data to bear on the debate between direct reference theorists and Fregeans: "Now Jones, as it happens, has correctly identified Smith's referent, since the man on television is the man on the train; but he has failed to understand Smith's utterance. It would seem that, as Frege held, some 'manner of presentation' of the referent is, even on referential uses, essential to what is being communicated" (p. 357).

This compresses a few steps of reasoning: it first applies something like 1 to infer that a proposition having the form ⟨a, the property of being a stockbroker⟩ where a is the man who is both on television and the train, is not the whole meaning of Smith's utterance, since the entertaining of that property by Jones appears insufficient to render the case one of communication. Because Loar goes on to offer a positive proposal about what Smith's utterance meant (indeed, the case was presented precisely to support this positive proposal), we infer that the following case is tacitly proffered as the second member of a minimal pair in which 4 is the first member:

5. Smith and Jones are unaware that the man being interviewed on television is someone they see on the train evert morning and about whom, in

[4]One notices that these attributions of luckyness have neither exactly the status of intuition nor of theory; I take them to be a but of proto-theory

that latter role, they have just been talking. Smith says 'He is a stockbroker', intending to refer to the man on television; Jones takes Smith to be referring to the man on the television.

Loar thinks that the difference between 4 and 5 is clearly that, in 4, the audience, Jones, thinks of the man under a different mode of presentation than does the speaker, Smith, whereas in 5 Smith and Jones think of the man of whom stockbrokerhood is being predicated under the same mode of presentation.

Buchanan (2014) charges that Loar's case is not actually evidence against direct reference views about the semantics of singular terms. In 4, per Buchanan, Smith might well have wanted Jones to arrive at the proposition ⟨a, the property of being a stockbroker⟩ (where a is the man who is both on television and the train), but he wanted him to do so on the basis of the observation that a salient man in the signalling context was the man on the television. Instead, Jones reached the same proposition but not on the basis of this feature.

Here, Buchanan is developing a point also made in Grice (1969), that an analysis of communication like the one give in the previous section, does not provide jointly sufficient conditions because it clearly fails to rule out lucky cases[5]. Speakers intend not just that their audience entertain p on the basis of

[5] "Comm.', the initial characterization of communication that Buchanan criticizes, looks slightly different than the characterization given by A-C above. One reason is superficial;

the signal and understand this to have been the intention of the speaker, but that they entertain p on the basis of *some particular features* of the signal and understand the speaker to have intended this. Using a term favored by Schiffer (see e.g. Schiffer (2017)), Buchanan refers to these features as *inference base features*, or *ib-features* (pp. 63-64).

Buchanan's point, though deployed against a particular instance of 1 and 3's application, offers a very general objection to these principles' legitimacy: communication can fail even where the audience ends up entertaining the desired proposition, so neither of these principles goes through. One way of thinking of Buchanan's point is to say that, in addition to an *object condition* on communication, which is a type of condition that puts constraints on what proposition(s) the audience must entertain, there is also a *process condition* which is a type of condition that puts constraints on how the audience arrives at the entertaining of the content specified by the object condition. Loar's mistake, on this gloss, is to have interpreted an intuitively failed case of communication as having failed due to a violation of the object condition, when really the failure was due to a violation of the process condition.

Peet (2017) points out though that Buchanan's ib-feature requirement

---

Buchanan is really giving conditions on an agent's intending to communicate, and so effectively centers the analysis of communication on the speaker, whereas A-C center on the audience. My B condition however, based on Sperber and Wilson's characterization of communication, does not map perfectly onto Comm. There is obviously a much more extensive discussion to be had about the how we should arbitrate between these views, but luckily the respects in which my baseline characterization and Buchanan's differ will not be significant for the point I ultimately want to make.

(IBFR), while working nicely to exclude some lucky cases, still admits others. He offer the following case as evidenceL

> 6. Smith and Jones are unaware that the man being interviewed on television is someone whom they see on the train every morning. Smith says 'He is a stockbroker', intending to refer to the man on television; Jones recognizes that Smith is drawing upon their common knowledge that there is a salient man on the television screen; but, seeing the similarity between the man on the television and the man whom they often see on the train, he thinks that Smith, who he assumes also recognizes the similarity, is talking about the man whom they see on the train. Now, Jones, as it happens, has correctly identified Smith's referent, since the man on television is the man on the train; but he has failed to understand Smith's utterance. (p. 381)

One might think that this doesn't show us that Buchanan is wrong to posit an IBFR on communication; this case just shows us that speakers' ib-feature intentions are normally more numerous or demanding than those we have associated with this particular case. Perhaps we just ought to have posited more ib-features in cases basically like Loar's (4-6) than we did above. As Peet sees it though, the problem with this line of thought is that "we can always construct a case in which the audience recognizes all of the speaker's intended ib-features, and follows the intended inferential path as far as it goes, but then deviates wildly in such a way that they could easily have failed to

recover the correct referent were it not for some coincidence" (Peet (2017), p. 381). The audience's routing their interpretive inferences through the elements of any posited set of ib-features, so long as that set specifies an inferential path that is less than maximally demanding–that is, so long as it stops short of specifying an exact and exhaustive set of inferential steps that the audience is to pass through–will not be enough to insulate them against communicative luck.

For the addition of IBFR to A-C to render these conditions jointly sufficient then, it would have to be the case that speakers do typically have such maximally demanding ib-feature intentions. But they don't. Here is a case adapted from one in Heck (2014) (p. 337) that Peet uses to make this point:

6. Della and Kirsi are waiting in line outside a restaurant. Out of the blue, Della says, "Man, he was funny." Kirsi knows exactly who she is referring to: a man who had been doing stand-up on a corner when they were walking over earlier.

What ib-features did Della have in mind in finding the referent for 'he'? That she, Della, was likely still to be thinking of the comedian? That the most salient funny person was the guy they had recently seen on that corner? One can imagine that Della might have constructed her utterance with one of these in mind (though it doesn't seem necessary that she had any preference between them), but even supposing she did, I at least don't have the

intuition that Kirsi's identifying the referent via the alternative route would have undermined their transaction's status as communication. Note that this doesn't amount to my being, after all, tolerant of a little luck in my cases of communication: it just doesn't strike me that Kirsi's process for arriving at the semantic value of 'he' wasn't particularly lucky. What it does mean is that a version of the IBFR which required associating maximally demanding ib-features to speakers is too strong, even as any version of an IBFR that was less than maximally demanding seems to leave room for communication-undermining luck.

To recap, 4 was a kind of Gettier case for communication; the case satisfied the conditions of our prior analysis of communication, but seemed not to be a case of communication due to some luckyness in it. This presents the project of offering a candidate fourth condition which, together with A, B, and C, will be jointly sufficient for a case's being one of communication. In pursuing this project, it is natural to ask whether the candidates that catch our eye are clear analogs of candidates that have likewise been put forward (and then cast aside) in the last half century of literature on the analysis of knowledge. Of course, simply because some candidate fourth condition on communication is analogous to a failed fourth condition on knowledge, does not entail that the former too will certainly be a failure. I think it's clear that communication and knowledge are analogous, but there are differences, and these differences may well be relevant to how well a candidate fourth condition fares. Nonetheless, I think the analogy between knowledge and communication

is strong enough that we might anticipate something like this. Peet's criticism of Buchanan's IBFR has already suggested that finding a satisfactory fourth condition may not be all that simple. But should we expect it to be as hard as solving the Gettier problem?

To play this idea out a bit, let's observe that an IBFR is closely related to the 'no-false-lemmas' (NFL) response to the Gettier problem (see e.g. Clark (1963)). NFL requires that the process one uses to form a belief passes through no falsehoods. As discussed when developing the analogy of condition C to TRUTH above, as TRUTH is to knowledge, so is the consistent-with-speaker-intentions relation to communication. Thus, it's to be expected that the communicative version of NFL would involve some sort of proscription around passing through *unintended* steps rather than false steps. And indeed, variations on an IBFR require that: the process used to reach P pass through all the elements of a certain set of steps (i.e. those specified by the speaker's intended ib-features), or that it pass through *only* elements of such a set, or through all and only these elements etc.

A failing of the NFL condition is that it doesn't exclude cases like Alvin Goldman's "fake barn country" example:

7. ...unknown to Henry, the district he has just entered is full of papier-mâché facsimiles of barns. These facsimiles look from the road exactly like barns, but are really just facades, without back walls or interiors, quite incapable of being used as barns. They are so cleverly constructed

that travelers invariably mistake them for barns. Having just entered the district, Henry has not encountered any facsimiles; the object he sees is a genuine barn. But if the object on that site were a facsimile, Henry would mistake it for a barn. (Goldman (1976), p. 773)

Goldman, and many with him, have the intuition that in this case Henry does not know that the structure in front of him is a barn, despite having a justified true belief that he moreover passed through no falsehoods in forming. So much for NFL.

Here's a stab at a communicative analog to Goldman's case, to see if it poses the same sort of problem to IBFR that 7 did to NFL:

8. The people in this town all suffer from an unusual neurological condition affecting their speech: they involuntarily utter the sentence "It's snowing!" over and over, doing so with no communicative intention whatsoever. It is not only possible, but a frequent occurrence, that visitors passing through believe residents to be trying to communicate when they are not. Janette has just stopped in town to get gas, knowing nothing about the place. A resident at the pump next to her looks around and, observing the weather, voluntarily (indeed, with communicative intentions) observes to her that"it's snowing!" Janette understands him, and correctly takes him to have intended to communicate in saying what he did.

Was this an instance of communication? My own intuitions about this are about as negative as they are for 4. This suggests that, along with the surface similarity between IBFR and NFL, they can be demonstrated to be inadequate fourth conditions on their respective target phenomena by analogous counterexamples.

A full inquiry into the analogies between the fourth conditions on knowledge and communication might look closely at whether seeming similarities persisted through the dialectic around other candidate fourth conditions, like security (Nozick (1981)) or safety (Sosa (1999)). As will be apparent, the point I want to make in §5 doesn't rely on their being no viable fourth condition on communication, but on whatever that viable condition is looking ad hoc from the point of view of the value of communication.

