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

Marieke Schouwstra*
University of Edinburgh
Marieke.Schouwstra@ed.ac.uk

Henriëtte de Swart
Utrecht University
h.deswart@uu.nl

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

*****Manuscript submitted to Cognitive Science*****
*****please do not cite without permission*****

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.

1 Introduction: production and interpretation of emerging language

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

*corresponding author

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). The silent gesture paradigm has been used to investigate several core features of language, such as how a communication system can be bootstrapped through iconicity (Fay, Arbib, & Garrod, 2013). In particular, the paradigm has provided notable insight into the origins of the ordering of Subject, Object and Verb in human language. 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 of their native language (Goldin-Meadow, So, Özyürek, & Mylander, 2008). Given the dominance of SOV in emerging language systems (e.g., Padden, Meir, Sandler, & 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 recent publications show that under certain circumstances SOV is not the dominant order (Meir, Lifshitz, İlkbasaran, & Padden, 2010; Langus & Nespor, 2010; Gibson et al., 2013; Hall, Mayberry, & Ferreira, 2013; Schouwstra & de Swart, 2014). Investigation of this variability in word order has sparked a debate about the 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 emergence of language. The silent gesture paradigm is relatively new, and many questions are still unanswered. However, two semantic distinctions, that between reversible 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 influence word order in emerging language.

Whether or not an event is reversible (typically events in which there are two animates, 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 SVO usage becomes more likely (although part of this effect is due to the native language of the participants; in SOV-dominant languages, participants only move away from SOV in embedded structures like the italic bit in 'woman says *fireman* *kicks girl*').[1] Various explanations for the phenomenon have been offered, but there is no conclusive evidence for whether the pattern is rooted in communicative or cognitive principles (or even potentially the result of modality-specific processes; (Gibson et al., 2013; Hall et al., 2013; Kline, Salinas, Lim, Fedorenko, & Gibson, 2017).

Another semantic effect on word order variation was observed by Schouwstra and 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 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 ('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 typically 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 important than the extension (object in the world). This makes the direct object more abstract, and possibly non-existent or non-specific.

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 extensional events. They observe that word order flexibility on the basis of such meaning differences in the verb do not exist in fully conventional languages, and argue that 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 properties of events) and build their improvised utterances flexibly, according to these biases. This position contrasts with previous hypotheses in which word order in emerging language systems is seen as something rigid rather than variable (Newmeyer, 2000; Goldin-Meadow et al., 2008). The distinction between intensional and extensional 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

[1] Note 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).

produce non-word sounds to convey information (Mudd, Kirby, & Schouwstra, 2018), a finding that is interesting given potential issues about modality specificity; see above, and Kline et al. (2017).

All in all, the improvised gesture paradigm can reveal pressures that are important 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 laboratory 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 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 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 Bayesian model.

1.1 Emerging language: production vs. interpretation

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 process with two directions: production and interpretation. These two directions may exert different pressures in the emergence of a language system (Burling, 2000; MacDonald, 2013).

The interpretation of strings in the silent gesture paradigm has received little attention, with the exception of two recent studies: one in which participants are asked to recognize the intended meaning of silent gesture strings in a timed forced choice setup (Langus & Nespor, 2010), and one in which participants are asked to choose an interpretation for ambiguous reversible events (Hall, Ahn, Mayberry, & Ferreira, 2015).

Langus and Nespor (2010) asked adult participants to watch video clips of gesture sequences describing simple transitive events through space in a two alternative forced 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 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 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 their native language. Langus and Nespor (2010) claim that this effect is due to the fact that in this task, participants disregard their computational system of grammar.

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 made up of an action and two animate participants, in three possible orders (Action-Participant1-Participant2, Participant1-Action-Participant2, Participant1-Participant2-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 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 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.

They conclude that interpretation of these ambiguous strings is governed by a semantic constraint, 'agent first', and they emphasise the difference between interpretation 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 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).

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 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 entirely clear how production and interpretation, in the absence of linguistic conventions, relate to each other. Hall et al. (2015) emphasise the difference between silent gesture production and interpretation. They postulate procedural, production-related constraints for production, and a semantic heuristic ('agent first') for interpretation.

Langus and Nespor (2010), on the other hand, emphasise the similarities between silent gesture production and interpretation: both are governed, not by grammatical rules, but by cognitive constraints.

