



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### From improvisation to learning

**Citation for published version:**

Motamedi, Y, Wolters, L, Naegeli, D, Kirby, S & Schouwstra, M 2022, 'From improvisation to learning: How naturalness and systematicity shape language evolution', *Cognition*, vol. 228, 105206.  
<https://doi.org/10.1016/j.cognition.2022.105206>

**Digital Object Identifier (DOI):**

[10.1016/j.cognition.2022.105206](https://doi.org/10.1016/j.cognition.2022.105206)

**Link:**

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

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Cognition

**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](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.





# From improvisation to learning: How naturalness and systematicity shape language evolution

Yasamin Motamedi<sup>a,\*</sup>, Lucie Wolters<sup>a</sup>, Danielle Naegeli<sup>b</sup>, Simon Kirby<sup>a</sup>, Marieke Schouwstra<sup>c</sup>

<sup>a</sup> Centre for Language Evolution, University of Edinburgh, UK

<sup>b</sup> DCC, Tilburg University, the Netherlands

<sup>c</sup> ILLC, University of Amsterdam, the Netherlands

## ARTICLE INFO

### Keywords:

Language evolution  
Word order  
Silent gesture  
Learning  
Systematicity  
Naturalness

## ABSTRACT

Silent gesture studies, in which hearing participants from different linguistic backgrounds produce gestures to communicate events, have been used to test hypotheses about the cognitive biases that govern cross-linguistic word order preferences. In particular, the differential use of SOV and SVO order to communicate, respectively, extensional events (where the direct object exists independently of the event; e.g., *girl throws ball*) and intensional events (where the meaning of the direct object is potentially dependent on the verb; e.g., *girl thinks of ball*), has been suggested to represent a *natural* preference, demonstrated in improvisation contexts. However, natural languages tend to prefer *systematic* word orders, where a single order is used regardless of the event being communicated. We present a series of studies that investigate ordering preferences for SOV and SVO orders using an online forced-choice experiment, where English-speaking participants select orders for different events i) in the absence of conventions and ii) after learning event-order mappings in different frequencies in a regularisation experiment. Our results show that natural ordering preferences arise in the absence of conventions, replicating previous findings from production experiments. In addition, we show that participants regularise the input they learn in the manual modality in two ways, such that, while the preference for systematic order patterns increases through learning, it exists in competition with the natural ordering preference, that conditions order on the semantics of the event. Using our experimental data in a computational model of cultural transmission, we show that this pattern is expected to persist over generations, suggesting that we should expect to see evidence of semantically-conditioned word order variability in at least some languages.

## 1. Introduction

All languages can use the ordering of the major constituents of subject (S), object (O) and verb (V) to signal who does what to whom. However, while all 6 possible combinations of S, O and V, are found cross-linguistically, two in particular — SOV and SVO are most common, used as the preferred word order in the majority of documented languages (Dryer, 2013; Napoli & Sutton-Spence, 2014). How has this strong typological trend come about?

Silent gesture research, in which hearing participants without knowledge of a sign language are asked to communicate using gesture

and no speech, has tried to shed light on the cognitive biases that shape our preferences for some word orders over others. For example, Goldin-Meadow, So, Özyürek, and Mylander (2008) found that speakers from different language backgrounds overwhelmingly produced SOV-like sequences when gesturing. Other work has suggested that the picture is somewhat more complex, with word order preferences suggested to be affected by, for example, event reversibility (Gibson et al., 2013; Hall, Mayberry, & Ferreira, 2013), animacy (Meir et al., 2014), saliency (Kirton, Kirby, Smith, Culbertson, & Schouwstra, 2021). Schouwstra and de Swart (2014) proposed differing word order preferences based on the semantics of the events being described. They asked Dutch- and Turkish-

\* Corresponding author at: Centre for Language Evolution, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK.

E-mail address: [Yasamin.Motamedi@ed.ac.uk](mailto:Yasamin.Motamedi@ed.ac.uk) (Y. Motamedi).