2.4 The Value of Knowledge

This exploration of the analogy between knowledge and communication turns now to question of their respective values. As concerns knowledge, the seemingly intractable issue of establishing a way of Gettier-proofing an analysis has important consequences for the issue of knowledge's value.

The history of epistemology has regarded knowledge as its central target of inquiry, and so at least implicitly suggested that knowledge is the most valuable type of epistemic state. But 1) why is knowledge more valuable than true belief? And 2) why is knowledge more valuable than justified true belief?

My real interest here will be in Jonathan Kvanvig's (2003) response to the second of these questions. But to understand how he responds to the second, we need to understand his answer to the first.

The first question goes back to at least Plato who, in the *Meno*, had Socrates and his interlocutor discuss why one would should prefer a guide who *knows* their way to Larissa over one who merely has true beliefs about how to get there. Where knowledge is understood, as it was by Plato, to be justified true belief, this question comes done to what value is added to a true belief by justification.

Answers must navigate an apparent dilemma. On the one hand, where the value of a belief's justification is derived from its being instrumental in reaching truth, a belief that is already true would seem to have nothing left to gain from justification (see e.g. Kvanvig (2003) pp. 44-60, Zagzebski (2004)); this gets called the "swamping problem". But, on the other hand, if the value of justification is not derived from that of truth somehow, its hard to say why the value in question is epistemic in nature.

Kvanvig, at least, thinks that certain accounts of justification's nature can provide satisfying accounts of its value and so survive this puzzle. One such account is a kind of subjective internalism, on which a belief is justified just in case it was formed according the agent's own standards for truth-conducive inferential patterns. How does this view escape the first horn of the above dilemma? Bear in mind that a belief's being subjectively justified doesn't necessarily make it more likely to be true; the agent might have

some wild personal standards for how to form true beliefs. Even for such an agent though, the epistemic value of their justification is secured because of its bearing a relationship to truth; justification still reflects the agent's "adopting intentional means to the goal of truth." Thus: "[this kind of] justification is extrinsically valuable in virtue of its relationship to the truth, though it is not instrumentally valuable on the basis of its relationship to truth." (Kvanvig (2003), p.200)

Setting aside the first problem now, the trickier problem is that of why knowledge is more valuable than justified true belief. The only difference between knowledge and true belief is that knowledge involves the satisfaction of some fourth condition to rule out Gettier cases. The exact nature of this fourth condition is, as discussed in the previous section, extremely elusive. Of the many candidate fourth conditions put forward, Kvanvig makes the following observation:

> In each case, such approaches offer something of value that might be used to explain the value of knowledge, but each such approach faces immediate difficulty concerning the nature of knowledge. Counterexamples to the initial formulation of the approach force alterations in the approach, and the alterations are guided exclusively by concern over the nature of knowledge, resulting in emendations of the original suggestion that appear entirely ad hoc from the point of view focusing on the question of the value of knowledge...When we look carefully at the variety of failed attempts to satisfy the twin desiderata concerning the nature and

> value of knowledge, we do not find signs of progress. We find, instead, a repeated pattern in which progress with respect to one desideratum is balanced by greater weakness with respect to the other. (Kvanvig (2003) pp. 138-139)

Kvanvig needn't be read as pessimistic about the possibility of eventually finding an fourth condition which, together with the three classical ones, generates all and only the right predictions about which cases count as knowledge. He just thinks that the evidence thus far suggests that this condition, whatever it is, will describe the presence of some highly gerrymandered feature that adds no clear value above that of justified true belief. So whereas he thinks that the first of the problems concerning the value of knowledge is answerable by some accounts of justification, Kvanvig thinks this second problem is unsolvable: knowledge really has no value over that of a proper subset of its components.

This conclusion leaves the throne empty, as it were. Some other phenomenon might be recognized as having supreme value among epistemic states, and the candidate Kvanvig puts forward is *understanding*. Understanding, like knowledge, is an attitude that is factive (i.e. to understand that p entails that p is true; to understand an object, o, entails having true beliefs about o). Additionally:

> ...understanding requires the grasping of explanatory and other coherence-making relations in a large and comprehensive body of

> information. One can know many unrelated pieces of information, but understanding is achieved only when informational items are pieced together by the subject in question. One might even propose a more radical thesis... Whereas knowledge can have as its object individual propositions, understanding may not. It may be that when understanding is achieved, the object of understanding is an 'informational chunk' rather than a number of single propositions. (Kvanvig (2003) p. 192)

But understanding is not just knowledge of a bunch of closely related propositions plus the knowledge that they *are* related: "although knowledge is incompatible with a certain kind of epistemic luck, understanding is not" (Kvanvig (2003), p. 199). Understanding, unlike knowledge, is invulnerable to being undermined by Gettier cases. Here is a case that Kvanvig thinks makes this idea intuitive:

> Consider, say, someone's understanding of the Comanche dominance of the souther plains of North America from the late seventeenth until the late 19th centuries. Suppose that if you asked this person any questions about this matter, she would answer correctly. Assume further that the person is answering from stored information; she is not guessing or making up answers, but is honestly averring what she confidently believes the truth to be. Such an ability is surely constitutive of understanding, and the experience of query and answer, if sustained for a long enough period of time, would generate convincing evidence that the person in

> question understood the phenomenon of Comanche dominance of the southern plains. But does she have knowledge? Ordinarily, yes; but it is not required. For, on the usual theories of knowledge, all those answers could be given from information possessed and still fail to be known to be true, because the answers might only be accidentally true. For example, most history books might have been mistaken, with only the correct ones being the sources of understanding in question, and with no basis in the subject for preferring the sources consulted over those ignored. (Kvanvig (2003), pp. 197-198)

Notwithstanding having been Gettiered in the process of acquiring all her true beliefs about Comanche history, Kvanvig thinks it's intuitive that this agent understands that history, even while lacking much knowledge about it.

Whether or not it's intuitive to use the term "understanding" to refer to this luck proof phenomenon strikes me as relatively unimportant to Kvanvig's goals. What's more important is that the phenomenon that the label "understanding" is here stuck to is one that strikes us as valuable. But nor should a conclusion of its value, or value relative to knowledge, be expected to be a deliverance solely of intuition. Knowledge, Kvanvig claimed, was only as valuable as justified true belief. Why is understanding more valuable than this? He explains that

> ...understanding is valuable because it is constituted by subjectively justified true belief across an appropriately individuated body

> of information that is systematized and organized in the process of achieving understanding, and subjectively justified true belief that is systematized in this way is valuable (p. 202)

We recall that subjective internalism was the view of justification that Kvanvig thought survived the challenge presented by the first value problem for knowledge. This sort of justification was deemed valuable because it amounted to the "adopting [of] intentional means to the goal of truth" and so bore a relationship, but a non-instrumental one, to the property of central epistemic concern, truth. The view about understanding then appears to be that it involves a layer of "intentional means to the goal of truth" on top of that involved in the justification of individual beliefs. Systematizing and organizing one's true beliefs is not necessarily instrumental in reaching the truth, but it is a way that one intentionally pursues the truth. Understanding a body of information thus involves an extra connection, or subjective commitment, to getting the truth, beyond that involved in the justification of all the true beliefs that constitute the nodes in that body of information.

2.5 The Value of Communication

As with knowledge, we can distinguish two different value problems for communication. The first asks what value it has for an audience to entertain the speaker's intended content, p, on the basis of the signal and with its being manifest to her that the speaker intended the signal to cause her to entertain that p (or however one wants to spell out the B condition), over and above the

value of their entertaining the speaker's intended content. That an audience has come to entertain what they have on the basis of the signal might be thought to be conducive to their entertaining the content the speaker had in mind, but where they were already doing this, the satisfaction of the B condition seems to add nothing; this is the swamping problem, re-emerging with respect to communication.

One might think though that the other horn of the dilemma we saw in the above presentation of the first value problem for knowledge doesn't have an analog here; whereas we were concerned that some attributions of value to justification wouldn't cast it as valuable in any distinctively *epistemic* way, what is the proprietary domain of value for communication and any one of its components whose value we might assess? I say that this is the domain of signaling. This domain, like the epistemic domain, is normative in that it is goal-oriented: in the epistemic domain the fundamental goal is truth, and in the signaling domain the fundamental goal is (something like) information transfer. If this is so, then we can articulate the second horn as follows: if the value of B isn't related to the transfer of information, it's not clear how its value is distinctively signaling-oriented.

As the goal of this paper is to sort of stick with an analogy and see how far it gets us, we now construct a response to this apparent dilemma that mirrors Kvanvig's concerning subjective internalist justification. Whatever the precisely correct specification of the B condition is, it will be something that isn't necessarily instrumental in information transfer, but which represents

an intentional commitment on the audience's part to facilitating information transfer. B thus adds a distinctively signal-oriented value to A and C.

On to the second problem of value for communication. If problems with luckyness are, as we suggested above, likely to be as persistent for communication as they are for knowledge, and if one finds Kvanvig's above response to the second value problem for knowledge compelling, it is natural to wonder how a similar response might play out with respect to communication. That is, perhaps we think that, the closer we get to a fourth condition on communication that genuinely excludes lucky cases, the more ad hoc this condition will come to seem from the point of view of communication's value. So we should take it that all the value in communication resided in the proper subset of its components enumerated in A-C above; just as Kvanvig concluded that knowledge is not a uniquely valuable epistemic state, we might conclude that communication is not a uniquely valuable signaling achievement.