We add crucial evidence to the question whether production and interpretation of improvised language are intrinsically similar or rather different from each other. Presenting a silent gesture interpretation experiment, along with a Bayesian computational model for the experimental task, we will point out in which respects production is crucially different from interpretation, in emerging language situations. Our starting point is the semantic differences between extensional and intensional events that are driving word order variability in silent gesture production (Schouwstra & de Swart, 2014). We 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 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 influence the meaning an interpreter derives?'), but also quantitatively: by specifying computational principles that sub-serve interpretation of silent gesture under uncertainty, 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 align with the pattern of results observed independently in production experiments: our results suggest skewed but defeasible event-class-conditional word-order preferences, whose effects on silent-gesture interpretation may be mediated by more general principles of inference under uncertainty.

2 Experiment: improvised gesture interpretation

To test if the order of constituents influences the way in which participants interpret 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 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

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.

results in two possible interpretations: 'witch climbs house' (an extensional event), and 'witch builds house' (an intensional event). We construed videos in two possible orders, SOV and SVO.[2] We hypothesised, based on the production results in Schouwstra and de Swart (2014), plus the similarities between production and interpretation found in Langus and Nespor (2010), that the gesture order had an influence on interpretation, and predicted that, when engaged in a dual forced choice task (that presents 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, and vice versa.

2.1 Method

2.1.1 Participants

Forty one native speakers of Dutch (16 male, 25 female) were recruited from the 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 Istanbul, Turkey. None of the participants received monetary compensation.

[2] Two example videos ('princess sleeps-on / dreams-of book') are included in the supporting material.

2.1.2 Materials

We created video clips showing three gestured elements: an actor, a patient and an 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 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 sequences: 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 material (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 SVO videos, while at the same time, each ambiguous action occurred only once per version.

Four filler items were created: videos of gesture sequences with unambiguous actions (two intensional and two extensional). For each ambiguous video, two line drawings 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 depicting the same actor and patient, but a different action.

2.1.3 Procedure

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 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 presented in random order, and the order was different for each participant. The two answer possibilities were shown before each video and again afterwards.[3] The order of the two answer possibilities was randomly determined. The experiment took about ten minutes to complete.

[3]Showing 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.

2.2 Analysis and results

The data were analysed using a logit mixed effects regression, implementing the *lme4* package (Bates et al., 2015) in R (R Core Team, 2014). Our model analysed the fixed effects of gesture-order and country (both sum coded) on the interpretation. Participant was included as random intercept[4] and random slopes of gesture-order were included for item. The model revealed that participants were slightly more likely 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 ($\beta = 0.414$, $SE = 0.103$, $p < 0.001$), but no effect of country ($\beta = 0.001$, $SE = 0.086$, $p = 0.984$).[5] Figure 2 depicts the proportions of videos interpreted as extensional events, by gesture order.

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

2.2.1 Baseline study

To further investigate the overall preference for extensional interpretations, and to establish a baseline measure for each individual ambiguous action (independent of word order) we carried out an additional experiment. This experiment presented participants with the ambiguous action gestures only, instead of strings containing actor, action and patient gestures. Data was collected online (N=40; all participants except one were native speakers of English), on a crowdsourcing platform (Crowdflower; see `www.crowdflower.com`).

Like in the full-string experiment, participants were presented with ambiguous experimental 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 interpretations, presented as line drawings. Subsequently, the participant observed a video

[4] Including 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 (χ^2=0.000, p=.99).

[5] Upon 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: $\beta = 0.302$, $SE = 0.075$, $p < 0.001$.

