



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Interpreting silent gesture

Citation for published version:

Schouwstra, M, De Swart, HE & Thompson, B 2019, 'Interpreting silent gesture: Cognitive biases and rational inference in emerging language systems', *Cognitive Science*, vol. 43, no. 7, e12732. <https://doi.org/10.1111/cogs.12732>

Digital Object Identifier (DOI):

[10.1111/cogs.12732](https://doi.org/10.1111/cogs.12732)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Cognitive Science

Publisher Rights Statement:

This is the peer reviewed version of the following article: Schouwstra, M. , Swart, H. and Thompson, B. (2019), Interpreting Silent Gesture: Cognitive Biases and Rational Inference in Emerging Language Systems. *Cogn Sci*, 43: e12732. doi:10.1111/cogs.12732, which has been published in final form at <https://onlinelibrary.wiley.com/doi/full/10.1111/cogs.12732>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Interpreting Silent Gesture: cognitive biases and rational inference in emerging language systems

Marieke Schouwstra*
Centre for Language Evolution
University of Edinburgh
Marieke.Schouwstra@ed.ac.uk

Henriëtte de Swart
Utrecht Institute of Linguistics OTS
Utrecht University
h.deswart@uu.nl

Bill Thompson
Social Science Matrix
University of California, Berkeley
billdthompson@berkeley.edu

April 3, 2019

Abstract

Natural languages make prolific use of conventional constituent-ordering patterns to indicate ‘who did what to whom’, yet the mechanisms through which these regularities arise are not well understood. A series of recent experiments demonstrates that, when prompted to express meanings through silent gesture, people bypass native language conventions, revealing apparent biases underpinning word order usage, based on the semantic properties of the information to be conveyed. We extend the scope of these studies by focusing, experimentally and computationally, on the *interpretation* of silent gesture. We show cross-linguistic experimental evidence that people use variability in constituent order as a cue to obtain different interpretations. To illuminate the computational principles that govern interpretation of non-conventional communication, we derive a Bayesian model of interpretation via biased inductive inference, and estimate these biases from the experimental data. Our analyses suggest people’s interpretations balance the ambiguity that is characteristic of emerging language systems, with ordering preferences that are skewed and asymmetric, but defeasible.

*corresponding author

1 Introduction

2 1.1 Language production without conventions: evidence from silent gesture

4 When people do not have an existing set of linguistic rules to use to communicate,
they use principles for structuring their utterances that are independent of their native
6 language. This has been observed in lab experiments where naïve adult participants
are asked to describe simple events using only gesture and no speech (silent gesture).
8 The silent gesture paradigm has been used to investigate several core features of lan-
guage, such as how a communication system can be bootstrapped through iconicity
10 (Fay, Arbib, & Garrod, 2013). In particular, the paradigm has provided notable in-
sight into the origins of the ordering of Subject, Object and Verb in human language.¹
12 For instance, it has been shown that when people describe transitive actions through
space in this paradigm, they prefer SOV word order, irrespective of the dominant order
14 of their native language (Goldin-Meadow, So, Özyürek, & Mylander, 2008). Given
the dominance of SOV in emerging language systems (e.g., Padden, Meir, Sandler,
16 & Aronoff, 2010), it has been suggested that SOV may have been important in the
emergence of language in humans (Newmeyer, 2000; Givon, 1997). However, more
18 recent publications show that under certain circumstances SOV is not the dominant or-
der (Meir, Lifshitz, İlkbasaran, & Padden, 2010; Langus & Nespors, 2010; Gibson et al.,
20 2013; Hall, Mayberry, & Ferreira, 2013; Schouwstra & de Swart, 2014; Schouwstra,
2017). Investigation of this variability in word order has sparked a debate about the
22 mechanisms that play a role when people communicate in the absence of a shared
linguistic system, and, indirectly, about the conventionalisation of word order in the
24 emergence of language. The silent gesture paradigm is relatively new, and many ques-
tions are still unanswered. However, two semantic distinctions, that between reversible
26 and non-reversible events and that between extensional and intensional events, have
been studied in some detail, and provide a picture of how semantic information can

¹We recognise that in improvised gesture, where there are no linguistic conventions for word order, the (syntactic) terms Subject, Object, and Verb are not meaningful. We will use them as a convenient shorthand for the more appropriate terms Agent, Patient and Action.

28 influence word order in emerging language.

1.2 Semantic properties influence constituent structure in silent 30 gesture production

Whether or not an event is reversible (typically events in which there are two animates,
32 such as ‘woman kicks man’) influences the word order that is used (Gibson et al., 2013;
Hall et al., 2013). The usage of SOV ordered strings drops for reversible events, and
34 SVO usage becomes more likely.² Various explanations for the phenomenon have been
offered, but there is no conclusive evidence for whether the pattern is rooted in com-
36 municative or cognitive principles (or even potentially the result of modality-specific
processes; (Gibson et al., 2013; Hall et al., 2013; Kline, Salinas, Lim, Fedorenko, &
38 Gibson, 2017).

Another semantic effect on word order variation was observed by Schouwstra and
40 de Swart (2014), who compared two semantic classes of transitive events: extensional
and intensional events. The former is a class of events in which a direct object is
42 manipulated in an action through space, similarly to the motion events used by Goldin-
Meadow et al. (2008). Some examples are throwing (‘pirate throws guitar’) or carrying
44 (‘princess carries ball’) events. Intensional events (e.g., ‘pirate searches for guitar,’
‘princess thinks of ball’, but also ‘cook hears violin’ and ‘witch builds house’) are typ-
46 ically described using intensional verbs, and for the interpretation of such descriptions,
the intension (meaning) of their arguments, and in particular the direct object, is more
48 important than the extension (object in the world). This makes the direct object more
abstract, and possibly non-existent or non-specific.

50 Schouwstra and de Swart (2014) show that in silent gesture, participants prefer to
use SVO word order over SOV for intensional events, and SOV order over SVO for
52 extensional events. They observe that word order flexibility on the basis of such mean-
ing differences in the verb do not exist in fully conventional languages, and argue that

²Note that more recently, it was argued that rather the effect might be the result of a preference to describe human participants first, and this would mean that reversibility is not the crucial factor (Kocab, Lam, & Snedeker, 2018; Meir et al., 2017). These studies discuss evidence from silent gesture and emergent sign languages, but the effect is well established in spoken language production too (Branigan, Pickering, & Tanaka, 2008).

54 they are typical for situations where there are no (or where there is only a limited set
of) linguistic conventions: people use their cognitive biases (rooted in the semantic
56 properties of events) and build their improvised utterances flexibly, according to these
biases. This position contrasts with previous hypotheses in which word order in emerg-
58 ing language systems is seen as something rigid rather than variable (Newmeyer, 2000;
Goldin-Meadow et al., 2008). The distinction between intensional and extensional
60 events turns out to be of influence on constituent order, not only in the gestural domain,
but also in the vocal domain, as shown in a study in which participants improvise to
62 produce non-word sounds to convey information (Mudd, Kirby, & Schouwstra, 2018),
a finding that is interesting given potential issues about modality specificity; see above,
64 and Kline et al. (2017).

All in all, the improvised gesture paradigm can reveal pressures that are important
66 when there is no system of linguistic conventions in place, and thus help reveal the
process that takes us from no language to full linguistic regularity in a controlled labo-
68 ratory setting. This setting allows us to study not only improvised production, but also
other processes that play a role in language use, such as interpretation, communicative
70 interaction, cultural transmission. In this paper we will take one step from improvised
gesture production toward full linguistic systems, by focusing on the *interpretation* of
72 improvised gesture, and comparing it to its production. We will do this by employing
a novel combination of a silent gesture experiment and an experimentally-informed
74 Bayesian model.

1.3 Silent gesture: production vs. interpretation

76 If silent gesture is to offer a comprehensive test ground for communication without
existing conventions, it should not only concern production, as communication is a
78 process with two directions: production and interpretation. These two directions may
exert different pressures in the emergence of a language system (Burling, 2000; Mac-
80 Donald, 2013).

The interpretation of strings in the silent gesture paradigm has received little atten-
82 tion, with the exception of two recent studies: one in which participants are asked to

recognise the intended meaning of silent gesture strings in a timed forced choice setup
84 (Langus & Nespors, 2010), and one in which participants are asked to choose an interpretation for ambiguous reversible events (Hall, Ahn, Mayberry, & Ferreira, 2015).

86 Langus and Nespors (2010) asked adult participants to watch video clips of gesture sequences describing simple transitive events through space in a two alternative forced
88 choice task. Participants, native speakers of Italian (SVO) and Turkish (SOV), saw video clips in all possible orderings of S,O and V. Both groups of participants showed
90 fastest reaction times for SOV ordered video clips, which shows that, like in production, SOV order is preferred in improvised gesture comprehension, independently of the
92 dominant order of the native language of the observer. In other words, when naive observers are presented with improvised gesture, they by-pass the dominant patterns of
94 their native language. Langus and Nespors (2010) claim that this effect is due to the fact that in this task, participants disregard their computational system of grammar.