Well, goes one line of criticism, whoever said it was? Knowledge has been a central target of inquiry in philosophy for the discipline's entire history, hence the case for there existing a presumption of its importance that it might be interesting to reject. Why think there's any such general presumption with respect to communication, a phenomenon whose philosophical consideration has so often been secondary to an interest in more central phenomena like meaning? Here it's worth noting that the first value problem for knowledge gets off the ground in the Meno without appealing to a philosophical tradition of revering knowledge; the folk seem to care a lot about knowledge, and it's

from their concern that the presumption of knowledge's unique value derives its force. I think it's likewise clear that the folk care a lot about communication, and so I think it's counterintuitive enough to be worth noting if communication turns out not to be uniquely valuable.

Let's now consider what the analog of understanding might be with respect to signaling[6]; understanding was true justified belief about a bunch of related pieces of information, plus some grasping of the connections between these, but without an anti-luck criterion. Let's stipulatively label our targeted analog *comprehension*. Comprehension involves in fact satisfying A-C with respect to a bunch of closely related signaling events– a "chunk" of signaling events, to recall Kvanvig's language– and grasping the connections between these.

Now, the kind of entity which it is natural to identify as such a chunk is a *discourse*. And one natural suggestion about what it is to grasp the connections between the moves in a discourse is that this amounts to tracking the information structure of that discourse. But just as there were no constraints put on the kind of connections between individual facts the appreciation of which could constitute understanding, we needn't confine comprehension to

[6]An unfortunate confusion is liable to arise because 'understanding' is a natural term to use for what it is an audience does i successful cases of communication, whereas here I want to talk about understanding qua putatively privileged epistemic phenomenon. Now what the audience does in cases of successful communication may incidentally be an instance of understanding in Kvanvig's sense (indeed, it seems he doesn't think his sense is a technical one), but it's important here to recall that we're examining an analogical relationship between what communicatively successful audiences have and knowledge, not an identity relationship between what these audiences do and understanding

an appreciation of the kind of relations between discourse moves that normally make their way into an account of information structure. A recognition of deep motivational connections between moves in a discourse too might contribute to comprehension.

Comprehension will subvert communication as the most valuable state in the domain of signaling, because it involves a greater intentional commitment to information transfer. This follows from the fact that drawing information-structural and deep-motivational connections between moves in a discourse is a way of committing oneself to information transfer.

2.6 Conclusion

In much of this paper, I have plowed ahead with the analogy between knowledge and communication while doing relatively little to point out any disanalogies. I think this has been a legitimate enterprise; assigning oneself a position (that these two things are next to perfectly isomorphic) and observing what kind of moves one would want to make in order to press that claim is one way of getting an initial sense of the dialectical possibilities in a region. Further thinking in this area will naturally involve probing this analogy more critically.

Chapter 3

This Machine Kills Fascists: Detecting Propaganda with Formal Models of Mass Discourse Structure

3.1 Introduction

Democracies' primary defense against existentially threatening propaganda is inoculation of their constituents against propagandistic manipulation. But in any attempt to perform this inoculation, a familiar problem emerges: those with differing ideologies will disagree about which messaging is threatening in this way. If, as many suggest, propaganda is essentially that which undermines a polity's ability to rationally deliberate on questions of public importance, this disagreement is about what rational deliberation looks like. Those on the political right may argue that the left's rationale for a minimum wage is so faulty as to count as undermining rational deliberation; those on the left may suspect the same of the right's rationales for mandatory minimum sentencing for non-violent drug crimes. It would be useless to suggest that any faction be privileged in their ability to dub others propagandistic merely for promoting kinds of inference which they find unsound; such a designation would influence the public only along ideological lines and so systematically fail to reach those actually in thrall to propaganda– whether this be a subset

of the public or virtually all of it in some way or another. So much, one might then think, for the prospects of a non-partisan effort to identify propaganda.

I propose a way of identifying propaganda that is less vulnerable to this problem. This method argues that a certain kind of mass discourse polarization is functionally propagandistic. Drawing this connection between polarization, a structural feature of mass discourse, and propaganda allows us to abstract away from the particular content of political messaging in our designations of propaganda. It is therefore a way of identifying propaganda that is, at least in theory, immune to the siren's song of partisan condemnation.

In what follows, I first develop this conceptual link between propaganda and polarization (§2). I then elucidate the notions of mass discourse and of discourse structure alluded to above (§3). I connect these two notions with a novel way of representing and analyzing the structure of mass discourses. From here I lay out my proposed formalism for modeling mass discourse structure and introduce two indices which represent the levels of propagandicity, of the kind I characterize in §2, in a modeled mass discourse (§4). I then connect the operationalization of propaganda at work in my model to the norm of public justification and to propaganda's non-epistemic dimensions (§5). I close with a prospectus on the next steps of the project.

3.2 Propaganda and Polarization

Many are likely to feel about propaganda as Justice Stewart did about pornography: we don't have a definition ready at hand but we "know it when

we see it." Those cases of propaganda that we tend to take as paradigmatic work via conspicuous mechanisms: explicit assertion; exaggeration of negative or positive traits in the subject of visual depiction; emotionally evocative slogans. Nazi posters caricature Jews and proclaim that, *Der ist Schuld am Kriege!* (The war is his fault). Contemporary pro-natalist campaigns in Italy depict a white, plausibly ethnically Italian woman, one hand on abdomen, one holding an hourglass, and declare that, *La bellezza non ha età, la fertilità sì* (Beauty has no age, fertility does). Reconstruction-era ads from the former confederate states of the US depict black freedmen indolently consuming the fruits of white laborers' toil; as sexually aggressing against white women; or else as succumbing to lynching at the hands of a klansman in vaguely Greco-Roman dress.

What theorists have tended to identify as distinctive in both subtle and flat-footed forms of propaganda is its effect on both the information and the methods that polity-members use in their process of political belief formation. On what Jason Stanley calls the "classical sense," propaganda is "*manipulation of the rational will to close off debate*" (Stanley (2015), p. 48, italics Stanley's own). Jacques Ellul characterizes propaganda as "A way to obtain power thanks to the support of psychologically manipulated groups or masses, or to use this power in putting pressure on the masses" (Ellul (1973), translation mine). And Randal Marlin defines it as "the organized attempt through communication to affect belief or action or inculcate attitudes in a large audience in ways that circumvent or suppress an individual's adequately formed,

rational, reflective judgment" (Marlin (2002), p. 22).

That there is some kind of connection between polarization and propaganda is suggested by the fact that polarization too is often distinguished by its characteristic negative epistemic effects. To make the case for their connection really compelling, we'll drill down a bit on the nature of political polarization.

3.2.1 Varieties of Political Polarization

Political polarization is very generally thought of as political division or disagreement that is *deep*, *intractable*, and *systemic*. Exact operationalizations of this notion vary concerning the traits with respect to which agents of a polarized society are clustered into their estranged camps[1]. A few sample operationalizations, falling into three broad categories, are as follows:

1. Network Polarization

    (a) Two or more distinct clusters of polity-members exist with respect to political information sourcing (e.g. Davis & Dunaway (2016)).

    (b) Two or more distinct clusters of polity-members exist with respect to inter-agent communication (e.g. Pattie & Johnston (2016)).

2. Doxastic Polarization

[1]A cluster is a concentration of agents associated with the same value vis-à-vis a categorical variable, or associated with relatively similar values vis-à-vis a continuous variable

(a) The existence of a two-peaked (bimodal) rather than single-peaked (unimodal) distributions of policy views across a political spectrum (e.g. Fiorina et al. (2005), Ansolabehere et al. (2006), Glaeser & Ward (2006), Fiorina & Abrams (2008))

(b) Individual political agents' ideological commitments are (on average) extreme relative to an appropriate reference point (e.g. perhaps Abramowitz & Saunders (2008)).

3. Affective Polarization

(a) Individual political agents show, on average, high levels of antipathy toward opposing ideologies and/ or those who hold them relative to an appropriate reference point (e.g. Iyengar & Westwood (2015)).

Moreover, each of these operationalizations can be applied to the members of a polity very generally, or to particular groups, such as political elites, media outlets, or the political laity. Media polarization is particularly interesting, because where two or more distinct clusters of media outlets/ major information venues exist with respect to ideological skew, this looks like doxastic polarization with respect to members of the media themselves. But the media's doxastic polarization is clearly closely related to a form of network polarization for members of the polity more generally who may access media outlets selectively and who rely upon them for their political information.

Many of these kinds of polarization will be partially co-constitutive, correlated or causally intertwined, and whether and how high levels of polarization within a polity are concerning, or even interesting, will depend on which operationalization(s) we work with.

Beyond the varieties of polarization described above, I suspect there is another variety of polarization worth attending to in what I will be calling *deliberative polarization*. On my characterization, deliberative polarization is a state of affairs in which clusters of agents within a population (say that of an entire polity) think though some target issue or range of issues in highly divergent ways. Put slightly differently, deliberative polarization is a particular pattern of *deliberative strategy variation* (DSV) across a population.

To be clear, I am staking no claim to deliberative polarization's being the right analysis of polarization *simpliciter*, but to its being a kind of polarization whose study is motivated by the reasons typically mustered to motivate the study of polarization more generally. I take it that there are two broad classes of reasons to be interested in political polarization. First, there is the worry that political polarization has negative *effects* concerning the efficiency and responsiveness of a polity's governing structures: legislative gridlock, an erosion of the separation of powers, and, together with the fact that many American judicial positions are filled by election, an undermining of the judiciary's counter-majoritarian character (Abramowitz & Saunders (2008); Fiorina et al. (2005); Davis & Dunaway (2016); Pattie & Johnston (2016)). Second, there is the concern that polarization just *constitutes* a breach in some

vital feature of a polity.

Something like this latter concern is one I take to follow from certain versions of deliberativism in democratic theory. The basic notions of deliberative democratic theory are familiar ones. One fairly typical expression of the view is offered by Jürgen Habermas as follows:

> ...the attempt to arrange a society democratically...is a question of finding arrangements which can ground the presumption that the basic institutions of society and the basic political decisions would meet with unforced agreement of all those involved, if they could participate, as free and equal, in discursive will-formation (Habermas (1979), p. 186).