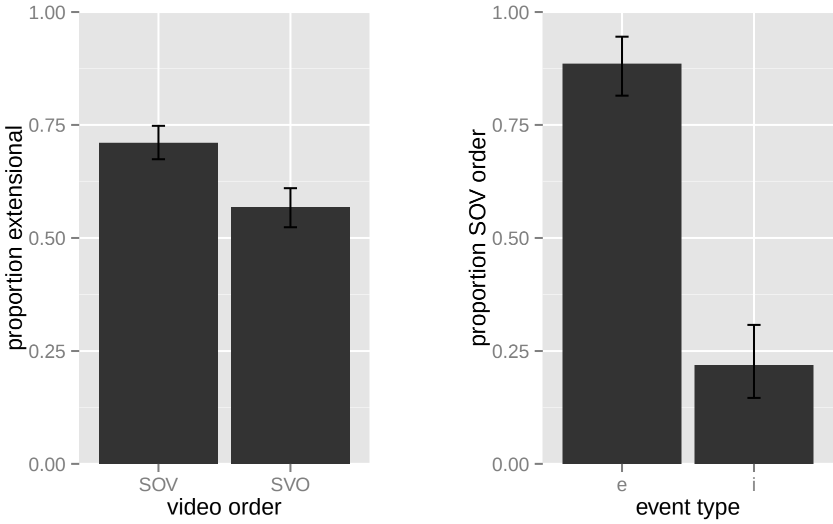

Figure 2: Main interpretation results are shown in left hand graph. Mean proportions of videos interpreted as extensional event are shown for SOV and SVO video orders. Error bars indicate 95% confidence intervals. The graph on the right hand side shows the production results from Schouwstra & de Swart (2014) for comparison. Proportion of strings in SOV order are displayed by event type (extensional and intensional). Error bars indicate 95% confidence intervals.

of an ambiguous action gesture; they then saw the two line drawings again, and were 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. Because the task was a two alternative forced choice task, there were no missing data points.

To compare the overall preference for extensional events in the full-string experiment to that in the verb-only experiment, we merged the data sets, and ran a logit mixed effects regression on the to model interpretation, with Experiment (verb-only vs full-string, 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 reflected in the model intercept: β = 0.953, SE =0.368, p <0.01. Crucially, no significant effect of experiment was observed (β = -0.215, SE =0.2659, p <0.419). From this we conclude that the two experiments saw no difference in the level of preference for extensional events.

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.[6] A linear mixed effects model was performed on these normalised values, taking gesture-order and country 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 Country as a predictor (χ^2=7.99, p<.001), while no significant difference was found between the full model and the model that omitted country as a predictor (χ^2=0.21, p=.90).

2.3 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, figure **??** depicts the effect of meaning on word order, taken from the production results presented in (Schouwstra & 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 expect similarly striking asymmetry in interpretation. This expectation may be misleading, because interpretation involves reasoning under uncertainty. If participants are accounting for this uncertainty,

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

then the impact of the word order biases may be dampened.

For example, the interpretation task – perhaps more so than the production task – is implicitly interactive: participants are interpreting the behaviour of another speaker. Participants have no knowledge of the speaker's linguistic system, and may be accounting for this uncertainty when making their decisions. A learner following these principles may be forced to consider disfavoured ordering systems that would be unlikely to play a role in the participant's own spontaneous productions: for all but the 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 asymmetries 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 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 ideal observer to account for the fact that the gesturer may be holding completely different coins - a more abstract consideration which could lead to uncertainty. In addition 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 a role during production).

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

3 Model: A computational Analysis of Gesture Interpretation

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 the experimental literature reviewed in section 1.1 provides intriguing hypotheses –

such as the hypothesis that independent heuristics drive production and interpretation (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 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 similar 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 discussed here and elsewhere in realted literature: the model can be tested against the experimental data, and can act as a benchmark against which alternative accounts can be contrasted.

The central abstraction in our analysis is that participant behaviour can be productively 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 employing these preferences when reasoning about the gesture orderings produced by another individual – in contexts where the intended meaning is unknown and must be reverse-engineered.

The Bayesian framework provides a natural model for this division of labour. This approach to statistical inference specifies a simple formula describing how a rational learner should update its beliefs about the nature of an unobserved mechanism responsible 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 underlying 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 underpinning Bayesian inductive inference align with human learning in many psychological domains (Chater, Oaksford, Hahn, & Heit, 2010; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). With respect to language, models of probabilistic rational inference have been applied to numerous aspects of linguistic structure (Chater & Manning, 2006), including word order generalisations in artificial grammar learning (Culbertson & Smolensky, 2012), and have been used to model the pragmatic principles underpinning production and interpretation of speech (Goodman & Frank, 2016;

Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013), but have not previously been explored as a model for the learning mechanisms that sub-serve improvised gesture.

Interpretation of improvised, not-yet-conventionalised communication is a particularly exciting focus for computational modeling of this sort because, in terms of the structure of the computational problem facing the interpreter, it has a distinctive character that is a-typical of linguistic communication: the interpreter is – knowingly - 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 which the learner is explicitly taught new conventions. Whether and how learners 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 important 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 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 our conclusions.