speaking participants to produce gestures to communicate two types of events: extensional events, involving the manipulation of a direct object, usually with movement through space (e.g. *throw*, *carry*) and intensional events, where the meaning of the arguments, and the meaning of the direct object in particular, are interpreted in relation to the event itself, such as creation events like *bake* and *paint* and perception events like *think*, *imagine* and *dream*. That is, the direct object does not necessarily exist independently of the event.<sup>1</sup> They found that participants from both language backgrounds produced gesture sequences with different orders, conditioned on the semantics of the event — extensional events were produced most frequently with SOV-like order, while intensional events were produced with SVO-like order. We, among others (Goldin-Meadow et al., 2008; Schouwstra, Smith, & Kirby, 2020), suggest that this pattern represents a **natural** preference, reflecting the preferences for ordering patterns at the level of the item (here individual events) that occur in the absence of conventions. Proposals for why SOV and SVO orders represent a natural preference have been made based on a cognitive bias to present entity information (i.e. agents and patients) before relational information (i.e. actions; Gentner & Boroditsky, 2001). For intensional events, the entity information becomes more relational and is therefore presented later, resulting in an SVO preference (Schouwstra & de Swart, 2014). An iconic account was proposed by Christensen, Fusaroli, and Tylén (2016). In a study that elicited gestures for creation events (a sub-type of intensional events), they argued that SOV and SVO respectively can iconically represent the structure of different events. For events such as “the doctor eats the cake” (i.e. extensional events) both the subject and direct object must be co-present for the event to take place, while for events such as “the doctor bakes a cake” (i.e. intensional events), the event must take place for the direct object to exist. The two accounts have in common that they postulate a direct relation between meaning and structure. A different explanation is offered by (e.g.) Kline Struhl, Salinas, Lim, Fedorenko, and Gibson (2017) and Hall and colleagues (Hall, Ahn, Mayberry, & Ferreira, 2015; Hall et al., 2013), who suggest that production related biases are responsible for word order alternation in silent gesture.

While evidence of this type of semantically-conditioned word order has recently been discovered in two sign languages, Brazilian Sign Language (Libras; Napoli, Spence, & de Quadros, 2017) and Nicaraguan Sign Language (NSL; Flaherty, Schouwstra, & Goldin-Meadow, 2018), differential ordering patterns based on verb semantics is not commonly found cross-linguistically.<sup>2</sup> Rather, languages tend to use one word order across different events, regardless of the semantic properties of those events, which we refer to as **systematic** ordering. The question remains then, how we get from the natural ordering preference found in improvisation tasks to the systematic ordering preference found most commonly cross-linguistically.

Natural ordering occurs in improvisation tasks when participants are asked to communicate without existing conventions, while languages in the real-world are conventional systems that have been learnt by language users over many generations. We suggest that learning may play a role in shifting ordering preferences from natural to systematic order,

<sup>1</sup> In the literature, intensional transitive verbs have been defined as verbs in which the object argument is understood more in terms of its meaning than in terms of its reference. This has as a consequence that the objects of intensional verbs are potentially non-existent or nonspecific (Forbes, 2020), although there are several potential ways to distinguish intensionals from other verbs (Saul, 2002). The class of verbs has caused semanticists to rethink how verbs and their arguments combine, and has challenged conceptions of meaning in which reference plays a central role (Moltmann, 2020; Schwarz, 2020). Some have claimed a satisfactory definition of meaning in natural language depends solely on a correct analysis of intensional verbs (D'Ambrosio, 2019).

<sup>2</sup> There is evidence that similar semantic distinctions can condition form (including constituent order) in sign languages – e.g. psych-verbs (denoting an emotional state) in Sign Language of the Netherlands (NGT; Oomen, 2017) and animacy in Swedish Sign Language (SSL; Bjerva & Börstell, 2016).

with both ordering preferences representing biases that are at play in different contexts. For example, a body of previous work has shown that, in learning tasks, adult participants **regularise** the input they receive (Culbertson, Smolensky, & Legendre, 2012; Ferdinand, Kirby, & Smith, 2019; Saldana, Smith, Kirby, & Culbertson, 2018; Smith et al., 2017; Smith & Wonnacott, 2010). In the face of unpredictable variation, participants in these studies reduce variability for a particular item or category in the output they produce. For example, Smith and Wonnacott (2010) used an iterated learning paradigm to demonstrate that, through learning over generations of participants, systems that had different variants to mark plurality became more regular, with systems on the whole becoming more predictable. Regularisation behaviours have been shown in different levels of language (Saldana et al., 2018), and in both linguistic and non-linguistic domains (Ferdinand et al., 2019). However, a shift away from naturalness preferences would not necessarily constitute reduction of unpredictable variation, but a move away from semantic conditioning. Instead, the preference for systematic ordering patterns demonstrates a reduction in variation across all categories, including conditioned variation reflected in the preference for natural orders.