Deliberative democracy, as a normative theory, says that policy decisions being reached (or reachable) via deliberation among a relevant set of political actors is either conducive to, or a necessary condition on, the legitimacy of that policy. That this characterization is vague in some spots and disjunctive in others is an artifact of the attempt to be maximally ecumenical. Most theories of deliberative democracy don't posit the actual formation of a consensus as a necessary condition on legitimacy; rather, such a consensus is a sort of regulatory ideal, or else it is the hypothetical capacity to provide (theoretically) acceptable reasons to the relevant agents which is prized. The allusion to "the relevant set of political actors" reflects the fact that deliberative ideals don't settle the question of whether polities should organize

themselves into representative or direct democracies, and so the matter of exactly *whose* deliberation is required is left up for grabs. And while some will regard the satisfaction of the deliberation condition (however these antecedent choice points are resolved) as a hard condition on the legitimacy of a polity, others, sometimes called "instrumentalists," will argue that satisfaction of this condition is merely conducive to the achievement of political legitimacy.

I don't here offer anything like an arbitration among the versions of normative deliberative theory offered historically in Rousseau, Kant and Mill, or more recently by Rawls, Habermas, and many others. My project is the very modest one of suggesting that any level of commitment to a version of deliberativism motivates an interest in deliberative polarization. And some level of commitment is seemingly widespread. It is often claimed that, as a matter of *descriptive* fact, a hallmark of modernity is that the capacity to give justificatory reasons for policy decisions is what individuals regard as necessary for the legitimacy of these decisions (see e.g. Habermas (1979)). This amounts to some aspects of *normative* deliberative democratic theory being widely endorsed.

When considering why deliberativism motivates an interest in deliberative polarization, it will be helpful to have a concrete example of deliberative strategy variation in mind. Consider the following sets of rival arguments:

1. (a) A fetus may have a right to life, but a right to life does not mean a right to use someone else's body without

their consent. So a mother is not obliged to carry a fetus to term. So abortion is permissible.

(b) A fetus is not a person, and only persons are eligible for ethical consideration. A fetus, in itself, is thus not eligible for ethical consideration. So abortion is permissible.

2. (a) Possessing a full complement of human genetic material is all that's required to have a right to life. A fetus is fully, genetically human. So abortion is impermissible.

(b) What makes killing wrong, when it is wrong, is that one is depriving a creature of their future. A fetus has a future. So abortion is impermissible.

The first pair of arguments arrive at the same conclusion, as do those of the second pair. But within each pair, it's evident that that conclusion is reached in different ways; here we see deliberative strategy variation with ultimate doxastic convergence. It's moreover evident that any discourse between the proponent of a pair-1 argument and the proponent of pair-2 argument would need to engage in some reconstruction of their respective background dialectics in order to figure out where they fundamentally diverged; the bits of argumentation given here are clearly a ways downstream from such a dialectical divergence.

It's pertinent here to observe that a common feature of the most intractable-seeming debates on issues of public importance, is what Alasdair MacIntyre

went so far as to call "the conceptual incommensurability of the rival arguments" (1981, p. 8). Macintyre observes that, among arguments like those enumerated above,

> Every one of the arguments is logically valid or can be easily expanded to be made so; the conclusions do indeed follow from the premises. But the rival premises are such that we possess no rational way of weighing the claims of one as against another (1981, p. 8)

I tend to think that saying we have *no* way to engage in this weighing goes a bit far; it's quite true, however, that doing so requires referring the debate back to the abstractions that undergird it, and that doing so often poses practical problems. In any case, MacIntyre raises the worry that deliberators who are attending to the contours of such rival arguments, and who fail to trace the dialectic back, will be left uncertain about how to bring the reasons mustered by each into contact with one another. One might find oneself compelled by arguments from each of the above pairs, but, lacking any principled procedure for deciding which is the more compelling, forgo the formation of an opinion altogether or, as one suspects is more likely, break the stalemate through the use of some *unprincipled* method. For such deliberators, it's then not clear how the process of weighing reasons against one another can, by itself, conclude in anything like the "discursive will-formation" envisioned by Habermas and other deliberativists.

I think that certain patterns of deliberative strategy variation endanger deliberativist norms for reasons beyond those raised in the above passage from MacIntyre. Whereas he was concerned about individuals who were more or less receptive to all of the above arguments, as far as they went, I am worried about those who are so unexposed to other deliberative strategies that, when confronted with them, they can not even see the pertinence of the mustered reasons to the issue at hand; while MacIntyre's agent was befuddled by her very neutrality, mine is limited by her conditioned partiality.

I'll make this suggestion more vivid with a concrete example. Consider debate around health care policy in the contemporary United States. Simplifying greatly, rhetoric on the left suggests that the decisive issue in this policy domain is what level of health care people have a right to; from there, remaining questions concern method of delivery and political expediency of implementation strategies. On the right, the decisive question remains which responsibilities the government has to arrange for citizens' health care and, in turn, which coercive powers to do this arranging should be ceded to it. While the right and the left might be distinguished by the conclusions they characteristically reach about health policy questions (i.e. regimes of extensive government regulation up to and including a single-payer system, as opposed to "market-based" or entirely free-market responses), this attention to their respective deliberative starting places sheds light on why these debates are seemingly intractable.

Geographical and media polarization increase the chances that an agent

may be regularly exposed to only one of these deliberative strategies concerning health care. Such an agent, confronted by someone propounding, or tacitly relying upon, an alternative strategy, may just fail to see the reasons proffered in this alternative strategy as relevant at all to the policy question at hand.

A liberal American, after tuning into Fox News on a whim, or hazarding a chat with her Republican coworker, might think "My god, why are we talking about the dangers of creeping authoritarianism in a debate about health care? Our current health care system is killing people, so it's obvious that we need Medicare for all. The rest is an idiotic diversion." While this is a caricature, it includes two features which I suspect are characteristic of real such encounters: 1) the agent regards the opposite side's putative justifications as almost non-sequiturs, and 2) she begins to suspect that only stupidity ("idiotic") or bad faith (a "diversion" is being executed) can explain what's going on. The latter feature is not essential to my argument here, but is suggestive of a way that deliberative and affective polarization are connected. The first though constitutes the sort of threat to deliberative democratic ideals that deliberative polarization really poses: it renders large swaths of constituents insensible to the dialectics that animate each other's policy preferences. Crucially, deliberative polarization consists not just in constituents disagreeing (doxastic polarization) or disliking each other (affective polarization), or even in their simply not speaking to each other (some variety of network polarization) but in their being ill-equipped to engage deliberatively even should they end up in a conversation with one another.

3.2.2 Deliberative Polarization as Propaganda

My claim will be this this: a deliberatively polarized mass discourse just *is* an instance of propaganda. This will strike some as sounding strange, since we normally take it that that the sort of things that can count as propaganda are individual expressive acts (e.g. the making of an utterance, the posting of a poster) not whole discourses. However, it will be distinctive of the type of the type of propaganda I introduce here, *Deliberative Polarization Propaganda (DPP)*, that its instances are structured bodies of expressive acts (i.e. discourses), rather than individual such acts. It will nonetheless remain natural to think of individual expressive acts and unstructured bodies of these acts as instances of DPP, though their status as such will be derivative.

Distinguishing DPP as a type of propaganda is not intended as a stipulative extension of the meaning of the term "propaganda", but as a substantive proposal about which things in the world count as propaganda if the characterizations of propaganda that are already commonplace, such as those by Stanley, Ellul and Marlin considered above, are correct.

Deliberative polarization per se doesn't limit the strategies one uses in deliberation about a key policy question; we can easily imagine a non-polarized public where individuals' deliberative horizons excluded valuable ways of thinking through a problem. Nor does deliberative polarization mean that more agents will deliberate in less rational ways. But deliberative polarization is a state of affairs responsible for closing off debate by providing conditions amenable to the development of incongruous deliberative strategies

between different sections of the population. It threatens epistemic integrity of the group, rather than that of the individual.

What I've tried to do in this section is render explicit the notion of polarization that I say is functionally propagandistic. In §5 I will provide further arguments in favor of this approach.

Next though, I develop a formally precise ways of characterizing deliberative polarization. Building on existing models of information structure in non-mass discourses, I develop a method of representing the ways in which agents in a polity may deliberate about a particular policy issue. I then offer a method of mass discourse analysis that effectively measures the degree to which agents' "deliberative scripts" on an issue converge or diverge.

3.3 Discourses, Mass and Count

The phenomenon I will be calling mass discourse is a collection of expressive acts (e.g. utterances). But which ones? Some things that one might take as the target of an inquiry into mass discourse are as follows:

1. Everything said within certain geographical or temporal boundaries
2. Everything said by the mass *media* and/or elites
3. Everything said *publicly*, by anyone

It is this third option that I have in my cross-hairs. To be clear though, it's not that I think there exists one single mass discourse either; "mass dis-

course," as I'll use the phrase, is (not unintuitively) a mass term, not a count term. We can talk about a given quantity of mass discourse, drawing the boundaries (e.g. geographical, topical and temporal) however we like, just as we can talk about a particular quantity of water in a glass. And as with "water," we can if we like use "mass discourse" as a count term when there are some salient partitioned portions of it to refer to (e.g. to a courier bearing bottles of Aquafina, "Bring the waters over here;" by a scholar of comparative politics: "The mass discourses of India and China are importantly different."). But where no salient portion is specified, the meaning of "public discourse" defaults to referring to the undistinguished mass rather than a portion of it.

What it is for something to be said "publicly" is of course a remaining question. Rather than a nuanced elucidation of our, arguably historically specific, notion of publicity (see e.g. Habermas (2015)), my characterization of publicity here is operationalized for practical use by the social scientist engaged in mass discourse analysis who just wamts to know when to include a case in their body of data[2]. Here then, what is public is what is easily accessible (perhaps also actually widely accessed); it includes media messaging; policy documents; statements disseminated by political figures, corporations or not-for profits; blogs, television, Twitter, Facebook; and so on.