3.1 Interpretation through Bayesian Inference

3.1.1 Interpreting Gestures

Our model casts gesture interpretation as probabilistic inductive inference from an ordered 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:

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

where $p(g)$ is simply a normalising constant[7]. Learners' a priori expectations about the probability of each event type, $p(m)$, can be captured with a single parameter λ, 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 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 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 (Goodman & Frank, 2016), and is just as important in situations like ours where no existing linguistic conventions are established.

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 gestures composed of a single verb and object (SVO and SOV)[8]. Here p_{ext} is the probability of employing SVO to express an *Extensional* event: $p(g = \text{SVO}|m = \text{Ext}) = p_{ext}$; likewise p_{int} is the probability of using SVO to express an *Intensional* event: $p(g = \text{SVO}|m = \text{Int})$. The probabilities of employing SOV for Extensional and Intensional events respectively are $p(g = \text{SOV}|m = \text{Ext}) = 1 - p_{ext}$ and $p(g = \text{SOV}|m = \text{Int}) = 1 - p_{int}$.

An underlying system of associations $\vec{p}$ is tacitly assumed in equation 1, since $p(g|m)$ 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 $\vec{p}$. 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 both extensional and intensional events using SVO (i.e. $p_{ext} = p_{int} \approx 1$) or SOV

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

[8]For 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.

(i.e. $p_{ext} = p_{int} \approx 0$). Likewise, these ordering patterns could be in free variation ($p_{ext} = p_{int} = 1/2$), strong complementary conditioned usage (i.e. $p_{ext} = 1, p_{int} = 0$ or $p_{ext} = 1, p_{int} = 0$), weaker complementary usage (i.e. $p_{ext} = 1 - p_{int}$), or anything in between. We aim to compute a probability model for the decisions of a learner who accounts for this uncertainty.

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 decision[9]. 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 uncertainty inherent in the learner's observations. Under these assumptions the probability of choosing meaning m after observing gesture g is:

$$p(m|g) = \iint_{\vec{p}} p(m|g, \vec{p})p(\vec{p}) \ \mathrm{d}p_{ext}\,\mathrm{d}p_{int} \tag{2}$$

Here $p(m|g, \vec{p})$ is given by equation (1), with the conditioning on $\vec{p}$ made explicit. 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 verb. These biases impose probabilistic preferences on the space of possible associiation systems, and can be modelled with the Beta distribution (see Appendix A for details): $p(\vec{p}) = p(p_{ext})p(p_{int}) = \mathrm{Beta}(p_{ext}; \alpha_{ext}, \beta_{ext}) \cdot \mathrm{Beta}(p_{int}; \alpha_{int}, \beta_{int})$. The shape and strength of these preferences are determined by the prior parameters $\alpha_{ext}, \beta_{ext}, \alpha_{int}$, and β_{int}. An intuitive way to view these parameters is as *psuedo-counts*, or counts that

[9]Technically, we assume the learner considers all systems $\vec{p}$ that could have generated the observed gesture. An infinitesimely 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.

are added to the observed counts when predicting the probability of an outcome (α 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 experimental 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 assosciation systems ($\alpha_{ext} = \beta_{ext} = \alpha_{int} = \beta_{int} = 1$), is unsurprisingly the poorest predictor of the experimental data, affording participants' responses a combined log-likelihood of -256.98: 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 preference for Extensional events.

In order to ask whether the experimental result is being driven by general preferences 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 ($\alpha_{ext} = \beta_{ext} = \alpha_{int} = \beta_{int} = 1$) but fit λ to the experimental data. The maximum-likelihood estimate is $\hat{\lambda} = 0.68$: this model affords the data a combined log-likelihood of -225. The model-fit suggests a slight overall preference for Extensional events independent of 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. We found of λ to be one of the most consistent parameter estimates in our model. Though M1 is a better fit to the data than M0, it nevertheless fails to capture the assymetry between responses to SVO and SOV gestures.

These versions of the model suggest that – to account for the pattern of experimental results – the learning model we have described must include a non-neutral 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