96 Hall et al. (2015) focused on the interpretation of reversible events, and come to very different conclusions. They showed participants silent gesture strings that were
98 made up of an action and two animate participants, in three possible orders (Action-Participant1-Participant2, Participant1-Action-Participant2, Participant1-Participant2-
100 Action). Each order was ambiguous: it was not made explicit which participant had the role of agent and which patient. For each string, participants were asked to choose an
102 interpretation from two line drawings: one with the first participant in the role of agent, and one with the second participant in the role of agent. They found that participants
104 take the element mentioned first in the gesture string to be the agent, i.e., ‘woman man push’ is interpreted most robustly with the woman in the role of the pusher.

106 They conclude that interpretation of these ambiguous strings is governed by a semantic constraint, ‘agent first’, and they emphasise the difference between interpretation
108 and production: the latter is motivated by production constraints—i.e., gesturers will often use their own body to take on roles of the event participants, and using SOV
110 word order involves more ‘role switches’ than using SVO order, which makes SVO more fluent than SOV (Hall et al., 2013; Hall, Ferreira, & Mayberry, 2014).

112 To summarise, silent gesture investigates the cognitive constraints that play a role

when a system of linguistic conventions is not in place. Investigating production and
114 interpretation of silent gesture can help us gain insight into how these two processes
contribute to an emerging linguistic system. From what we have seen above it is not
116 entirely clear how production and interpretation, in the absence of linguistic conven-
tions, relate to each other. Hall et al. (2015) emphasise the difference between silent
118 gesture production and interpretation. They postulate procedural, production-related
constraints for production, and a semantic heuristic ('agent first') for interpretation.
120 Langus and Nespors (2010), on the other hand, emphasise the similarities between silent
gesture production and interpretation: both are governed, not by grammatical rules, but
122 by cognitive constraints.

We add crucial evidence to the question whether production and interpretation of
124 improvised language are intrinsically similar or rather different from each other. Pre-
sented a silent gesture interpretation experiment, along with a Bayesian computational
126 model for the experimental task, we will point out in which respects production is cru-
cially different from interpretation, in emerging language situations. Our starting point
128 is the semantic differences between extensional and intensional events that are driving
word order variability in silent gesture production (Schouwstra & de Swart, 2014). We
130 ask if participants will use these semantic principles when they *interpret* silent gesture,
and what this can tell us about their underlying biases. Our interpretation experiment is
132 the first to mirror a silent gesture production task, and this allows us to investigate the
link between meaning and word order, not only qualitatively ('does word order influ-
134 ence the meaning an interpreter derives?'), but also quantitatively: by specifying com-
putational principles that sub-serve interpretation of silent gesture under uncertainty,
136 we are able to reason backwards from experimental results to a quantitative estimate
of the cognitive biases guiding word order usage. Our estimates of participants' biases
138 align with the pattern of results observed independently in production experiments: our
results suggest skewed but defeasible event-class-conditional word-order preferences,
140 whose effects on silent-gesture interpretation may be mediated by more general princi-
ples of inference under uncertainty.



Figure 1: The figure depicts different stages of an ambiguous action being acted out. This action can be interpreted as ‘build’ or as ‘climb’. The experiment investigates if the order of the constituents in a gesture sequence has an influence on the interpretation of such ambiguous actions.

142 2 Experiment: improvised gesture interpretation

To test if the order of constituents influences the way in which participants interpret
144 gesture strings, we presented participants with video clips of gesture strings with an
ambiguous action (verb) gesture plus its two arguments. An example of an ambiguous
146 action gesture is shown in figure 1. This gesture can be interpreted as a climbing action,
but also as a building action. Together with the constituents ‘witch’ and ‘house’, this
148 results in two possible interpretations: ‘witch climbs house’ (an extensional event), and
‘witch builds house’ (an intensional event). We construed videos in two possible or-
150 ders, SOV and SVO.³ We hypothesised, based on the production results in Schouwstra
and de Swart (2014), plus the similarities between production and interpretation found
152 in Langus and Nespors (2010), that the gesture order would have an influence on inter-
pretation, and predicted that, when engaged in a dual forced choice task (that presents
154 the two possible interpretations as answer options), participants would be more likely to
interpret SVO ordered gesture strings as intensional events than as extensional events,
156 and vice versa.

³Two example videos (‘princess sleeps-on / dreams-of book’) are included in the supporting material.

2.1 Method

158 2.1.1 Participants

Forty one native speakers of Dutch (16 male, 25 female) were recruited from the
160 Utrecht University library in Utrecht, the Netherlands, and forty native speakers of
Turkish (12 male, 28 female) were recruited from the Bogazici University library in
162 Istanbul, Turkey. Note that Dutch is an SVO language in main clauses, while Turkish
is SOV. None of the participants received monetary compensation.

164 2.1.2 Materials

We created video clips showing three gestured elements: an actor, a patient and an
166 action. The three elements for each video were recorded separately, and for each video
clip, three fragments were concatenated using white flash transitions. The *actions* in
168 each video were ambiguous: they could be interpreted as an extensional verb, or an
intensional verb. For each ambiguous action, we created two ambiguous gesture se-
170 quences: one in SVO order and one in SOV order, resulting in 12 pairs of videos. Note
that for each pair of differently ordered strings, we used exactly the same video ma-
172 terial (but ordered differently). The twelve pairs of ambiguous strings were randomly
distributed over two versions such that each version consisted of 6 SOV videos and 6
174 SVO videos, while at the same time, each ambiguous action occurred only once per
version.

176 Four filler items were created: videos of gesture sequences with unambiguous ac-
tions (two intensional and two extensional). For each ambiguous video, two line draw-
178 ings were made, that represented the two alternative interpretations for the ambiguous
items. For each filler, we created one line drawing depicting the right answer, and one
180 depicting the same actor and patient, but a different action.

2.1.3 Procedure

182 The participants were shown videos on a laptop screen in a two alternative forced
choice task; pictures of the corresponding intensional and extensional events were

184 shown as the two answer possibilities. First, two practice items with unambiguous
verbs were shown, followed by the ambiguous items and fillers. The items were pre-
186 sented in random order, and the order was different for each participant. The two
answer possibilities were shown before each video and again afterwards.⁴ The order of
188 the two answer possibilities was randomly determined. The experiment took about ten
minutes to complete.

190 2.2 Analysis and results

The data were analysed using a logit mixed effects regression, implementing the *lme4*
192 package (Bates et al., 2015) in R (R Core Team, 2014).⁵ Our model analysed the
fixed effects of gesture-order and native language (both sum coded) on the interpreta-
194 tion. Participant was included as random intercept⁶ and random slopes of gesture-order
were included for item. The model revealed that participants were slightly more likely
196 to choose an extensional response, as indicated by the model intercept: $\beta = 0.759$,
 $SE = 0.415$, $p = 0.067$. A significant effect of gesture-order was found ($\beta = 0.414$,
198 $SE = 0.103$, $p < 0.001$), but no effect of native language ($\beta = 0.001$, $SE = 0.086$,
 $p = 0.984$).⁷ Figure 2 depicts the proportions of videos interpreted as extensional
200 events, by gesture order.

Accuracy for the filler items was almost at ceiling level, with 98% overall accuracy,
202 and at most 1 wrong answer per participant.

⁴Showing the answer possibilities before the gesture videos was necessary, because a pilot experiment suggested that the task was too hard when we did not show the answers first.

⁵All data and code are provided in the supplementary material folder.

⁶Including random slopes of gesture order resulted in high correlations between fixed and random effect; moreover, the model that implements both random slopes does not reveal an improved fit over the model that was eventually used ($\chi^2=0.000$, $p=.99$).

⁷Upon re-analysis of the video clips we decided to exclude two videos from the results: 'Pirate drops/searches ball' and 'Girl kisses/thinks of doll'. These two videos differ from the others in the sense that the ambiguous actions they depict consist of two sub-gestures, (a 'drop'-gesture followed by a 'search' gesture for the former, and a 'think of' gesture followed by a 'kiss' gesture for the latter) whereas for all other ambiguous actions, only one gesture is used. Including the two deleted item in the analysis still yields significant main effect of gesture-order: $\beta = 0.302$, $SE = 0.075$, $p < 0.001$.