We suggest that the preference for naturalness appears in the absence of conventions, and reflects the gesturer's ordering preference for individual items. However, once conventions are established, the relations between form (here, order) and meaning (here, event) can be viewed as a *system* of mappings, leading to the simplest system to learn — systematic ordering in which a single order is used for all events (Culbertson & Kirby, 2016). To investigate our hypotheses, we conducted a set of online experiments in which participants are shown gesture sequences for extensional and intensional events. The gesture sequences they see are identical, differing only in the ordering of the constituent parts, which appear in either SOV or SVO order. Across four studies, participants are asked to select the gesture sequence which best conveys the target event, either in a task where participants have no previous experience with the gesture sequences (i.e. no conventional mappings) or in a learning task in which participants are first shown mappings between events and gesture sequences in different frequencies. Finally, we model the cultural evolution of the ordering preferences from our two learning experiments, to understand how ordering patterns might evolve through transmission to new learners over many generations. Across our experiments, we focus on 3 main measures that characterise the output participants produce: i) *regularisation*, which we define as the reduction of variation within a category, ii) *systematicity*, defined as the reduction in variation across categories, and iii) *naturalness*, which we define as an item-level preference at play in the absence of communicative conventions. We hypothesise that word order alternation based on intentionality is rooted in general cognitive preferences rather than production related constraints. Therefore we predict that in experiments 1a and 1b, in which participants select word orders in the absence of conventions, but without producing any gestures themselves, we will replicate the finding from Schouwstra and de Swart, in which participants prefer semantically-conditioned orders that reflect the structure of extensional and intensional events. In contrast, when participants are given variable input, we expect them to regularise that input, reducing variation for each event type. We also expect that learning will lead to a stronger preference for systematic languages that reduce variability across all events, and that the preference for natural, semantically-conditioned ordering patterns will substantially reduce following learning, reflecting the typological tendencies found in natural languages.

## 2. Experiment 1a: ordering preferences for events in the absence of conventions

### 2.1. Methods

Participants were recruited ( $N = 160$ ) from the crowdsourcing

platform Prolific for an online task in which they were shown two videos of gesture sequences describing an event, and asked to select which video they thought best conveyed the event. The experiment took 4 min to complete and participants were paid £1 for participation. We filtered participants using Prolific's screening options, including participants who were native English speakers and who had not completed a similar task posted by the authors. Participants were randomly assigned to one of two conditions in which they were shown gestures to describe either an extensional or an intensional event.

In the experiment, participants were presented with a line drawing (Fig. 1) showing one of two events: either *nun-throws-ukelele* (extensional event) or *nun-thinks-about-ukelele* (intensional event). The line drawing was shown in the upper middle part of the screen, with two videos showing the gesture descriptions positioned below the image, side by side (shown in Fig. 2). The two videos showed identical gesture descriptions with 3 iconic constituent gestures depicting, respectively, the actor, the action and the patient involved in the event. The only difference between the two gesture descriptions is the order in which the 3 gestures appear. One video showed the gestures in Actor-Action-Patient order (SVO), while the other showed the gestures in Actor-Patient-Action order (SOV), giving a total of 4 gesture description videos (2 for each event) with each video being 4.5 s in length (videos can be viewed at <https://osf.io/b9nm6/>). During the task, the two videos onscreen played in a continuous loop and were synchronised with each other, with the timing of each element within each video being such that the point of segmentation between S, V, and O is the same in both. The location of each video (left or right) was randomised for each participant.

Participants were asked to select which of the two gesture descriptions best conveyed the event shown in the image by clicking on the video to make their selection. Following selection, participants were presented with a screen showing a slider with a gesture description video at each end, the location of each video consistent with the selection task from the previous screen (Fig. 2). Participants were told to move the slider to indicate the strength of their preference for the selection they had made previously.

Following completion of the task, we excluded participants who a) made a selection in the selection task more quickly than the combined length of the two videos (9 s), suggesting that they did not attend to both completely, and/or b) expressed a preference in the slider task that was inconsistent with their choice in the selection task (i.e. expressed a preference for the SVO video when they had selected the SOV video in the selection task). This resulted in 31 excluded participants, leaving a total of 129 participants ( $N_{\text{extensional}} = 63$ ,  $N_{\text{intensional}} = 66$ ).

Data wrangling, visualisation and analysis was completed for experiment 1a and throughout using R (R Core Team, 2013). We used a logistic regression to model the responses from the selection task (as how often participants selected the SVO-variant), and a linear model to analyse responses from the slider task, which were transformed to represent the strength of preference for the selected variant. We include event-type as a deviation-coded predictor (extensional = -0.5, intensional = 0.5), such that the intercept represents the overall preference for the SVO-ordered sequence. Our method and analysis plan was pre-registered prior to data collection on the Open Science Framework (<https://osf.io/b9nm6/>).

## 2.2. Results

Fig. 3 shows the proportion of participants who selected the SVO-ordered sequence in the selection task for each event type. Our findings indicate that we do not find a significant preference overall for the SVO-ordered sequence ( $\beta = -0.35$ ,  $SE = 0.19$ ,  $z = -1.86$ ,  $p = 0.06$ ), but preference for the SVO-ordered sequence is greater when participants see the intensional event compared to the extensional event ( $\beta = 1.45$ ,  $SE = 0.38$ ,  $z = 3.78$ ,  $p < 0.001$ ). As in the click task, participants preference for the SVO-ordered sequence of the SOV sequence was higher

when they saw the intensional event compared to the extensional event ( $\beta = 0.20$ ,  $SE = 0.05$ ,  $t = 4.32$ ,  $p < 0.001$ ).