I will be offering a model of mass discourse that builds on models of non-mass discourse. Non-mass discourse, or what I will (nodding once again

[2]Notably, I think a more critical characterization of publicity is necessary when giving an account of *public opinion*– it's just that I don't touch that phenomenon here.

to the grammatical properties of the expression) call "count discourses" or "c-discourses," is the variety of discourse about which most linguistic theorizing has been done. "C-discourse" is really a name for the kinds of conversations that we have on a day-to-day basis; they typically, though not necessarily, involve a relatively small number of people and typically, though not necessarily, involve contributions from multiple speakers. The category of c-discourse is not sharply delineable, but nor is providing such a delineation necessary for this project. All that's important is that the phenomenon be grasped so that it can be contrasted with that of *mass* discourse.

Another sort of discourse worth mentioning in order to distinguish it from my target variety of mass discourse is that associated with the post-structuralist tradition in general, and with Michel Foucault in particular. This is all the more necessary since this type is typically referred to simply as *discourse*, without any qualifying adjectives. In the Foucauldian sense of the term, a discourse is a set of those key "statements" (another common term given a technical meaning) which act to delimit the very conception of a practice or subject matter. Although Foucault describes the entities he calls statements as "events," they are not token utterances or written statements. On his technical sense, statements are bits of normative content, and so are a *type* of message rather than a token instance of messaging. They are events only in that they are specific to a time period. The token utterances or amalgams thereof we find in c-discourses (as characterized above) can be said to participate in or partake of the statements that constitute a discourse, and they

can even amount to what Foucault calls "discourse formation" by propagating new such statements (Foucault (2002)). The point to be clear on, however, is that what I'll call "mass discourse" is not the same thing as what those in the Foucauldian tradition are getting at.

3.3.1 C-Discourse Structure

Conversations are more often than not arranged around topics. Having a sense of what the present topic is helps interlocutors know what it would be relevant to add, and also helps them derive implicatures and resolve ambiguities in others' utterances. In formal pragmatics, the relation between and among a discourse's constitutive topics and utterance, is referred to as the discourse's *structure*. To date, discourse structure has been a phenomenon of interest primarily to linguists and philosophers of language (e.g. **?**). Part of the upshot of this paper is that it must come to be of significant interest to social epistemologists and political philosophers, as well as those in the social sciences who study media effects on public opinion, and political behavior more broadly.

One modeling apparatus used for this task is one that organizes discourse into the question that is under discussion at a given time, and which registers the relationships between successive and contemporaneous questions. A few key notions of this model are as follows:

- *Questions:* Thought of here as *sets* of the propositions that would count as their answers (Groenendijk & Stokhof (1984b), **?**, Hamblin (1973b)).

- *Full and partial answers:* A partial answer entails the truth or falsity of at least one element of the set of alternative propositions denoted by the question to which it is an answer. A full answer entails the truth or falsity of every member of that set.
- *Sub- and Super-questions:* One question, q1, is a subquestion of another, q2, iff any complete answer to q2 entails a complete answer to q1. Subquestions are sometimes characterized as *strategies* for answering their superquestion.

I here provide a full characterization of the model I will be using for c-discourse[3], though for most intents and purposes, the above bullet points are sufficient. For a given c-discourse, D:

$\text{InfoStr}_D = \langle M, Q, A, C, Acc, CG, QUD \rangle$.

1. **M** is the *set* of moves in D
2. **Q** is the *set* of questions in D (Q $\subseteq$ M)
3. **A** is the *set* of answers in D (A $\subseteq$ M)
4. **C** is a *function* from M to date and time stamps, yielding for each m $\in$ M the date and time at which it was made
5. **Acc** is the *set* of accepted moves in the discourse (Acc $\subseteq$ M)

[3]This model is closely based on that in Roberts 2012. In other work I motivate and introduce some extensive alterations, but as these aren't relevant here I stick with the more basic model.

6. **CG** is a *function* from M to sets of propositions, yielding for each m $\in$ M the common ground of D just prior to the making of m.

7. **QUD** is a function from M to ordered subsets of Q $\cap$ Acc such that, for all m $\in$ M:

   (a) For all q $\in$ Q $\cap$ Acc, q $\in$ QUD_n(m) iff:

      i. C(q) $<$ C(m)

      ii. CG(m) doesn't entail an answer to q and q has not been determined to be unanswerable.

I want briefly to mention an alternative way that one might model discourse structure and explain why this is not the method I employ in this project. Quantitative methods in sociolinguistics and in the kind of discourse analysis that takes place across disciplines as diverse as political science, cultural studies, sociology, economics and literary studies, have developed various ways of automating the separation of discourse into parts with different topics, and of analyzing features like the changes in, and co-occurence between, different topics (see e.g. Blei (2012)). The notion of topichood exploited by those models based on Latent Dirichlet Allocation (LDA) is quite different from the notion of topichood that arises from the literature using QUDs to structure discourse. For those using LDA, a topic is a distribution across some vocabulary, and sections of text are probabilistically assigned topics based on how the term frequencies within them match the distributions associated with

each topic. Using QUD models, the topic of conversation is not characterized by term frequency or distributions of term frequencies across the text, but by the notion of a question currently under discussion. No doubt because of the difficulty involved in the automation of this method for coding corpora, the notion of topichood and structure developed in the above literature is not yet used in any extensive corpus analysis. Nonetheless, I employ it in the current project because, unlike LDA-based models, it captures relations between topics beyond that of, say, proximity in a text. Models which construe information structure in terms of sub- and super-questions nicely capture the contours of deliberative strategies and so render tractable the project of high-resolution comparison between rival such strategies.

3.3.2 Mass Discourse Structure

On my conception, a mass discourse is partially constituted by a collection or sample of c-discourses. A key insight is that, for our model to draw to the foreground the sort of structure that we're interested in, we must think of a mass discourse's constituent c-discourses as retaining their internal structure even as they are aggregated into a mass. This point is more vivid if we consider the alternative. We might have considered mass discourse as sort of a large vessel into which a variety of smaller vessels, c-discourses, emptied their contents. If what we wanted to know was, for instance, how frequently an expression was used in a certain population and across a certain time interval, this approach would be perfectly fine; total expression frequency in a

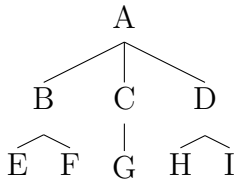
mass discourse is a property left intact even if the internal structure of its c-discourses is not. But what we in fact want to know is how the deliberative strategy enacted in individual c-discourses, a property of these c-discourses discernible at the level of their information structure, varies across a sample. For this reason, the model of mass discourse that I'll be offering is parasitic on the above model of c-discourses.

With this model of c-discourse structure in hand, we can posit a way to compare the structures of individual c-discourses–the task for which our mass discourse model is being developed. What we'll be particularly interested in doing is comparing c-discourses that both have a certain question on their QUD stack at some point. Specifically, we'll be interested in comparing the sets of sub-questions that are laid on top of this shared question, as this is what reflects the differences and similarities in the way that the shared superquestion is deliberated about in each c-discourse.

Notice that, at the point of any move in a discourse, m, $QUD(m)$ delivers the set of questions on the QUD stack at the moment immediately antecedent to m, and the contents of this set are ordered through application of the C function, which associates each element with its time stamp, among which the natural (i.e. chronological) order obtains. Although this chronological ordering doesn't rule out any two questions on the QUD stack having the *same* time stamp, the idealizations associated with at least Roberts' approach to the QUD model do rule this out. As a result, for any m, there is a total order on $QUD(m)$. Again because of the assumptions of the Roberts model,

this order reflects the sub-question/ super-question relationship in that each question is a subquestion of the question of the one preceding it (should there be any such) and a super-question of the one following it (should there be any such). All of this means that the state of a QUD stack (i.e. the totally ordered set that is the output of the QUD function) can always be represented as a tree structure in which the daughter relationship corresponds to the sub-question relationship, and that such a structure will only ever feature a unary branching structure, since binary or greater branching would suggest a merely partial order on the QUD stack, which has been ruled out:

Trees representing *synchronic* discourse structure will all feature such unary branching. But notice that when we engage in conversation about some vexing question, q, we may pursue multiple subquestions to q consecutively, and that each of these first-generation subquestions may in turn have multiple consecutive subquestions of their own, and so on. A tree representation of our *diachronic* deliberation about q will involve greater than unary branching structure, in which some questions bear sister, rather than mother or daughter, relations to one another:



A full deliberative strategy concerning a question, α, is going to be most apparent by looking at the α-region of such a diachronic tree representation. A comparison between diachronic discourse structures is then what we're interested in here. As the first pieces of a formal characterization of how to do this, suppose that we have a sample of c-discourses, D. I introduce the following function, *Strat*, which, for any discourse, d, which is within our sample, D, and any question, α, yields the ordered set of questions which occur at some level of sub-question in the diachronic q-deliberation which takes place in d[4]:

Strat is a *function*

from $\{\langle \text{d}, \alpha\rangle | d \in D$ and $\exists x(\alpha \in QUD_d(x)$ and $x \in M_d)\}$

to $\{\text{q: q}\in Q_d \cap Acc_d$ and $\exists x(\alpha \in QUD_d(x)$ and $x \in M_d)$ and $C(q) > C(\alpha)\}$

Note that $Strat(d, \alpha)$ is undefined if d at no point addresses the question α. We then have a convenient way of formally specifying those discourses in our sample which address a particular question using the *Pert* (as in pertinent) function as follows:

[4]I haven't included here a formal specification of the way that the nested ordering on generations of sub-questions is maintained. The requisite formalism is relatively straightforward but not terribly interesting. It has been omitted just for the purposes of streamlining the model's elaboration.