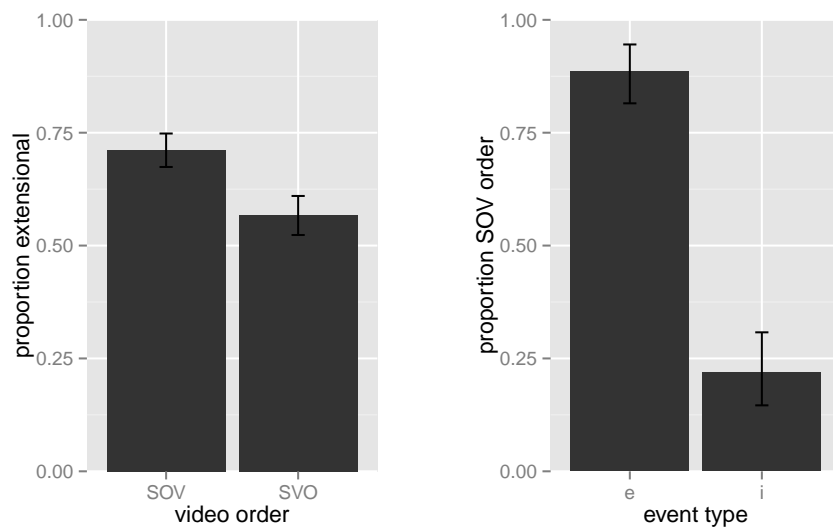


Figure 2: **On the left:** main interpretation results. Mean proportions of videos interpreted as extensional event are shown for SOV and SVO video orders. Error bars indicate 95% confidence intervals. The results show that participants were more likely to interpret SOV ordered videos as extensional than SVO videos. **On the right:** production results from Schouwstra & de Swart (2014) for comparison. Proportion of strings in SOV order are displayed by event type (extensional and intensional). Error bars indicate 95% confidence intervals. These results show that in production participants strongly prefer SOV for extensional events, and SVO for intensional events.

2.3 Baseline study

204 To further investigate the overall preference for extensional interpretations, and to es-
205 tablish a baseline measure for each individual ambiguous action (independent of word
206 order) we carried out an additional experiment. This experiment presented partici-
207 pants with the ambiguous action gestures only, instead of strings containing actor, ac-
208 tion and patient gestures. Data was collected online (N=40; all participants except
209 one were native speakers of English), on a crowdsourcing platform (Crowdfunder; see
210 www.crowdfunder.com).

Like in the full-string experiment, participants were presented with ambiguous ex-
212 perimental items (10) and fillers (4), presented in random order in a two alternative
forced choice task. For each trial, the participant would first see the two possible inter-
214 pretations, presented as line drawings. Subsequently, the participant observed a video
of an ambiguous action gesture; they then saw the two line drawings again, and were
216 asked to select the drawing that they thought best matched the gesture in the video.

Ten experimental responses plus four filler responses per participant were collected.
218 Because the task was a two alternative forced choice task, there were no missing data
points.

2.4 Results: preference for extensional events

220 To compare the overall preference for extensional events in the full-string experiment
221 to that in the verb-only experiment, we merged the data sets. We then ran a logit mixed
222 effects regression to model interpretation, with Experiment (verb-only vs full-string,
223 null-coded) as fixed effect, and random intercepts and random slopes of experiment
for item. A preference for extensional interpretations in the Verb experiment was re-
224 flected in the model intercept: $\beta = 0.953$, $SE = 0.368$, $p < 0.01$. Crucially, no significant
225 effect of experiment was observed ($\beta = -0.215$, $SE = 0.2659$, $p < 0.419$). From this
226 we conclude that the two experiments saw no difference in the level of preference for
227 extensional events.

230 **2.5 Results: model with baseline values**

To use the verb-only experiment results as a by-item baseline for the original (full-string) study, we calculated the proportion of extensional interpretations for each item, resulting in ten baseline values. We incorporated these baseline values into the data for the full-string experiment, by creating normalised responses per trial: we took numeric conversions for the responses per trial (1 for extensional and 0 for intensional), and subtracted the baseline value (based on the item ID), adding 1 to the resulting value.⁸ A linear mixed effects model was performed on these normalised values, taking gesture-order and native language as fixed effects, and random intercepts for item and participant, as well as random slopes for Gesture-order on item. The full model revealed a significantly better fit than the reduced model which only had native language as a predictor ($\chi^2=7.99$, $p<.001$), while no significant difference was found between the full model and the model that omitted native language as a predictor ($\chi^2=0.21$, $p=.90$).

244 **2.6 Discussion**

There are two main conclusions we can draw from the experimental results. First of all, the order in which the ambiguous gesture strings were presented did indeed influence the way they were interpreted by participants: a video clip was more likely to be interpreted as an extensional event when it was presented in SOV order than when it was presented in SVO order, and vice versa. The fact that variability between SOV and SVO is picked up as a cue for interpretation shows that this variability, as it occurs in production, matters for communication.

The second important conclusion is that – in comparison to the results of the production experiment – the effect of word order on meaning in interpretation is modest. For comparison, the right hand graph in figure 2 depicts the effect of meaning on word order, taken from the production results presented in Schouwstra and de Swart (2014). What does this quantitative difference tell us about the nature of word-order biases in this context? Given the striking asymmetries in production, it is tempting to

⁸The latter was done to ensure all values were above 0. The resulting values were all between 0 and 2.

258 expect similarly striking asymmetry in interpretation. This expectation may be mis-
leading, because interpretation involves reasoning under uncertainty. If participants
260 are accounting for this uncertainty, then the impact of the word order biases may be
dampened.

262 For example, the interpretation task – perhaps more so than the production task –
is implicitly interactive: participants are interpreting the behaviour of another speaker.
264 Participants have no knowledge of the speaker’s linguistic system, and may be ac-
counting for this uncertainty when making their decisions. A learner following these
266 principles may be forced to consider disfavoured ordering systems that would be un-
likely to play a role in the participant’s own spontaneous productions: for all but the
268 most strongly biased learners, this could lead to a scenario in which low-level ordering
preferences can drive striking asymmetries in improvised production, but these asym-
270 metries are attenuated by uncertainty during interpretation. To apply a classic analogy:
production can be likened to repeatedly flipping a weighted coin to decide SOV or
272 SVO, one for Extensional and one for Intensional events, where the bias of the coin
corresponds to a low-level semantic bias; interpretation, on the other hand, forces an
274 ideal observer to account for the fact that the gesturer may be holding completely dif-
ferent coins - a more abstract consideration which could lead to uncertainty. In addition
276 to these considerations, any *a priori* bias the observer has toward one event class over
the other could dilute the influence of word-order biases (in a way that would not play
278 a role during production).

These factors, which we will discuss in greater detail below, may break the direct
280 link between biases evident in production and their impact on interpretation. Drawing
conclusions about word-order biases from interpretation implicitly assumes a model of
282 participants’ decisions. In the next section, we develop an explicit model, and use the
model to estimate participants’ biases from our experimental data.

284 **3 Model: A computational Analysis of Gesture Inter-** **pretation**

286 The role of word-order biases in interpretation of improvised gestural communication
has, to the best of our knowledge, received no formal attention whatsoever. While
288 the experimental literature reviewed in section 1.3 provides intriguing hypotheses –
such as the hypothesis that independent heuristics drive production and interpretation
290 (Hall et al., 2015) – there remains no general computational framework for deriving
and testing their quantitative predictions. Here we present a model which allows us
292 to test a simple model of interpretation against the experimental data. Our model is
based around the the idea that non-conventional gestural communication recruits simi-
294 lar biases to production, but the effects of those biases may be mediated by uncertainty.
Our approach is to lay out a simple computational model which formalises the logic
296 discussed here and elsewhere in related literature: the model can be tested against the
experimental data, and can act as a benchmark against which alternative accounts can
298 be contrasted.

The central abstraction in our analysis is that participant behaviour can be pro-
300 ductively broken down into two components: a set of preferences or dispositions that
favour the use of particular orderings in particular contexts; and a procedure for em-
302 ploying these preferences when reasoning about the gesture orderings produced by
another individual – in contexts where the intended meaning is unknown and must be
304 reverse-engineered.

The Bayesian framework provides a natural model for this division of labour. This
306 approach to statistical inference specifies a simple formula describing how a rational
learner should update its beliefs about the nature of an unobserved mechanism respon-
308 sible for generating an observed set of data: under this perspective, the task of a learner
(e.g. a language learner) is to evaluate competing hypotheses about the nature of the un-
310 derlying mechanism in light of the data observed (Perfors et al, 2011). In particular, the
framework allows us to explicitly model biases as *prior distributions*. The principles
312 underpinning Bayesian inductive inference align with human learning in many psy-

chological domains (Chater, Oaksford, Hahn, & Heit, 2010; Griffiths, Chater, Kemp,
314 Perfors, & Tenenbaum, 2010). With respect to language, models of probabilistic ra-
tional inference have been applied to numerous aspects of linguistic structure (Chater
316 & Manning, 2006), including word order generalisations in artificial grammar learning
(Culbertson & Smolensky, 2012), and have been used to model the pragmatic princi-
318 ples underpinning production and interpretation of speech (Goodman & Frank, 2016;
Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013), but have not previously
320 been explored as a model for the learning mechanisms that sub-serve improvised ges-
ture.