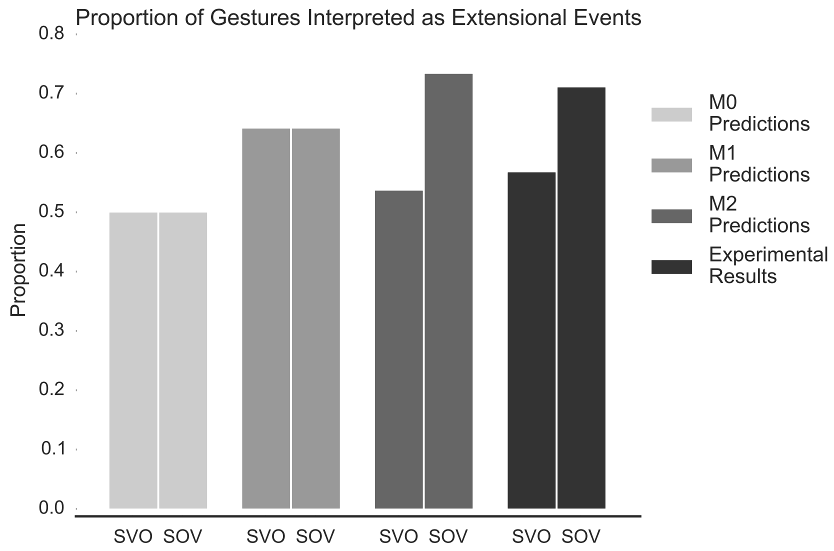

Figure 3: Comparison of model predictions and experimental results. The biased model (M2), but not the unbiased (M0) or the event-biased (M1) model, predicts both experimental results: asymmetric responses to SOV and SVO gestures, and an overall preferences for Extensional events. Model predictions show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

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' constituent ordering preferences are approximately complementary across event-types: *however strongly I prefer one ordering pattern for Extensional events, that's how 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.

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 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 equivelant or near equivelant values for $p(m|g)$, so the experimental data cannot choose between alternative priors reliably [10]. A natural solution is to reduce the number of 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 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 are met (more details below).

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 the data a combined log-likelihood of -215.977, correctly predicting participants' chosen interpretation with average probability 0.77. Figure 3 demonstrates the close correspondence between the model's predictions and participants' responses in our experiment.

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 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 a major concern. First, whilst we aren't able to identify a single best-fitting prior in the four-dimensional case, we are able to identify the maximum of the data likelihood function in this model (achievable under multiple "best-fitting" priors). Crucially, this maximum value is identical to the maximum value achieveable under the two-parameter "complementary priors" model (-215.977). In other words, the assumption of complementary biases does not reduce the likelihood of the data, suggesting that we should prefer the two-parameter version on grounds of parsimony anyway.

Second, we also analysed alternative assumptions within the restriction that only two parameters define the prior, and found these to be inferior. For example, rather than assuming "complementary" priors accross event types ($\alpha_{ext} = \beta_{int}, \alpha_{int} = \beta_{ext}$), 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 by two

[10] This 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.

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” assumption: respectively, the maximum of the likelihood function in these models is -216.92 and -225.03. Taken together, these analyses suggest that the ”complimentary priors” assumption is justified over alternatives, both practically and theoretically, so we will proceed to focus on this case.

3.3.2 Inferred Priors

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, and SVO for Intensional events. Second, the prior demonstrates a bias toward *regularity*: consistent usage of the favoured ordering is preferred over variable usage (probability 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 preference for regularity – motivated by simplicity principles and thought to be relevant to cognition in general – that has been proposed in various linguistic (Reali & Griffiths, 2009a; Smith & Wonnacott, 2010; Culbertson & Smolensky, 2012) and non-linguistic (Ferdinand, Thompson, Smith, & Kirby, 2013) domains.[11] Third, the prior expresses preferences that are skewed but *weak*; it encodes asymmetric ordering preferences, but these defeasible preferences could be easily overturned by observing contradictory data 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, 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. In the inferred prior, $s = 2.08$, suggesting just a handful of contradictory data-points could lead the learner to entertain disfavoured systems $\vec{p}$.

[11] Note 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.