Pert is a *function*

from $\{\langle D, \alpha\rangle | D$ is the set of c-discourses in a mass discourse sample$\}$

to $\{x | x \in D$ and $Strat(x, \alpha)$ is defined$\}$

To qualitatively compare the deliberative strategies regarding some question, α, in two c-discourses, d_1 and d_2, one then compares the output of the *Strat* function for each. Of primary concern here though is the ability to quantitatively determine their similarity. I introduce the *CStrat* function to do this, which is definable as follows:

CStrat is a *function*

from $\{\langle d_1, d_2, \alpha\rangle | d_1, d_2 \in D$ and $\exists x \exists y(\alpha \in QUD_{d1}(x)$ and $x \in M_{d1}$ and $\alpha \in QUD_{d2}(y)$ and $x \in M_{d2})\}$

to the value given by

$$\frac{|Strat(d_1, \alpha) \cap Strat(d_2, \alpha)|}{0.5(|Strat(d_1, \alpha)| + |Strat(d_2, \alpha)|)}$$

CStrat assigns to any pair of discourses a similarity score vis-à-vis their deliberation about a particular question. With this similarity metric in hand, we are then in principle positioned to cluster discourses in a sample on the basis of their similarity scores. Very generally, a cluster is a concentration of points associated with the same value vis-à-vis a categorical variable, or associated with relatively similar values vis-à-vis a continuous variable. Clustering is a routine statistical task, though one which can be executed in a number of ways.

The exact clustering algorithm right for the task is not an issue I pursue here. What is important for our purposes is that we've set up the necessary input for the clustering task (i.e. developed a similarity metric) and that we go on, as I do in the next section, to say what interesting conclusions we can reach on the basis of this task's output.

A last component of the mass discourse model I'm developing here is the semi-technical notion of an agent *accessing* a discourse. To access a discourse is to witness it in some capacity, whether as a participant or a passive onlooker; I access the discourse I engage in with my friend over the lunch table, and I also access the discourse between Presidential candidates when I stream a televised debate.

That last piece squared away, the full model of mass discourse structure which our above work has helped us develop is as follows. Let some particular sample of mass discourse be called B. MassInf$_B$= $\{A, D, R, Strat, CStrat, Pert\}$ such that:

1. **A** is the *set* of agents present in B.
2. **D** is the *set* of c-discourses present in B.
3. **R** is a *relation* between A and D, such that $\mathrm{a}_n\mathrm{Rd}_n$ just in case a_n is witness to d_n.
4. **Strat** is a *function*

   from $\{\langle \mathrm{d}, \alpha\rangle | d \in \mathrm{D}$ and $\exists x(\alpha \in QUD_d(x)$ and $x \in M_d)\}$

to {q: q$\in Q_d \cap Acc_d$ and $\exists x(\alpha \in QUD_d(x)$ and $x \in M_d)$ and $C(q) > C(\alpha)$}

5. **CStrat** is a *function*

   from $\{\langle d_1, d_2, \alpha\rangle | d_1, d_2 \in D$ and $\exists x \exists y(\alpha \in QUD_{d1}(x)$ and $x \in M_{d1}$ and $\alpha \in QUD_{d2}(y)$ and $x \in M_{d2})\}$

   to the value given by $\frac{|Strat(d_1,\alpha) \cap Strat(d_2,\alpha)|}{0.5(|Strat(d_1,\alpha)|+|Strat(d_2,\alpha)|)}$

6. **Pert** is a *function*

   from $\{\langle D, \alpha\rangle | D$ is the set of c-discourses in a mass discourse sample$\}$

   to $\{x | x \in D$ and $Strat(x, \alpha)$ is defined$\}$

3.4 Measuring Deliberative Polarization

In what follows, I informally characterize two dimensions of deliberative polarization propaganda, and then demonstrate how the model developed above allows us to articulate indices for these two dimensions. These two dimensions are *siloization* and *insularity*. Deliberative polarization on a policy question requires that a mass discourse feature c-discourses which vary widely in the deliberative strategies they enact with respect to this question. This is what I'll be calling siloization. But it also requires that individual agents' access profiles are skewed toward discourses that enact some deliberative strategies rather than others. This is what I'll call insularity.

3.4.1 Silos

Siloization is the degree to which the c-discourses on an issue are split between different deliberative strategies.

For a sample of mass discourse, B, subsets S_1 and S_2 of D_B such that $|S_1| = n$ and $|S_2| = m$ are silos with respect to a question, q iff:

$$\left[\binom{n}{2}^{-1}\left(\sum_{i=d_1S_1}^{d_nS_1}\sum_{j=d_1S_1}^{d_nS_1} CStrat(i,j,q)\right)\right] > \left[\frac{1}{nm}\left(\sum_{i=d_1S_1}^{d_nS_1}\sum_{j=d_1S_2}^{d_mS_2} CStrat(i,j,q)\right)\right]$$

and

$$\left[\binom{m}{2}^{-1}\left(\sum_{i=d_1S_2}^{d_mS_2}\sum_{j=d_1S_2}^{d_mS_2} CStrat(i,j,q)\right)\right] > \left[\frac{1}{nm}\left(\sum_{i=d_1S_1}^{d_nS_1}\sum_{j=d_1S_2}^{d_mS_2} CStrat(i,j,q)\right)\right]$$

This first equation represents a state of affairs where the average similarity in deliberative scripts with respect to q among elements of S_1 is greater than the average similarity in deliberative scripts between elements of S_1 and S_2, and the second a state of affairs where the average similarity in deliberative scripts with respect to q among elements of S_2 is greater than the average similarity in deliberative scripts between elements of S_1 and S_2.

Discourses which feature silos with respect to a given question may do so to differing degrees. We can quantify such differences using a Siloization Index (SI). The SI is the average difference between a) how similar elements of each respective silo are to their silo-mates and b) how similar they are to

their non-silo-mates. For a sample of mass discourse, B, which has silos S_1 and S_2 with respect to a question, q such that $|S_1| = n$ and $|S_2| = m$:

$$\mathbf{SI}_{B,q,S_1,S_2} =$$

$$\frac{1}{2}\left[\left[\binom{n}{2}^{-1}\left(\sum_{i=d_1S_1}^{d_nS_1}\sum_{j=d_1S_1}^{d_nS_1} CStrat(i,j,q)\right)\right] + \left[\binom{m}{2}^{-1}\left(\sum_{i=d_1S_2}^{d_mS_2}\sum_{j=d_1S_2}^{d_mS_2} CStrat(i,j,q)\right)\right]\right] - \left[\frac{1}{nm}\left(\sum_{i=d_1S_1}^{d_nS_1}\sum_{j=d_1S_2}^{d_mS_2} CStrat(i,j,q)\right)\right]$$

This definition of silohood is such that a given discourse may have multiple pairs of subsets that constitute silos with respect to a single question. Using the above, we can also characterize a notion of maximal silohood:

Maximal Silos: For a sample of mass discourse, B, with subsets S_1 and S_2 of D_B such that $|S_1| = n$ and $|S_2| = m$, S_1 and S_2 are maximal silos with respect to a question, q iff:

1. $S_1 \cup S_2 = Pert(D, q)$[5]
2. There are no silos of B with respect to q that generate a higher SI than do S_1 and S_2.

This gives rise to an intuitive notion of a Maximal Siloization Index, which is the SI for a question in a sample of mass discourse calculated based

[5]There is good reason to ask whether we should really want all q-addressing discourses to be sorted into silos, rather than allowing some discourses to go unsiloed. Without something like this provision however, B's maximal SI might be based on relatively small clusters of c-discourses, and would fail to capture the more general distribution across D. This issue is one I leave unsettled here.

on that discourse's maximal silos with respect to that question. We might adopt the convention of understanding an SI to be a maximal SI whenever the silos go unspecified in its subscript, i.e. $\mathrm{SI}_{B,q}$ is the maximal SI for B and q.

3.4.2 Insularity

By itself, a high Siloization Index isn't necessarily an indication that the public is split between widely divergent deliberative scripts on an issue, and so isn't a good measure of DPP. To know whether the public is split between widely divergent deliberative scripts, we need to know more about who is accessing discourses in the respective silos. Because a given agent can, in principle, access an unlimited number of discourses, there's nothing preventing them from accessing all silos. The additional information one wants in order know whether high siloization actually amounts to a significant split in the way that members of the public deliberate about an issue (i.e. DPP) has to do with which agents access the discourses of each silo. A measure of this feature is what I'll call the Insularity Index (II). For a particular question, q, and a set of agents A_B, the Insularity Index is the proportion of agents in A_B who access *any* discourse that addresses q, for whom it's the case that more than, say, 80% of their accessed q-discourses come from just one silo. For a sample of mass discourse, B, with silos (S_1, S_2...S_n) with respect to the question q at time t:

$\mathbf{II}_{B,q}$=

$$\frac{\left|\left\{a : a \in A_B \text{ and } \exists n\left(\frac{|\{d{:}d \in S_n \text{ and } aRd\}|}{|\{d{:}d \in Pert(D_B,q) \text{ and } aRd)\}|} > 0.8\right)\right\}\right|}{|\{a : a \in A_B \text{ and } \exists d(d \in Pert(D_B, q) \text{ and } aRd)\}|}$$

We will assume, as above, that where no silos are specified in the subscript of II this is the insularity score for the maximal silos of B vis-à-vis q. The exact threshold of 80% is more or less arbitrary, and any implementation of this index could customize this aspect. Of note is that a sample of discourse with no silos regarding a question would have no insularity score with respect to that question.

Together, Siloization and Insularity Indices give a sense of how divided a group of agents is with respect to their deliberation on an issue. In the paradigm developed above, these scores are then indicators of Deliberative Polarization Propagandicity.