322 Interpretation of improvised, not-yet-conventionalised communication is a partic-
ularly exciting focus for computational modelling of this sort because, in terms of the
324 structure of the computational problem facing the interpreter, it has a distinctive char-
acter that is atypical of linguistic communication: the interpreter is – knowingly -
326 largely or completely in the dark with respect to the gesturer’s linguistic system. This
distinguishes improvised gesture from typical artificial language learning scenarios in
328 which the learner is explicitly taught new conventions.

In other words, the improvised gesture interpreter, who we know has certain pro-
330 duction preferences, is faced with data from a producer, but it is unknown to this inter-
preter if the producer was acting according to a system of conventions. Whether and
332 how interpreters account for this uncertainty, and accordingly lean on their own biases
in the absence of helpful evidence about the gesturer, is an open question with impor-
334 tant implications for emerging language systems. By constructing an inferential model
for the experimental task at hand, we can make inroads on this question in a simple
336 problem where, thanks to existing results (Schouwstra & de Swart, 2014), we already
have a good impression of people’s biases in production, allowing cross-validation of
338 our conclusions.

3.1 Interpreting Gestures through Bayesian Inference

Our model casts gesture interpretation as probabilistic inductive inference from an or-
dered gesture g to an unobserved intended meaning m . Given the principles of Bayesian

inference, we model selection of a meaning as a random sample from the posterior distribution over meanings given an observed gesture, $p(m|g)$: the learner arrives at posterior beliefs by combining its *prior expectations* $p(m)$ about the relative probability of meanings m , and the *likelihood* of observing gesture g if m were the true intended meaning. Under this model, the probability of choosing a meaning m as the intended meaning behind an observed gesture g is given by:

$$p(m|g) = \frac{p(g|m)p(m)}{p(g)}, \quad (1)$$

340 where $p(g)$ is simply a normalising constant⁹. Learners' a priori expectations about the probability of each event type, $p(m)$, can be captured with a single parameter λ ,
 342 such that $\lambda = p(m = \text{Extensional}) = 1 - p(m = \text{Intensional})$. However, the likelihood $p(g|m)$ of observing a gesture g in the event that the gesturer were expressing meaning
 344 m is not inherently specified by that meaning. Rather, it reflects the gesturer's system for associating meanings and ordering patterns. To interpret the utterances of another
 346 speaker, we must make some assumption about the speaker's system for *producing* utterances. This principle has been central to models of pragmatic language processing
 348 (Goodman & Frank, 2016), and is just as important in situations like ours where no existing linguistic conventions are established.

350 3.2 Probabilistic Conditional Word-order Usage

Let $\vec{p} = (p_{ext}, p_{int})$ be a simple probabilistic model describing preferential usage, conditional on semantic properties of the verb, of the two possible orderings for subject-first
 352 gestures composed of a single verb and object (SVO and SOV)¹⁰. Here p_{ext} is the probability of employing SVO to express an *Extensional* event: $p(g = \text{SVO}|m = \text{Ext}) = p_{ext}$;
 354 likewise p_{int} is the probability of using SVO to express an *Intensional* event: $p(g =$

⁹The constant $p(g) = p(g|m = \text{Ext})p(m = \text{Ext}) + p(g|m = \text{Int})p(m = \text{Int})$ captures the degree of evidence conveyed by the gesture, summed over both possible hypothesised event types.

¹⁰For brevity, we will call these *S-First* gestures. The model describes the computations that underpin usage of just these two orderings: though more are possible, we are interested primarily in the balance of SVO and SOV, and as such we ignore alternatives, though note that the model could easily be extended to reserve probability mass for alternative orderings. This is a reasonable simplification since our experiment concerned just SOV and SVO.

356 SVO $|m = \text{Int}$). The probabilities of employing SOV for Extensional and Intensional
 events respectively are $p(g = \text{SOV}|m = \text{Ext}) = 1 - p_{ext}$ and $p(g = \text{SOV}|m = \text{Int}) =$
 358 $1 - p_{int}$.

An underlying system of associations \vec{p} is tacitly assumed in equation 1, since
 360 $p(g|m)$ is a function of \vec{p} . In our experiment, which featured no labelled examples
 or feedback, participants faced an inherent uncertainty about the gesturer’s system \vec{p} .
 362 For example, the gesturer could be speaking a language that does not condition the
 ordering of verbs and their objects on this semantic distinction, consistently expressing
 364 both extensional and intensional events using SVO (i.e. $p_{ext} = p_{int} \approx 1$) or SOV
 (i.e. $p_{ext} = p_{int} \approx 0$). Likewise, these ordering patterns could be in free variation
 366 ($p_{ext} = p_{int} = 1/2$), strong complementary conditioned usage (i.e. $p_{ext} = 1, p_{int} = 0$
 or $p_{ext} = 0, p_{int} = 1$), weaker complementary usage (i.e. $p_{ext} = 1 - p_{int}$), or anything
 368 in between. We aim to compute a probability model for the decisions of a learner who
 accounts for this uncertainty.

370 3.2.1 Accounting for Uncertainty about \vec{p}

One simple way to achieve this computationally is to model a learner who considers
 all possible systems \vec{p} , accounting for the implications each variant entails for her de-
 cision¹¹. Crucially, such a learner need not treat all \vec{p} s as equally plausible. We allow
 the computation to reflect a *weighted* sum, taken over a prior distribution $p(\vec{p})$ which
 specifies the learner’s biases over the space of possible systems. This is how we model
 the influence of inductive biases on inference whilst also accommodating the uncer-
 tainty inherent in the learner’s observations. Under these assumptions the probability

¹¹Technically, we assume the learner considers all systems \vec{p} that could have generated the observed ges-
 ture. An infinitesimally small subset of possible systems \vec{p} represent a mis-specified model for the gesturer
 under certain observations (observed gesture orderings). For example, if the learner observes an SVO gesture,
 the system $\vec{p} = (0, 0)$ is a mis-specified model of the world, since neither event type could have generated the
 data under this model: as a result, the posterior distribution $p(m|g, \vec{p})$ over event types is improper, being zero
 for both types of meaning, leading to $p(g) = 0$. So in equation (2), a misspecified model of the gesturer will
 make no contribution to the sum, even if it reserves probability mass under the prior $p(\vec{p})$, since $p(m|g, \vec{p})$
 will evaluate to zero whatever the meaning. We thank Simon Kirby for raising this point.

of choosing meaning m after observing gesture g is:

$$p(m|g) = \iint_{\vec{p}} p(m|g, \vec{p}) p(\vec{p}) \, dp_{ext} \, dp_{int} \quad (2)$$

Here $p(m|g, \vec{p})$ is given by equation (1), with the conditioning on \vec{p} made explicit.
 372 The quantity $p(\vec{p})$ can be understood to reflect the learner’s prior beliefs: cognitive
 biases for conditional association of ordering patterns and semantic properties of the
 374 verb. These biases impose probabilistic preferences on the space of possible associ-
 ation systems, and can be modelled with the Beta distribution (see Appendix A for
 376 details): $p(\vec{p}) = p(p_{ext})p(p_{int}) = \text{Beta}(p_{ext}; \alpha_{ext}, \beta_{ext}) \cdot \text{Beta}(p_{int}; \alpha_{int}, \beta_{int})$. The shape
 and strength of these preferences are determined by the prior parameters $\alpha_{ext}, \beta_{ext}, \alpha_{int},$
 378 and β_{int} . An intuitive way to view these parameters is as *pseudo-counts*, or counts that
 are added to the observed counts when predicting the probability of an outcome (α
 380 being the pseudo-count for SVO gestures, and β for SOV).

3.3 Results

3.3.1 Model Predictions

We analysed three versions of the model and compared their predictions to the exper-
 384 imental data (figure 3). A baseline *unbiased* version of the model (M0), in which we
 fixed neutral priors over meanings ($\lambda = 1/2$) and event-ordering association systems
 386 ($\alpha_{ext} = \beta_{ext} = \alpha_{int} = \beta_{int} = 1$), is unsurprisingly the poorest predictor of the exper-
 imental data, affording participants’ responses a combined log-likelihood of -256.98:
 388 this model predicts fifty-fifty interpretation responses to both SOV and SVO gestures,
 failing to capture the asymmetry in responses across ordering patterns, and the overall
 390 preference for Extensional events.