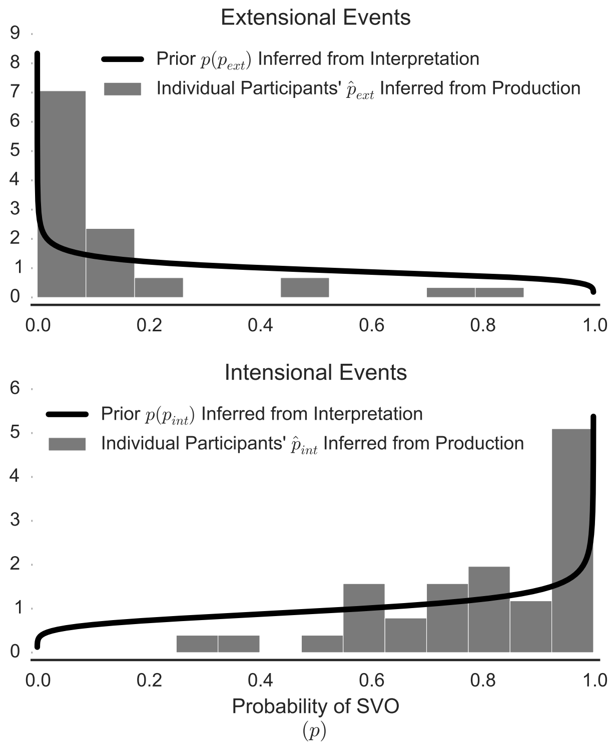

Figure 4: Lines show probability density functions for priors $p(p_{ext})$ (top) and $p(p_{int})$ (bottom) inferred from production data, superimposed on the (normalised) histograms of estimates of individual participants' $\hat{p}_{ext}$ (top) and $\hat{p}_{int}$ inferred from Schouwstra and de Swart's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

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 reasoning about the differences between production and interpretation in this context predicts that the shape of the inferred prior should be broadly compatible with the distribution of productions across participants, favouring most strongly the kinds of systems evidenced in production, but should also reserve some probability mass over a wider 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 **??**, these results show that our model is consistent with the differences we are attempting to explain between production and interpretation experimental results.

We analysed production data from (Schouwstra & de Swart, 2014)'s experiment and inferred maximum-likelihood estimates $\hat{p}_{ext}$ and $\hat{p}_{int}$ for each individual participant (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 interpretation. A correspondence between the distributions is clear: the biases we inferred from interpretation are consistent with the pattern of results observed independently in production. The prior favours the same strongly biased ordering systems that were most prominent in production, but also reserves some non-zero probability for alternative 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 low probability ordering systems are accounted for during interpretation, which dilutes the stronger asymmetry observed in production that was driven by favored, higher probability ordering systems.

It is also possible to directly compute the likelihood of the production data under 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 with average probability 0.77 for extensional gestures and 0.74 for intensional gestures.

3.4 Discussion

Our model provides one possible computational account for the main experimental finding that when interpreting gestures, participants used constituent ordering patterns 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, the basic proposal – that interpretation involves inference and estimation, and that the Bayesian framework provides a natural and useful model for understanding how learners 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 specific predictions about the posterior beliefs participants should entertain after observing labeled training examples, and it would be straightforward to construct experimental procedures that test these predictions. Likewise, plausible alternative explanations for asymmetry between production and interpretation could be formulated within this framework and directly compared.

Our model assumes that during interpretation, uncertain observers have principled 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 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 production and interpretation experiments. This explanation rests on the hypothesis that production in this particular scenario does not invoke the same abstract considerations (at least not to the same degree), but follows a lower level sampling process driven by favored ordering systems. These favoured ordering schemes will have high probability under the prior thanks to the same semantic biases, but we shouldnt expect that these biases are so strong as to rule out consideration of alternative ordering schemes. While the asymmetry may not hold for more interactive production scenarios in general, we believe this assumption is a conservative starting point which could easily be tested in future experiments that manipulate the degree of interactivity in production. An emerging body of research on the pragmatics of speech production and interpretation (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 modeling 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 interpretation decisions accordingly.

In the introduction section we saw that previous interpretation experiments have led to differing conclusions. Either, interpretation of improvised gesture, like production, by-passes the grammatical system, and prefers SOV order for simple transitive (extensional) events (Langus & Nespor, 2010), or production and interpretation each call for rather different explanations: a simple semantics based heuristic ('agent first') for interpretation, and specific production-related constraints (i.e., role conflict for reversible 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 improvisation 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 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 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.[12]

At the same time, there are important differences between silent gesture production and interpretation. While in both production and interpretation experiments, participants must improvise and cannot use their own language or any conventional language they know, the production experiment is more clearly than the interpretation experiment a situation that lacks linguistic conventions. Participants produce their gestures to the camera, and although there is an experimenter present, this experimenter is not 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 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 of interpretation through inductive inference, based around the principle that learners account for this uncertainty surrounding another individual's language use. The model suggests that participants' decisions – which varied by word-order but nevertheless portrayed uncertainty – may reflect motivated uncertainty in response to this unknown. Casting interpretation in this framework (Perfors, Tenenbaum, Griffiths, & Xu, 2011) connects improvised gesture with inference and estimation in other domains (e.g. Hsu, Chater, & Vitányi, 2011; Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Culbertson, Smolensky, & Legendre, 2012; Goldwater, Griffiths, & Johnson, 2009; Perfors, Tenenbaum, & Regier, 2011) through domain-general principles of inductive inference. Formulating questions about the emergence of linguistic rules/conventions in probabilistic 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 evidence. Going forward, we aim to explore principles for inductive inference further, in semi-supervised learning scenarios that systematically confirm or contradict 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 seemingly disfavoured ordering patterns be easily learned?* The computational principles