3.5 A Further Theoretical Advantage of the Approach

Now that my full approach to operationalizing Deliberative Polarization Propagandicity has been developed, I want briefly to mention a further advantage of the approach vis-à-vis its relations to certain theoretical commitments in democratic theory. Specifically, I want to address the capacity of this approach to reflect those dimensions of propaganda which go beyond the epistemic. Consider the American electorate. In the US, significant portions of the electorate lack even basic political knowledge concerning, for instance,

which party controls Congress, the composition of the Supreme Court, or the basic characteristics of "liberalism" or "conservatism"; this seems to have been the case for as long as polls assessing citizens' level of political knowledge have been administered (Converse (1964/ 2006), Lupia (2016)). The political beliefs of what are, in this literature, referred to as "unsophisticated" agents are marked by a low degree of temporal stability; a susceptibility to alteration by the language of opinion polls (i.e. priming and framing effects); and a low degree of what has come to be called "belief system constraint" (Converse (1964/ 2006), Zaller (1992)). Belief systems are thought of as interdependent beliefs where interdependence may amount to logical entailment between beliefs, or else something more like ideological coherence. Those with less political knowledge form new political beliefs under fewer such constraints. This means that they may not see the connection between, say, decreasing tax revenue and cutting social welfare spending, or between the former trade-off and the ideology of small-government conservatism.

Notably, party identification seems to be much more durable over time, and much more likely to constrain one's political behaviors and preferences. It's just that this party identification is not, as an optimist might conjecture from the armchair, reliably selected to match one's antecedent political beliefs. Of course, people do switch party loyalties; but even in these cases, it's been suggested, their doing so is a consequence of a new party tapping into their "group identity" (Achen & Bartels (2016)), as was the case with US Republicans' "Southern strategy" of the 60s and 70s. Social groups, be

they ethnic, religious, class-based or otherwise, then replace party allegiance as primary determinants of political behavior, but this paints no rosier picture of the electorate's rationality.

Those in a position to influence the behavior of mass publics know all this. They know, therefore, that speaking to voters' group loyalties is an effective way to influence their political behavior while bypassing the onerous task of rational persuasion. This picture of political psychology comes in weak and strong varieties, which I distinguish here for the sake of precision rather than any attempt to arbitrate between them. The weak view is that tapping into group-identities can cause voters to adopt certain policy beliefs, which in turn dispose them toward certain political behavior, without offering rationale bases for these beliefs. The strong thesis is that tapping into group identities can cause voters to take up a commitment to certain political behavior *without* this behavior being undergirded by *any* policy beliefs, though policy beliefs may be backfilled to justify one's behavioral commitments. And behavior, be it voting behavior or consumer behavior, is for most purposes all that it is in political or financial elites' interest to try to influence.

In recent work, Cory Wimberley nicely sums up the consequences of the (weak) view of group-identity political psychology for our characterization of propaganda:

> Propagandists are not epistemologists and are not focused on a battle to prove things true or false; they are governors seeking

> to alter the conduct of the public through tactical interventions into public opinion: What interest have they in truth or falsehood except as it produces the conduct that their clients want to see? (Wimberly (2017), p. 115)

Nor is this observation a new one; Jacques Ellul expressed a similar thought half a century ago:

> Very frequently propaganda is described as a manipulation for the purpose of changing idea or opinions of making individuals 'believe' some idea or fact, and finally of making them adhere to some doctrine–all matters of the mind... This is a completely wrong line of thinking: to view propaganda as still being what it was in 1850 is to cling to an obsolete concept of man and of the means to influence him; it is to condemn oneself to understand nothing about propaganda. The aim of modern propaganda is no longer to modify ideas, but to provoke action. (Ellul (1973), p. 25)

Rather than being messaging which disposes political agents to form certain political *beliefs* or conduct theoretical reasoning in a certain way about political matters, propaganda should be seen as propagating certain political *behavior*. This characterization of propaganda is strictly neutral between the strong and weak psychological theses sketched above; that is, whether or not political messaging does typically route through (rational) policy belief formation on its way to influencing behavior, where this messaging counts as propaganda, the influence on behavior is the point. Presupposing for the

moment a simple view of behavior's psychological genesis, I submit that the theorist of propaganda does well to regard propaganda as having effects on both beliefs *and* other attitudes, like desires.

Wimberley highlights propaganda's capacity to produce particular kinds of "subjectivity"; that is, to leverage our antecedent desires or interests in order to generate in us new, instrumental desires, and so remake us as ideal consumers (p. 111: 'Buy this fitness equipment and you will look like a Greek god in sixty days!'). But I have in mind a different way in which propaganda works on our non-doxastic attitudes. This is the alteration of *organizing structures* on our desires, such as the relationships we have to others.

Suppose my neighborhood wakes up one morning to posters on every telephone pole and bulletin board declaring that, because of some innate characteristic, certain community members are dangerous, scheming, and apt to carry disease. I hope that we would unequivocally denounce these messages, and that untargeted members of the community would offer support to those targeted. But what if some suggested that there *was*, after all, some grain of truth in what was said, that perhaps we *should* prevent our children associating with those of the dangerous group? Regardless of how this story continues, relationships between neighbors have already changed; some suspect that their neighbors are all the things suggested by the posters, and others come to mistrust their untargeted neighbors because they don't know who might have been thus taken in. Our relationships have been changed, indeed harmed, by propaganda.

That propaganda's effect on relationships may normally be mediated by beliefs (almost certainly true) is not a threat to my claim here; what is essential is that the *point* of propaganda is often to alter relationships, just as much as it is to instill certain beliefs. And the variety of relationship which propaganda might be said to characteristically undermine is that of civic friendship. This point is anticipated in recent literature which argues that the absence of co-deliberation on policy issues, in addition to undermining ideals of public reason in the way I elaborated on in §2, undermines relationships of civic friendship (Ebels-Duggan (2010); Lister (2013)). A notion of propagandicity that looks at the sort of latent deliberative networks existing, and not existing, among agents in a polity, as ours does, then tracks both beliefs and (crucial structures on) desires, the two crucial psychological antecedents to behavior which it is propaganda's ultimate goal to control. It is a strength of this approach that it proposes the forensic utility of tracking shifting deliberative relationships among polity members.

3.6 Conclusion

In the foregoing, I have characterized those features of propaganda that I think a discourse-structural analysis helps lay bare. I have also provided a way of conceiving of mass discourse as a network among count discourses and agents and precisified characterizations of those features of mass discourse which indicate system-wide propagandistic distortion. While regarding propaganda as a property of systems has been productive, it may well be objected

that this approach renders it questionable how to treat a diagnosis of propaganda. What if it turns out a sample of mass discourse has high siloization and insularity scores on many of its most pressing policy questions–what then? Even well-intentioned and relatively influential agents in such a system might wonder what unilateral actions they could take to alter such a massively systemic issue. Likewise, an noninfluential media consumer, wanting to guard themselves against manipulation, might be left at a loss. The basically forensic nature of the project carved out here is incomplete until methods of defense and counteroffensive are added to a method of detection. In the next stage of this project, I plan to offer a way to identify particular information venues as vectors of DPP; with this further diagnostic tool come a suite of therapeutic ones for both consumers and distributors of political information.

It is also worth noting that I have motivated the development of a model of deliberative polarization (of the kind that I think constitutes a form of propaganda) in terms of political concerns, but this tool has applications beyond the study of political polarization. Political issues are just one kind of issue with respect to which a population can be deliberatively polarized. Those with broadly Kuhnian or Feyerabendian views on the history and the philosophy of science might be seen as describing deliberative polarization as part of the mechanism of radical theory change. Likewise, the split between analytic and continental philosophers on issues with which they are both concerned is likely characterizeable partially in terms of deliberative polarization. These are applications I hope to explore in the future.

In the mean time, the central hope here has been that, in developing a way of identifying propaganda that abstracts away from the partisan contents of political messaging, we are empowered to discuss those first principles which bear on our ideals for our polity's discourse without being stymied by the very political polarization that may constitute this discourse's primary ailment.

Bibliography

Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, *70*(2), 542–555.

Achen, C., & Bartels, L. (2016). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton Studies in Political Behavior. Princeton University Press.

Anderson, A. A., & Belnap, N. D. (1975). *Entailment: The Logic of Relevance and Necessity*. Princeton University Press.

Ansolabehere, S., Rodden, J., & Jr., J. M. S. (2006). Purple america. *Journal of Economic Perspectives*, *20*(2), 97–118.

Arceneaux, K., & Johnson, M. (2013). *Changing Minds or Changing Channels?: Partisan News in an Age of Choice*. Chicago Studies in American Politics. University of Chicago Press.
URL `https://books.google.com/books?id=YZdyvQ9R\_hEC`

Arendt, H. (1951/1994). *The Origins of Totalitarianism*. Harcourt.

Armstrong, D. (1973). *Belief, Truth, and Knowledge*. Cambridge University Press.

Bar-Hillel, Y., & Carnap, R. (1953). Semantic information. In *Proceedings of the Symposium on Applications of Communication Theory*. Butterworth Scientific Publications.

Beaver, D., & Clark, B. (2008). *Sense and Sensitivity: How Focus Determines Meaning*. Wiley-Blackwell.

Begby, E. (2013). Semantic minimalism and the of communication". *Philosophical Studies*, *165*(3), 957–973.

Bezuidenhout, A. (1997). The communciation of de re thoughts. *Noûs*, *31*(2), 197–225.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Buchanan, R. (2010). A puzzle about meaning and communication. *Noûs*, *44*(2), 340–371.

Buchanan, R. (2013). Conversational implicature, communicative intentions, and content. *Canadian Journal of Philosophy*, *43*(5), 720–740.

Buchanan, R. (2014). Reference, understanding, and communication. *Australasian Journal of Philosophy*, *92*(1), 55–70.

Clark, M. (1963). Knowledge and grounds. a comment on mr. gettier's paper. *Analysis*, *24*(2), 46–48.

Converse, P. E. (1964/ 2006). The nature of belief systems in mass publics. *Critical Review*, *18*(1-3), 1–74.

Converse, P. E. (1987). Changing conceptions of public opinion in the political process. *The Public Opinion Quarterly*, *51*, S12–S24.

Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review*, *18*(1-3), 1–74.