In order to ask whether the experimental result is being driven by general pref-
 392 erences for one event type over another, and not by conditional associations between
 events and ordering patterns, we computed the predictions of a semi-biased version of
 the model (M1): here we fixed a neutral prior over association systems ($\alpha_{ext} = \beta_{ext} =$
 394 $\alpha_{int} = \beta_{int} = 1$) but fit λ to the experimental data. The maximum-likelihood estimate

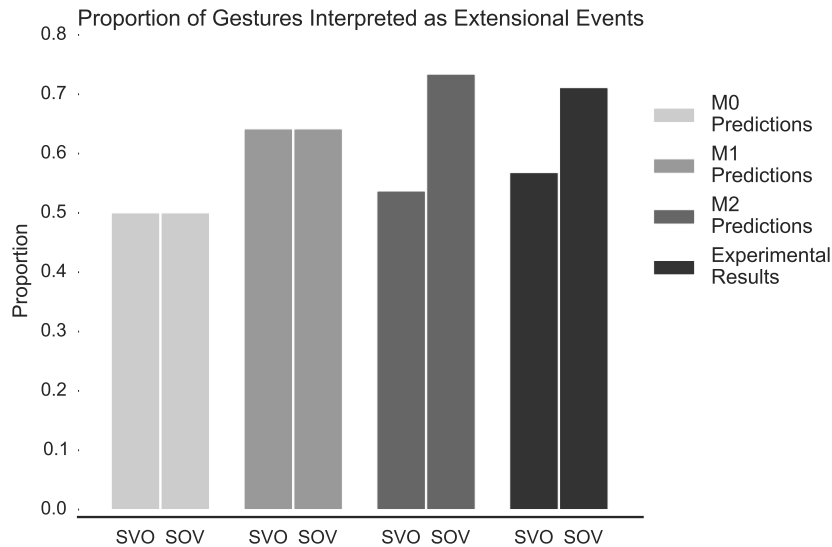


Figure 3: Comparison of model predictions and experimental results. The biased model (M2), but not the unbiased (M0) or the event-biased (M1) model, predicts both experimental results: asymmetric responses to SOV and SVO gestures, and an overall preference for Extensional events. Model predictions show the predicted probability of interpreting SVO/SOV gestures as Extensional events $p(m = \text{Ext} | g = \text{SVO}, \alpha, \beta)$ and $p(m = \text{Ext} | g = \text{SOV}, \alpha, \beta)$.

396 is $\hat{\lambda} = 0.68$: this model affords the data a combined log-likelihood of -225.00. The
 model-fit suggests a slight overall preference for Extensional events independent of
 398 gesture ordering, and this is reflected in the model’s predictions (figure 3). This bias is
 in line with the results of our baseline experiment presented above, in which an overall
 400 proportion of 0.68 of actions were interpreted as extensional events. We found of λ
 to be one of the most consistent parameter estimates in our model. Though M1 is a
 402 better fit to the data than M0, it nevertheless fails to capture the assymetry between
 responses to SVO and SOV gestures.

404 These versions of the model suggest that – to account for the pattern of exper-
 imental results – the learning model we have described must include a non-neutral
 406 preference for some systems of event-ordering association over others. We fit the full
 model (M2) to the experimental data, by inferring maximum-likelihood estimates (see
 408 Appendix A for details) for $p(\vec{p})$. Drawing inferences about the shape of this prior
 is challenging: but possible under the relatively weak assumption that participants’

410 constituent ordering preferences are approximately complementary across event-types:
412 *however strongly I prefer one ordering pattern for Extensional events, that’s how*
414 *strongly I prefer the alternative ordering pattern for Intensional events.* More formally, we limit the space of possibilities for $p(\vec{p})$ by assuming $p(p_{ext}) \sim \text{Beta}(\alpha, \beta)$ and $p(p_{int}) \sim \text{Beta}(\beta, \alpha)$, thereby reducing the parameter space to two dimensions rather than four.

416 This assumption may seem restrictive, but is justified by both theoretical and practical considerations. In practical terms, the space of possible priors defined by allowing
418 four freely varying parameters is too broad to make reliable inferences about their values given the model we defined and the available data: many possible priors lead to
420 equivalent or near equivalent values for $p(m|g)$, so the experimental data cannot choose between alternative priors reliably¹². A natural solution is to reduce the number of
422 model parameters to create a space of possible priors in which we can perform reliable inference. Moreover, this reduction can even be a desirable restriction if there are theoretical reasons to focus on a particular subspace of priors, and it is possible to check
424 that the reduction does not also lead to a dramatic reduction in the likelihood of the data (compared to the higher-dimensional model). In our case, both of these conditions
426 are met (more details below).

428 Fixing $p(m)$ at its maximum-likelihood value inferred from M2 ($\lambda = 0.68$), the maximum-likelihood parameter estimates for $p(\vec{p})$ are $\hat{\alpha} = 0.9$ and $\hat{\beta} = 1.18$, affording
430 the data a combined log-likelihood of -215.98, correctly predicting participants’ chosen interpretation with average probability 0.77. Figure 3 demonstrates the close
432 correspondence between the model’s predictions and participants’ responses in our experiment.

434 A natural concern is that we are building in the assumption that SVO and SOV are used to communicate the semantic distinction in a somewhat complementary way, by
436 assuming $p(p_{ext}) \sim \text{Beta}(\alpha, \beta)$ and $p(p_{int}) \sim \text{Beta}(\beta, \alpha)$. In addition to the practical issues raised above, there are a number of theoretical reasons that this should not be
438 a major concern. First, whilst we aren’t able to identify a single best-fitting prior in

¹²This is a common obstacle in model fitting, and is often referred to in technical terms as *weak identifiability*. Appendix D includes MCMC samples from the posterior distribution over these parameters.

the four-dimensional case, we are able to identify the maximum of the data likelihood
440 function in this model (achievable under multiple "best-fitting" priors). Crucially, this
maximum value is identical to the maximum value achievable under the two-parameter
442 "complementary priors" model (-215.98). In other words, the assumption of comple-
mentary biases does not reduce the likelihood of the data, suggesting that we should
444 prefer the two-parameter version on grounds of parsimony anyway.

Second, we also analysed alternative assumptions within the restriction that only
446 two parameters define the prior, and found these to be inferior. For example, rather
than assuming "complementary" priors across event types ($\alpha_{ext} = \beta_{int}, \alpha_{int} = \beta_{ext}$),
448 we could assume independent priors which are each defined by a single parameter,
such that $p(p_{ext}) \sim \text{Beta}(\alpha, \alpha)$ and $p(p_{int}) \sim \text{Beta}(\beta, \beta)$, or identical priors defined
450 by two parameters, so that $p(p_{ext}) \sim \text{Beta}(\alpha, \beta)$ and $p(p_{int}) \sim \text{Beta}(\alpha, \beta)$. Neither of
these assumptions can explain the data as well as the "complementary prior" assump-
452 tion: respectively, the maximum of the likelihood function in these models is -216.92
and -225.03. Maximum likelihood analysis favours the complementary priors model,
454 although the independent single-parameter model achieves relatively comparable log-
likelihood, and may therefore also be worthy of further investigation as an alterna-
456 tive description of participants' biases. Taken together, these analyses suggest that the
"complimentary priors" assumption is justified over alternatives, both practically and
458 theoretically, so we will proceed to focus on this case.

3.3.2 Inferred Priors

460 Figure 4 shows the inferred prior $p(\vec{p})$. First, the model suggests a clear asymmetry in
ordering preferences across event types: the prior favours SOV for Extensional events,
462 and SVO for Intensional events. Second, the prior demonstrates a bias toward *regular-*
ity: consistent usage of the favoured ordering is preferred over variable usage (prob-
464 ability density peaks close to 0 for Extensional events and 1 for Intensional events).
This aspect of the prior is in keeping with the *regularisation bias*: a general prefer-
466 ence for regularity – motivated by simplicity principles and thought to be relevant to
cognition in general – that has been proposed in various linguistic (Real & Griffiths,

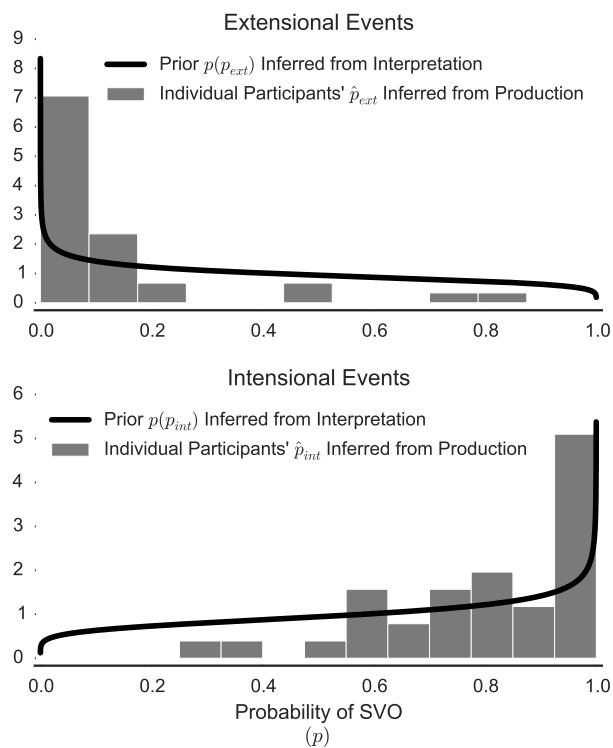


Figure 4: Lines show probability density functions for priors $p(p_{ext})$ (top) and $p(p_{int})$ (bottom) inferred from production data, superimposed on the (normalised) histograms of estimates of individual participants' \hat{p}_{ext} (top) and \hat{p}_{int} inferred from Schouwstra and de Swart's (2014) production data.