[12] In 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.

governing when and how people bring their biases to bear on language use under uncertainty are at present only superficially understood. Improvised gesture offers a rich testing ground for these questions, which we believe will be best understood through synthesis of experimental and computational analysis.

In the context of this paper, production and comprehension are studied separately, 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 (Pickering & Garrod, 2013). A logical next step is to extend the silent gesture paradigm to include communication (Christensen, Fusaroli, & Tylén, 2016) and cultural transmission through artificial generations of lab participants (Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2018; Schouwstra, Smith, & Kirby, 2016).

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 that the interpretation of silent gesture favours ordering preferences that are conditioned on meaning - similar to what was observed for silent gesture production. Fitting 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. 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 results from Schouwstra and de Swart (2014). On the other hand, the conditioned word order alternation may be easily overturned by contradictory linguistic observations. 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. However, it is well known that, under the right circumstances, weak inductive biases can shape regularities – sometimes even disproportionately strong regularities – over the course of cultural transmission (Kirby, Dowman, & Griffiths, 2007; Smith & Wonnacott, 2010; Griffiths & Kalish, 2007; Boyd & Richerson, 1985; Reali & Griffiths, 2009b). Understanding the cultural evolutionary forces that suppress this alternation in natural language is therefore a key priority for future research.

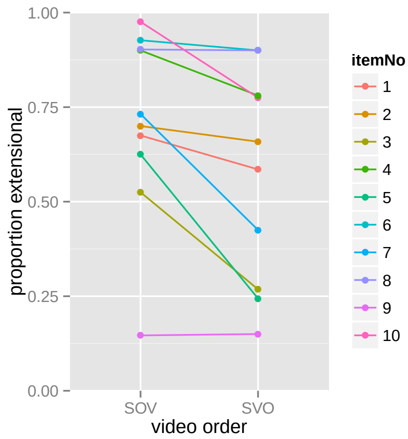

Figure 5: This shows the experimental results per item. The y axis shows the proportion of extensional interpretations.

5 Appendix A: Experimental stimuli and results by item

The following strings were used in the experiment (one ambiguous action per pair of 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 the proportion of extensional interpretations chosen in the verb-only experiment (see section 2.2.1).

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
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 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:

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

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 ith participant respectively, while N^{svo} and N^{svo} give the total number of SVO and SOV gestures observed, which did not vary accros participants. For the main two-paramter *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:

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

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

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

The first term in equations (B.2) and (B.3) gives $p(m = \text{Extensional}\,|\,g, \vec{p}, \lambda)$. The 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

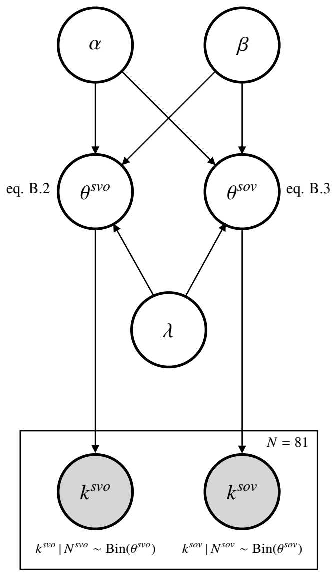

Figure 6: The model in graphical form, assuming *complementary priors* $p_{ext} \sim \text{Beta}(\alpha, \beta)$ and $p_{int} \sim \text{Beta}(\beta, \alpha)$.

in the paramters and the identity $\text{B}(\alpha, \beta) = \text{B}(\beta, \alpha)$ in the Beta function). Maximum likelihood estimates were obtained through numerical minimization of (the inverse of) eq. (B.1). Predictive probabilities reported throughout refer to the geometric mean of the combined log likelihood of all decisions. All optimization procedures reported were carried out using the Python library *Scipy*. Figure 6 shows the two-parameter version of the model in graphical form.