Davis, N. T., & Dunaway, J. L. (2016). Party polarization, media choice, and mass partisan-ideological sorting. *Public Opinion Quarterly*, *80*, 272 – 297.

Davis, W. (1998). *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory*. Clarendon Press.

Ebels-Duggan, K. (2010). The beginning of community: Politics in the face of disagreement. *Philosophical Quarterly*, *60*(238), 50–71.

Eberle, C. J. (2002). *Religious Conviction in Liberal Politics*. Cambridge University Press.

Ellul, J. (1973). *Propaganda: the Formation of Men's Attitudes*. A Vintage Book. Vintage Books.

Fernández, R., & Ginzburg, J. (2002). Non-sentential utterances: Grammar and dialogue dynamics in corpus annotation. In *Proceedings of the 19th international conference on Computational linguistics*, vol. 1, (pp. 1–7). Association for Computational Linguistics.

Fiorina, M., Abrams, S., & Pope, J. (2005). *Culture War?: The Myth of a Polarized America*. Great questions in politics series. Longman.

Fiorina, M., & Abrams, S. J. (2008). Political polarization in the american public. *Annual Review of Political Science*, *11*, 563–588.

Foucault, M. (2002). *Archaeology of Knowledge*. Classics Series. Routledge.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Ginzburg, J. (1994). An update semantics for dialogue. In *Proceedings of the Tilburg International Workshop on Computational Semantics*.

Ginzburg, J. (1995a). Resolving questions, part i. *Linguistics and Philosophy*, *18*(5), 459–527.

Ginzburg, J. (1995b). Resolving questions, part ii. *Linguistics and Philosophy*, *18*(6), 567–609.

Ginzburg, J. (1996). Dynamics and the semantics of dialogue. In J. Seligman, & D. Westerstahl (Eds.) *Language, Logic and Computation*. CSLI.

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Communication*. Oxford University Press.

Ginzburg, J., & Macura, Z. (2007). Lexical acquisition with and without metacommunication. In *Emergence of Communication and Language: Part III*, vol. 3, (pp. 287–301). Springer London.

Glaeser, L., Edward, & Ward, B. A. (2006). Myths and realities of american political geography. *Journal of Economic Perspectives*, *20*(2), 119–144.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, *73*(20), 771–791.

Grice, H. (1957). Meaning. *Philosophical Review*, *66*(3), 377–388.

Grice, H. (1967). Logic and conversation. In D. Davidson, & G. Harman (Eds.) *The Logic of Grammar*. Dickenson.

Grice, H. (1969). Utterer's meaning and intention. *Philosophical Review*, *78*(2), 147–177.

Groenendijk, J., & Stokhof, M. (1984a). *Studies on the semantics of questions and the pragmatics of answers*. Ph.D. thesis, Institute for Logic, Language & Computation (ILLC), University of Amsterdam.

Groenendijk, J., & Stokhof, M. (1984b). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam, Department of Philosophy.

Habermas, J. (1979). *Communication and the Evolution of Society*. Beacon Press.

Habermas, J. (1984). *The Theory of Communicative Action*, vol. 2. Beacon Press.

URL `https://books.google.com/books?id=JdCzHO3ErngC`

Habermas, J. (2015). *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society*. Wiley.

Habermas, J., & McCarthy, T. (1985). *The Theory of Communicative Action*, vol. 1. Beacon Press.
URL https://books.google.com/books?id=kuFhjNZuHTAC

Hamblin, C. L. (1973a). Questions in montague english. *Foundations of Language: International Journal of Language and Philosophy*, *10*(1), 41–53.

Hamblin, C. L. (1973b). Questions in montague english. *Foundations of Language*, *10*(1), 41–53.

Heck, R. (1995). The sense of communication. *Mind*, *104*(413), 79–106.

Heck, R. (2014). Semantics and context dependence: Towards a strawsonian account. In A. Burgess, & B. Sherman (Eds.) *Metasemantics: New Essays on the Foundations of Meaning*, (pp. 327–365). Oxford University Press.

Herman, E., & Chomsky, N. (2011). *Manufacturing Consent: The Political Economy of the Mass Media*. Knopf Doubleday Publishing Group.

Hyland, K. (2005). *Metadiscourse*. A&C Black.

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, *59*(3), 690–707.

Jordan, S., & Bowling, C. J. (2016). Introduction: The state of polarization in the states. *State and Local Government Review*, *48*(4), 220–226. URL http://dx.doi.org/10.1177/0160323X17699527

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.

Langton, R. (2009). *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*. OUP Oxford. URL https://books.google.com/books?id=N-ITDAAAQBAJ

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(339-359).

Lister, A. (2013). *Public Reason and Political Community*. Bloomsbury Research in Political Philosophy. Bloomsbury Publishing.

Loar, B. (1976). The semantics of singular terms. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *30*(6), 353–377.

Lupia, A. (2016). *Uninformed : Why People Seem to Know So Little About Politics and What We Can Do About It.*. Oxford University Press.

MacIntyre, A. (2007). *After Virtue: A Study in Moral Theory, Third Edition*. University of Notre Dame Press.

Marlin, R. (2002). *Propaganda and the Ethics of Persuasion*. Broadview Press.

Murray, S. E. (2014). Varieties of update. *Semantics and Pragmatics*, *7*, 1–53.

Nacos, B., Shapiro, R., & Isernia, P. (2000). *Decisionmaking in a Glass House: Mass Media, Public Opinion, and American and European Foreign Policy in the 21st Century*. Rowman & Littlefield Publishers.
URL `https://books.google.com/books?id=ueCxAAAAQBAJ`

Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.

Page, B., & Shapiro, R. (2010). *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. American Politics and Political Economy Series. University of Chicago Press.
URL `https://books.google.com/books?id=3R9XFfX4bfcC`

Pagin, P. (2008). What is communicative success? *Canadian Journal of Philosophy*, *38*(1), 85–116.

Pattie, C., & Johnston, R. (2016). Talking with one voice? conversation networks and political polarisation. *British Journal of Politics and International Relations*, *18*(2), 482 – 497.

Paul, M. (1999). *Success in Referential Communication*. Springer.

Peet, A. (2017). Referential intentions and communicative luck. *Australasian Journal of Philosophy*, *95*(2), 379–384.

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford University Press.

Purver, M., Ginzburg, J., & Healey, P. (2001). On the means for clarification in dialogue. In J. van Kuppevelt, & R. Smith (Eds.) *Current and New Directions in Discourse and Dialogue*, (pp. 235–256). Kluwer.

Récanati, F. (1993). *Direct Reference: From Language to Thought*. Oxford, Blackwell Publishers.

Roberts, C. (1996). Information structure: Towards an integrated formal theory of pragmatics. In A. K. Je Hak Yoon (Ed.) *Ohio State University Working Papers in Linguistics (OSUWPL) Volume 49: Papers in Semantics*. The Ohio State University Department of Linguistics.

Roberts, C. (2004). Context in dynamic interpretation. In L. Horn, & G. Ward (Eds.) *The Handbook of Pragmatics*. Blackwell.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*(6), 1–69.

Roberts, C. (2012b). Information structure: Afterword. *Semantics and Pragmatics*, *5*(7), 1–19.

Rooth, M. (1985). *Association with focus*. Ph.D. thesis, University of Massachusetts Amherst Department of Linguistics.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116.

Schiffer, S. (2017). Intention and convention in the theory of meaning. In B. Hale, C. Wright, & A. Miller (Eds.) *A Companion to the Philosophy of Language*. John Wiley & Sons, 2nd ed.

Simons, M., Beaver, D., Roberts, C., & Tonhauser, J. (2010). What projects and why. In N. Li, & D. Lutz (Eds.) *Semantics and Linguistic Theory (SALT 20)*, (pp. 309–327). CLC Publications.

Simons, M., Beaver, D., Roberts, C., & Tonhauser, J. (2013). Toward a taxonomy of projective content. *Language*, *89*(1), 66–109.

Simons, M., Beaver, D., Roberts, C., & Tonhauser, J. (2017). The best question: Explaining the projection behavior of factives. *Discourse Processes*, *54*(3), 187–206.

Sosa, E. (1999). How to defeat opposition to moore. *Noûs*, *33*(s13), 141–153.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Blackwell.

Sperber, D., & Wilson, D. (2015). Beyond speaker's meaning. *Croatian Journal of Philsoophy*, *15*(44), 117–149.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.) *Syntax and semantics, volume 9: Pragmatics*, (pp. 315–322). Academic Press.

Stalnaker, R. (2002a). Common ground. *Linguistics and Philosophy*, *25*, 701–721.

Stalnaker, R. (2002b). Common ground. *Linguistics and Philosophy*, *25*(5-6), 701–721.

Stanley, J. (2015). *How Propaganda Works*. Princeton University Press.

Tonhauser, J. (2012). Diagnosing (not-)at-issue content. In *Proceedings of Semantics of Under-represented Languages of the Americas (SULA) 6*, (pp. 239–254). GLSA.

van Rooij, R. (2003). Negative polarity items in questions. *Journal of Semantics*, *20*, 239–273.

von Fintel, K., & Gillies, A. S. (2011). 'might' made right. In A. Egan, & B. Weatherson (Eds.) *Epistemic Modality*, (pp. 108–130). Oxford University Press.

Ward, G. (2004). Equatives and deferred reference. *Language*, *80*(2), 262–289.

Wimberly, C. (2017). The job of creating desire: Propaganda as an apparatus of government and subjectification. *Journal of Speculative Philosophy*, *31*(1), 101–118.

Zagzebski, L. T. (1996). *Virtues of the Mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

Zagzebski, L. T. (2004). The search for the source of epistemic good. *Metaphilosophy*, *34*(1-2), 12–28.

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press. URL https://books.google.com/books?id=mLghAwAAQBAJ

Zwicky, A. M., & Sadock, J. M. (1975). Ambiguity tests and how to fail them. In J. P. Kimball (Ed.) *Syntax and Semantics 4*, (pp. 1–36). Academic Press.