468 2009a; Smith & Wonnacott, 2010; Culbertson & Smolensky, 2012) and non-linguistic
 (Ferdinand, Thompson, Smith, & Kirby, 2013) domains.¹³ Third, the prior expresses
 470 preferences that are skewed but *weak*; it encodes asymmetric ordering preferences, but
 these defeasible preferences could be easily overturned by observing contradictory data
 472 about \vec{p} . A common measure for the strength of preferences imposed by prior beliefs
 modelled using the Beta distribution is the *effective sample size* (ESS): $s = \alpha + \beta$. If,
 474 as is common, the prior is viewed as expressing a set of *imaginary* data-points, then
 the ESS reflects their number, and thus their power to over-rule observed data-points.
 476 In the inferred prior, $s = 2.08$, suggesting just a handful of contradictory data-points
 could lead the learner to entertain disfavoured systems \vec{p} .

¹³Note that the kind of regularity observed here is *conditioned* regularity. Had there not been an asymmetry in ordering preference (the first aspect of the prior discussed above), then regularisation would have pushed the system towards one word order.

478 A simple way to test the credibility of the model is to ask how well its predictions
generalise to *production*, having been inferred from *interpretation* only. Our reason-
480 ing about the differences between production and interpretation in this context predicts
that the shape of the inferred prior should be broadly compatible with the distribu-
482 tion of productions across participants, favouring most strongly the kinds of systems
evidenced in production, but should also reserve some probability mass over a wider
484 range of possible systems \vec{p} than those which were most prominent in production. This
is what we find. Together with the results presented in figure 3, these results show
486 that our model is consistent with the differences we are attempting to explain between
production and interpretation experimental results.

488 We analysed production data from Schouwstra and de Swart (2014)'s experiment
and inferred maximum-likelihood estimates \hat{p}_{ext} and \hat{p}_{int} for each individual partici-
490 pant (see Appendix C for details). Figure 4 shows the (normalised) histograms of these
estimates, superimposed on the priors $p(p_{ext})$ and $p(p_{int})$ we inferred from interpreta-
492 tion. A correspondence between the distributions is clear: the biases we inferred from
interpretation are consistent with the pattern of results observed independently in pro-
494 duction. The prior favours the same strongly biased ordering systems that were most
prominent in production, but also reserves some non-zero probability for alternative
496 ordering systems that were not prominent in production. This is consistent with our
hypothesis that the same biases play a role in production and interpretation, but that
498 low probability ordering systems are accounted for during interpretation, which dilutes
the stronger asymmetry observed in production that was driven by favoured, higher
500 probability ordering systems.

It is also possible to directly compute the likelihood of the production data under
502 the prior inferred from participants' interpretations (see appendix C or details): this
analysis shows that the model correctly predicts the use of SVO and SOV in production
504 with average probability 0.77 for extensional gestures and 0.74 for intensional gestures.

3.4 Discussion

506 Our model provides one possible computational account for the main experimental
finding that when interpreting gestures, participants used constituent ordering patterns
508 as a cue to meaning. The model we described is a first-approximation to the inferences
that underpin production and interpretation of improvised communication. However,
510 the basic proposal – that interpretation involves inference and estimation, and that the
Bayesian framework provides a natural and useful model for understanding how learn-
512 ers bring their biases to bear on this uncertainty – is not tied to these experimental
conditions or this particular model. For example, the inferential model makes spe-
514 cific predictions about the posterior beliefs participants should entertain after observ-
ing labeled training examples, and it would be straightforward to construct experimen-
516 tal procedures that test these predictions. Likewise, plausible alternative explanations
for asymmetry between production and interpretation could be formulated within this
518 framework and directly compared.

Our model assumes that during interpretation, uncertain observers have principled
520 motivation to fall back on a more abstract layer of knowledge – a *prior* over possible
ordering systems – which can in theory dilute the lower-level ordering biases evident
522 in participants’ responses in a matched production experiment. We have suggested that
this principle offers an explanation for the difference in effect we observe between pro-
524 duction and interpretation experiments. This explanation rests on the hypothesis that
production in this particular scenario does not invoke the same abstract considerations
526 (at least not to the same degree), but follows a lower level sampling process driven by
favoured ordering systems. These favoured ordering schemes will have high probabilit-
528 ity under the prior thanks to the same semantic biases, but we shouldn’t expect that
these biases are so strong as to rule out consideration of alternative ordering schemes.
530 While the asymmetry may not hold for more interactive production scenarios in gen-
eral, we believe this assumption is a conservative starting point which could easily be
532 tested in future experiments that manipulate the degree of interactivity in production.
An emerging body of research on the pragmatics of speech production and interpreta-
534 tion (Goodman & Frank, 2016) provides a road map for these kinds of questions.

In general, we hope our analysis can motivate further experimental and computational efforts to illuminate how individuals use and process improvised communication systems under uncertainty. Computational modelling will be a crucial component in understanding how production and interpretation interact during communication and learning to shape the dynamics of an emerging language, particularly as those forces play out in populations. Having experimentally-informed computational accounts of these processes is an important step in that direction.

4 General Discussion

In the silent gesture paradigm, people are forced to communicate while they cannot rely on an existing language system: they have to improvise. Previous work has shown that when people improvise, there are some general principles for the organisation of their utterances: they prefer SOV word order for simple transitive events that involve motion through space, but they switch to other orders for other kinds of events. This kind of meaning based word order alternation is not generally observed in fully conventionalised languages.

In this paper we have looked at the interpretation of silent gesture, and compared it to silent gesture production. We used a laboratory experiment as well as a computational model to investigate the mechanisms that underpin the emergence of linguistic rules, particularly how language production and language comprehension relate to each other. We started from the observation (Schouwstra & de Swart, 2014) that when people improvise, the organisation of their utterances depends on their semantic properties: extensional and intensional events give rise to SOV and SVO word orders respectively. Using a silent gesture interpretation experiment, we showed that a similar connection between meaning and form is present in the interpretation of improvised gesture. However, the effect in interpretation appeared modest in comparison to production. In the second part of the paper we proposed an explanation for this: when people interpret improvised gesture, they face an inherent uncertainty about the gesturer's linguistic system. An ideal learner would account for this uncertainty, and shape her interpreta-

tion decisions accordingly.

564 In the introduction section we saw that previous interpretation experiments have led
to differing conclusions. Either, interpretation of improvised gesture, like production,
566 by-passes the grammatical system, and prefers SOV order for simple transitive (exten-
sional) events (Langus & Nespors, 2010), or production and interpretation each call for
568 rather different explanations: a simple semantics based heuristic ('agent first') for in-
terpretation, and specific production-related constraints (i.e., role conflict for reversible
570 events) for production (Hall et al., 2015). In this paper we used the combination of an
experiment and Bayesian modelling to obtain a more detailed picture of the improvi-
572 sation situation. With our experiment we showed that in silent gesture interpretation
(like in its production), meaning type and structure are connected in a way that is not
574 generally observed in existing languages. The heuristic that we have focused on in
this paper (the one that connects SOV to extensional and SVO to intensional events) is
576 different from the one that is described by (Hall et al., 2015), but they are both clearly
semantics based, and certainly compatible with one another.¹⁴

578 At the same time, there are important differences between silent gesture production
and interpretation. While in both production and interpretation experiments, partici-
580 pants must improvise and cannot use their own language or any conventional language
they know, the production experiment is more clearly than the interpretation experi-
582 ment a situation that lacks linguistic conventions. Participants produce their gestures
to the camera, and although there is an experimenter present, this experimenter is not
584 engaged in the improvisation task. In the interpretation experiment, participants are not
alone in the improvisation act: they observe another person's linguistic behaviours, and
586 they may entertain the possibility that this person behaves according to a set of existing
or emerging rules or tendencies in production. We constructed a computational model
588 of interpretation through inductive inference, based around the principle that learners
account for this uncertainty surrounding another individual's language use. The model
590 suggests that participants' decisions – which varied by word-order but nevertheless

¹⁴In fact, there is no reason to assume that the 'agent first' heuristic does not play a role in the *production* data discussed in (Hall et al., 2015), because the data does not provide counter evidence against this principle. The principle alone is simply not enough to explain the word order patterns.

portrayed uncertainty – may reflect motivated uncertainty in response to this unknown.
592 Casting interpretation in this framework (Perfors, Tenenbaum, Griffiths, & Xu, 2011)
connects improvised gesture with inference and estimation in other domains (e.g. Hsu,
594 Chater, & Vitányi, 2011; Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Culbertson,
Smolensky, & Legendre, 2012; Goldwater, Griffiths, & Johnson, 2009; Perfors, Tenen-
596 baum, & Regier, 2011) through domain-general principles of inductive inference. For-
mulating questions about the emergence of linguistic rules/conventions in probabilistic
598 models of cognition is fruitful because it explicitly addresses how learners represent
and reason about the uncertainty that surrounds other minds in the absence of helpful
600 evidence. Going forward, we aim to explore principles for inductive inference fur-
ther, in semi-supervised learning scenarios that systematically confirm or contradict
602 the biases we inferred: e.g. *how much evidence must learners observe before they
are confident in their estimate of another individual's linguistic system? Can seem-
604 ingly disfavoured ordering patterns be easily learned?* The computational principles
governing when and how people bring their biases to bear on language use under un-
606 certainty are at present only superficially understood. Improvised gesture offers a rich
testing ground for these questions, which we believe will be best understood through
608 synthesis of experimental and computational analysis.