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 ith participant are:

$$\hat{p}_{ext} = k_i^{ext} / N_i^{ext} \tag{C.1}$$

$$\hat{p}_{int} = k_i^{int} / N_i^{int}, \tag{C.2}$$

where k_i^{ext} and k_i^{int} give the number of Extensional and Intensional events expressed using SVO, and N_i^{ext} and N_i^{int} are the total number of S-First gestures the ith participant produced for Extensional and Intensional events, respectively. When we report the 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 by the inferred parameters – the beta-binomial compound distribution.

8 Appendix D: Independent Beta Priors Analysis

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 the model which allows four free parameters (with $\lambda = .68$ fixed). Samples were collected under a uniform prior using an ensemble sampler with 250 walkers (Foreman-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 erratic: the model parameters are only weakly identifiable given the experimental data. We ommit pairwise correlation plots because they are largely uninformative. Note however 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 preferences, though we caution against overinterpreting parameter estimates in this version of model.

Acknowledgements

The authors would like to thank Marianne Smit, Nicky Mariën, Caglar Yali, Anouschka 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 British Academy, grant number pf130034; BT was funded by the ERC, grant number 283435.

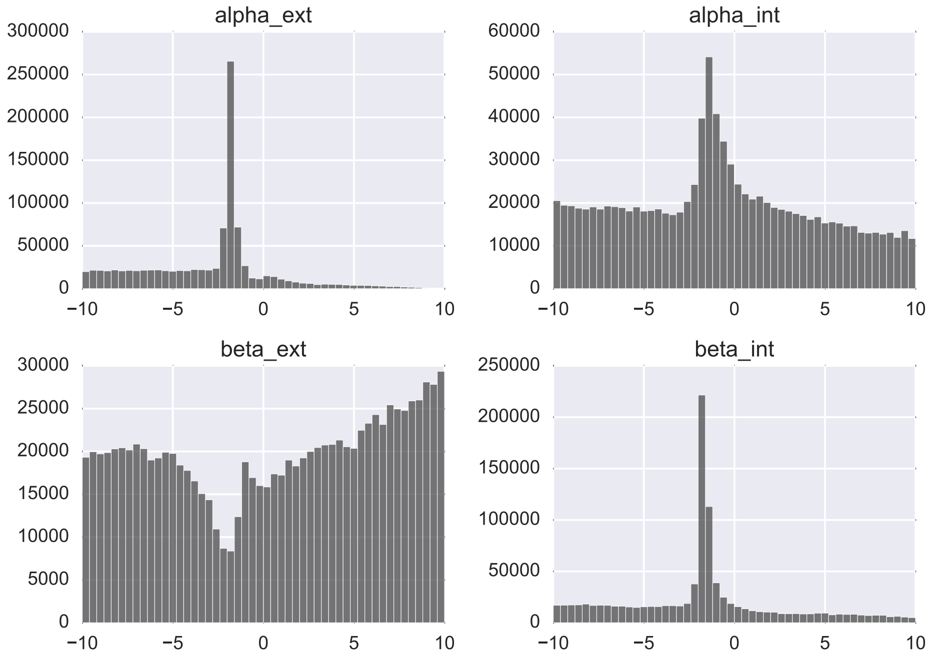

Figure 7: Posterior marginal samples in the *independent priors* version of the model, which allows four free paramters to determine prior distributions with independent shapes.

References

References

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., 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*, *125*(925), 306–312.

Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological* *science*, 0956797612463705.

Givon, T. (1997). On the co-evolution of language, mind and brain. *Evolution of communication*, *2*(1), 45–116.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, *105*(27), 9163–9168.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: exploring the effects of context. *Cognition*, *112*(1), 21–54.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in cognitive sciences*, *14*(8), 357–364.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive science*, *31*(3), 441–80.

Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of memory and language*, *81*, 16–33.

Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2014). Investigating constituent order change with elicited pantomime: A functional account of svo emergence. *Cognitive science*, *38*(5), 943–972.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17.

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.

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.

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.

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.

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive psychology*, *60*(4), 291–318.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in psychology*, *4*, 226.

Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lepic, 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.

Meir, I., Lifshitz, A., İlkbasaran, 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).

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2018, Apr). *The emergence of spatial modulation in artificial sign languages.* PsyArXiv.

Mudd, K., Kirby, S., & Schouwstra, M. (2018). Improvised word order biases are 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 urlhttp://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., & 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 international conference (evolang xi). evolang xi, new orleans, united states* (Vol. 20, p. 24).

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