In the context of this paper, production and comprehension are studied separately,
610 but in real life, production and interpretation are not separated as strictly. In natural
interactive situations, they are always combined, and often even done at the same time
612 (Pickering & Garrod, 2013). A logical next step is to extend the silent gesture paradigm
to include communication (Christensen, Fusaroli, & Tylén, 2016) and cultural trans-
614 mission through artificial generations of lab participants (Motamedi, Schouwstra, Smith,
Culbertson, & Kirby, 2018; Schouwstra, Smith, & Kirby, 2016).

616 Together, the experiment and the model presented here clarified our thinking about
the mechanisms at play when a new language system emerges. The experiment showed
618 that the interpretation of silent gesture favours ordering preferences that are condi-
tioned on meaning - similar to what was observed for silent gesture production. Fitting
620 a computational model to the experimental data allowed us to estimate these ordering

preferences: they appear to be skewed and asymmetric across event types, but weak.
622 This implies that they lead to stronger conditioning of order on meaning when there are
no linguistic observations, a pattern that is confirmed by the silent gesture production
624 results from Schouwstra and de Swart (2014). On the other hand, the conditioned word
order alternation may be easily overturned by contradictory linguistic observations.
626 This observation appears consistent with the fact that there are no languages in which
word order is conditioned on verb type as it appears to be in silent gesture production.
628 However, it is well known that, under the right circumstances, weak inductive biases
can shape regularities – sometimes even disproportionately strong regularities – over
630 the course of cultural transmission (Kirby, Dowman, & Griffiths, 2007; Smith & Won-
nacott, 2010; Griffiths & Kalish, 2007; Boyd & Richerson, 1985; Real & Griffiths,
632 2009b). Understanding the cultural evolutionary forces that suppress this alternation in
natural language is therefore a key priority for future research.

634 **5 Appendix A: Experimental stimuli and results by item**

The following strings were used in the experiment (one ambiguous action per pair of
636 verbs). The experimental results plotted by item can be found in figure 5. Item numbers
in the table correspond with those in the figure, and the baseline values correspond to
638 the proportion of extensional interpretations chosen in the verb-only experiment (see
section 2.3). All experiment items, as well as the raw data and code for analysis are
640 available on <https://osf.io/tfqcp/>.

item	description	baseline value
1	Princess smashes / carves vase	0.73
2	Gnome cuts / draws pizza	0.65
3	Witch eats / wants banana	0.45
4	Witch decorates / paints table	0.63
642 5	Girl sleeps on / dreams of book	0.30
6	Princess talks to / talks about teddybear	0.90
7	Pirate throws / hears guitar	0.85
8	Cook stirs / smells	0.90
9	Gnome pats / feels book	0.45
10	Witch climbs / builds house	0.95

6 Appendix B: ML Estimation of Interpretation Model

644 Parameters

The model formulates each chosen interpretation as an independent Bernoulli trial over Intensional and Extensional interpretations. The likelihood of a given participant’s set of decisions is the product of two Binomial likelihoods, one for interpretations of SVO gestures and another for SOV gestures. The combined log-likelihood of the entire experimental data set D , taken over all n participants, as a function of model parameters $\Theta = (\alpha, \beta, \lambda)$ is:

$$\mathcal{LL}(D|\Theta) = \sum_{i=1}^n \ln [\text{Binomial}(k_i^{SVO}; \theta^{SVO}, N^{SVO})] + \ln [\text{Binomial}(k_i^{SOV}; \theta^{SOV}, N^{SOV})] \quad (\text{B.1})$$

where $\theta^{SVO} = p(m = \text{Ext.} | g = SVO)$ and $\theta^{SOV} = p(m = \text{Ext.} | g = SOV)$ are computed with equation (2), which is given in explicit form below. k_i^{SVO} and k_i^{SOV} give the number of SVO and SOV gestures interpreted as Extensional events by the i th participant respectively, while N^{SVO} and N^{SOV} give the total number of SVO and SOV gestures observed, which did not vary across participants. For the main two-parameter *complementary priors* version of the model, equation (2) can be written more explicitly than

the version in the main text. Separating the two gesture orderings:

$$\theta^{svo} = \int_0^1 \int_0^1 \frac{\lambda p_{ext}}{\lambda p_{ext} + (1 - \lambda)p_{int}} \frac{b}{\mathbf{B}(\alpha, \beta)^2} dp_{ext} dp_{int} \quad (\text{B.2})$$

$$\theta^{sov} = \int_0^1 \int_0^1 \frac{\lambda(1 - p_{ext})}{\lambda(1 - p_{ext}) + (1 - \lambda)(1 - p_{int})} \frac{b}{\mathbf{B}(\alpha, \beta)^2} dp_{ext} dp_{int} \quad (\text{B.3})$$

$$b = [(1 - p_{int}) \cdot p_{ext}]^{\alpha-1} \cdot [p_{int}(1 - p_{ext})]^{\beta-1} \quad (\text{B.4})$$

The first term in equations (B.2) and (B.3) gives $p(m = \text{Extensional} | g, \vec{p}, \lambda)$. The
646 second term in both gives the prior over ordering systems $p(\vec{p})$, which is a combination
of two Beta densities (the combination can be written this way thanks to the symmetry
648 in the parameters and the identity $\mathbf{B}(\alpha, \beta) = \mathbf{B}(\beta, \alpha)$ in the Beta function). Maximum
likelihood estimates were obtained through numerical minimisation of (the inverse of)
650 eq. (B.1). Predictive probabilities reported throughout refer to the geometric mean
of the combined log likelihood of all decisions. All optimisation procedures reported
652 were carried out using the Python library *Scipy*. Figure 6 shows the two-parameter
version of the model in graphical form.

654 7 Appendix C: ML Estimation of \vec{p} from Production

Maximum likelihood estimates for p_{ext} and p_{int} inferred from production data for the
 i th participant are:

$$\hat{p}_{ext} = k_i^{ext} / N_i^{ext} \quad (\text{C.1})$$

$$\hat{p}_{int} = k_i^{int} / N_i^{int}, \quad (\text{C.2})$$

where k_i^{ext} and k_i^{int} give the number of Extensional and Intensional events expressed
656 using SVO, and N_i^{ext} and N_i^{int} are the total number of S-First gestures the i th participant
produced for Extensional and Intensional events, respectively. When we report the
658 probability of the production data under the prior inferred from interpretation, we are

computing the marginal likelihood of the binomial data under the beta prior determined
660 by the inferred parameters – the beta-binomial compound distribution.

8 Appendix D: Independent Beta Priors Analysis

662 Figure 7 shows 500000 MCMC samples from the marginal posterior distributions for
 $\log(\alpha_{ext})$, $\log(\beta_{ext})$, $\log(\alpha_{int})$, and $\log(\beta_{int})$ in the "independent Beta priors" version of
664 the model which allows four free parameters (with $\lambda = .68$ fixed). Samples were col-
lected under a uniform prior using an ensemble sampler with 250 walkers (Foreman-
666 Mackey, Hogg, Lang, & Goodman, 2013) initialised uniformly at random in $[-10, 10]$.
We collected so many samples because the data likelihood surface in this model is er-
668 ratic: the model parameters are only weakly identifiable given the experimental data.
We omit pairwise correlation plots because they are largely uninformative. Note how-
670 ever the correspondence between the distributions for α_{ext} and β_{int} , which is broadly
consistent with the idea that $p(p_{ext})$ and $p(p_{int})$ encode somewhat complimentary pref-
672 erences, though we caution against overinterpreting parameter estimates in this version
of model.

674 Acknowledgements

The authors would like to thank Marianne Smit, Nicky Mariën, Çağlar Yalıt, Anouschka
676 van Leeuwen and Martijn van der Klis for their help with the experiment, and Simon
Kirby for his comments on a previous version of the model. MS was funded by the
678 British Academy, grant number pf130034; BT was funded by the ERC, grant number
283435.

680 References

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H.,
682 Dai, B., & Grothendieck, G. (2015). Package 'lme4'. *Convergence*, 12(1).

- Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.
- Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, 118(2), 172–189.
- Burling, R. (2000). Comprehension, production and conventionalisation in the origins of language. *The evolutionary emergence of language: Social function and the origins of linguistic form*, 27–39.
- Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10(7), 335–44.
- Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 811–823.
- Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146, 67–80.
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–9.
- Culbertson, J., & Smolensky, P. (2012). A Bayesian model of biases in artificial language learning: the case of a word-order universal. *Cognitive science*, 36(8), 1468–98.
- Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122(3), 306–29.
- Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, 37(7), 1356–1367.
- Ferdinand, V., Thompson, B., Smith, K., & Kirby, S. (2013). Regularization behavior in a non-linguistic domain. In M. Kanuff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science society*. Berlin: Cognitive Science Society.
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee :

- The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*,
714 *125*(925), 306–312.
- Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language
716 Games. *Science*, *336*(6084), 998–998.
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013).
718 A noisy-channel account of crosslinguistic word-order variation. *Psychological
science*, 0956797612463705.
- 720 Givon, T. (1997). On the co-evolution of language, mind and brain. *Evolution of
communication*, *2*(1), 45–116.
- 722 Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order
of events: How speakers of different languages represent events nonverbally.
724 *PNAS*, *105*(27), 9163–9168.
- Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word
726 segmentation: exploring the effects of context. *Cognition*, *112*(1), 21–54.
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Prob-
728 abilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.
- Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling
730 Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*(1),
173–184.
- 732 Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Prob-
abilistic models of cognition: exploring representations and inductive biases.
734 *Trends in cognitive sciences*, *14*(8), 357–364.
- Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with
736 bayesian agents. *Cognitive science*, *31*(3), 441–80.
- Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and com-
738 prehension show divergent constituent order preferences: Evidence from elicited
pantomime. *Journal of memory and language*, *81*, 16–33.
- 740 Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2014). Investigating constituent or-
der change with elicited pantomime: A functional account of svo emergence.
742 *Cognitive science*, *38*(5), 943–972.

- 744 Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17.
- 746 Hsu, A. S., Chater, N., & Vitányi, P. M. B. (2011). The probabilistic analysis of language acquisition: theoretical, computational, and experimental analysis. *Cognition*, *120*(3), 380–90.
- 748 Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(12), 5241–5.
- 750 Kline, M., Salinas, M. A., Lim, E., Fedorenko, E., & Gibson, E. (2017). Preprint-word order patterns in gesture are sensitive to modality-specific production constraints.
- 752 Kocab, A., Lam, H., & Snedeker, J. (2018). When cars hit trucks and girls hug boys: The effect of animacy on word order in gestural language creation. *Cognitive science*, *42*(3), 918–938.
- 754 Langus, A., & Nespors, M. (2010). Cognitive systems struggling for word order. *Cognitive psychology*, *60*(4), 291–318.
- 756 MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.
- 760 Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lopic, R., Ben-Basat, A. L., Padden, C., & Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, *158*, 189–207.
- 762 Meir, I., Lifshitz, A., Ilkbasaran, D., & Padden, C. (2010). The interaction of animacy and word order in human languages: A study of strategies in a novel communication task. In *Proceedings of the eighth Evolution of Language Conference* (pp. 455–456).
- 764 Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2018, Apr). *The emergence of spatial modulation in artificial sign languages*. PsyArXiv.com/p6zy4.
- 766 Mudd, K., Kirby, S., & Schouwstra, M. (2018). Improvised word order biases are
- 772

- not modality specific: evidence from non-linguistic vocalizations. In H. Little & A. Micklos (Eds.), *The proceedings of the Evolang XII Modality Matters workshop*. Online at http://hlittle.com/MM/allpaperpdfs/EvoLangMM_paper6.pdf.
- Newmeyer, F. J. (2000). On the reconstruction of ‘proto-world’ word order. *The evolutionary emergence of language: Social function and the origins of linguistic form*, 372–388.
- Padden, C., Meir, I., Sandler, W., & Aronoff, M. (2010). Against all expectations: Encoding subjects and objects in a new language. *Hypothesis A/hypothesis B: Linguistic explorations in honor of David M. Perlmutter*, 383–400.
- Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, 120(3), 302–21.
- Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118(3), 306–38.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329–347.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria.
- Reali, F., & Griffiths, T. L. (2009a). The evolution of frequency distributions: relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–28.
- Reali, F., & Griffiths, T. L. (2009b). The evolution of frequency distributions: relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–28.
- Schouwstra, M. (2017). Temporal structure in emerging language: From natural data to silent gesture. *Cognitive Science*, 41(S4), 928-940.
- Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131(3), 431–6.
- Schouwstra, M., Smith, K., & Kirby, S. (2016). From natural order to convention in silent gesture. In *The Evolution of Language: Proceedings of the 11th Interna-*

tional Conference (Evolang XI).

- ⁸⁰⁴ Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–9.

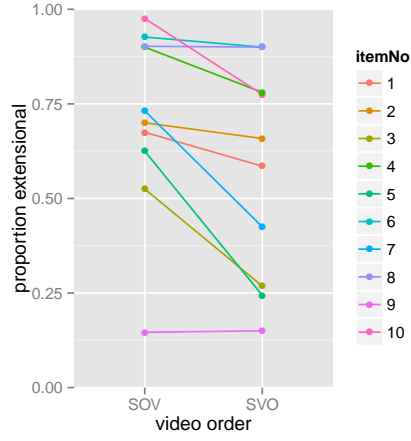


Figure 5: This shows the experimental results per item. The y axis shows the proportion of extensional interpretations.

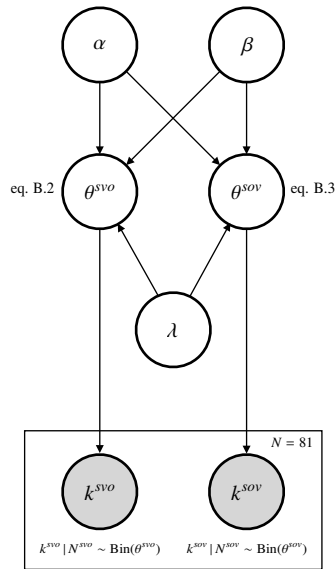


Figure 6: The model in graphical form, assuming complementary priors $p_{ext} \sim \text{Beta}(\alpha, \beta)$ and $p_{int} \sim \text{Beta}(\beta, \alpha)$.

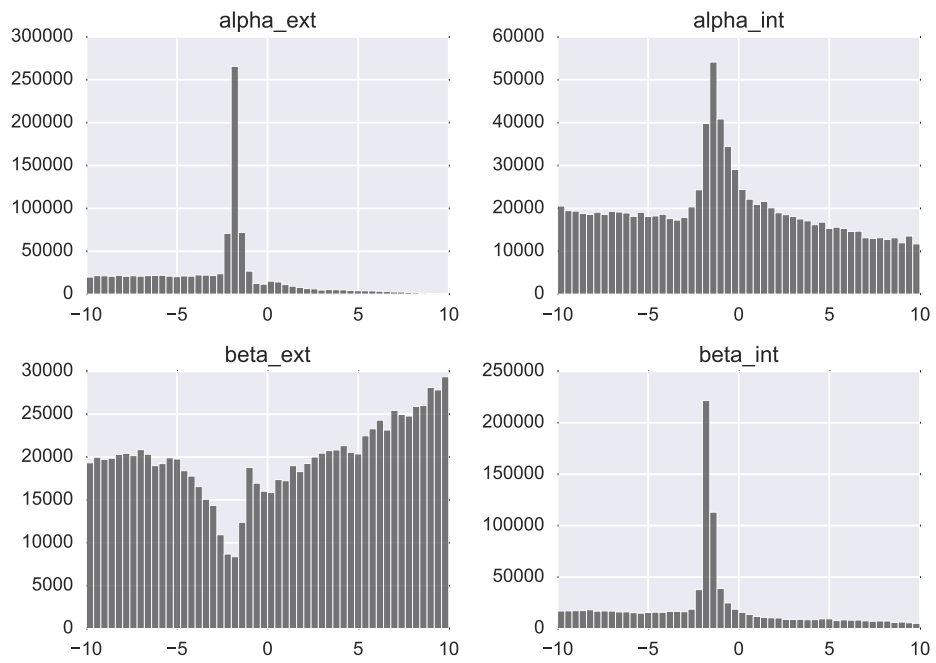


Figure 7: Posterior marginal samples in the *independent priors* version of the model, which allows four free parameters to determine prior distributions with independent shapes.