Our findings from experiment 1a therefore replicate in an online forced-choice selection task the natural ordering pattern observed in the gesture production task reported by Schouwstra and de Swart (2014), in which participants produced gestures for extensional events in SOV order and for intensional events in SVO order. This means that the preference is not dependent on production and is likely rooted in general cognitive biases. However, we only used one pair of events for the whole study. In experiment 1b, we test whether this preference holds across multiple extensional and intensional events.

## 3. Experiment 1b: ordering preferences for event categories in the absence of conventions

### 3.1. Methods

We recruited 162 participants from Prolific, according to the same payment and exclusion criteria described for experiment 1a. The experimental procedure was identical to experiment 1a, except that participants were randomly assigned to complete the task for one of eight events, 4 extensional events and 4 intensional events, given in Table 1. As in experiment 1b, we excluded participants who responded too quickly in the selection task, and whose slider task responses were inconsistent with their click task response, leaving 141 participants in total ( $N_{\text{extensional}} = 70$ ,  $N_{\text{intensional}} = 71$ ). The procedure and analysis of results was otherwise identical to that for experiment 1a.

### 3.2. Results

Our findings for experiment 1b are illustrated in Fig. 4. Our findings do not indicate a clear overall preference for the SVO-ordered gesture sequence compared to the SOV-ordered sequence ( $\beta = 0.31$ ,  $SE = 0.17$ ,  $z = 1.79$ ,  $p = 0.07$ ), though, across both event types, preference for the SVO sequence is higher than for experiment 1a ( $M_a = 0.43$ ,  $M_b = 0.57$ ). Consistent with experiment 1a, we find an increased preference for the SVO-ordered sequence for intensional events, compared to extensional events ( $\beta = 0.97$ ,  $SE = 0.35$ ,  $z = 2.77$ ,  $p < 0.001$ ), with this finding similarly reflecting in the strength of preference for the SVO sequence as indicated by the slider task ( $\beta = 0.14$ ,  $SE = 0.04$ ,  $z = 3.40$ ,  $p < 0.001$ ).<sup>3</sup>

### 3.3. Interim summary

We have used an online forced-choice selection task to replicate findings from a silent gesture production task (Schouwstra & de Swart, 2014). While a forced-choice selection task is not in itself equal to gesture production,<sup>4</sup> we demonstrate that constituent ordering preferences are conditioned on the semantics of the event in the absence of existing conventions. In both experiments 1a and 1b, participants demonstrated an increased preference for SVO-ordered gesture sequences when shown intensional events (in which the direct object is inherently linked to the event denoted by the verb), compared to extensional events (where the direct object exists independently of the event). Our experiment shows that the biases that drive this cannot be purely production related, and the naturalness preference must be rooted in general cognitive biases.

In experiments 2a and 2b, we investigate preferences for SVO and SOV-ordered gesture sequences for extensional and intensional events

<sup>3</sup> Stimuli, data and analysis files can be accessed at <https://osf.io/b9nm6>

<sup>4</sup> We have previously run a version of experiment 1a with different instructions, in which participants were first asked to imagine producing their own gestures to communicate the target event, before completing the forced-choice selection task. We find no difference in the results compared to the results we report here (see <https://osf.io/dpm34> for analysis of this study).

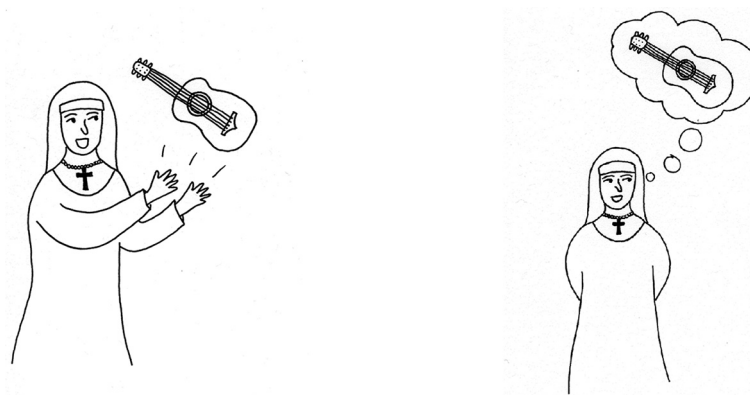


Fig. 1. Event images used in the experiment, showing an extensional event (left), nun-throws-ukelele, and an intensional event (right), nun-thinks-about-ukelele.

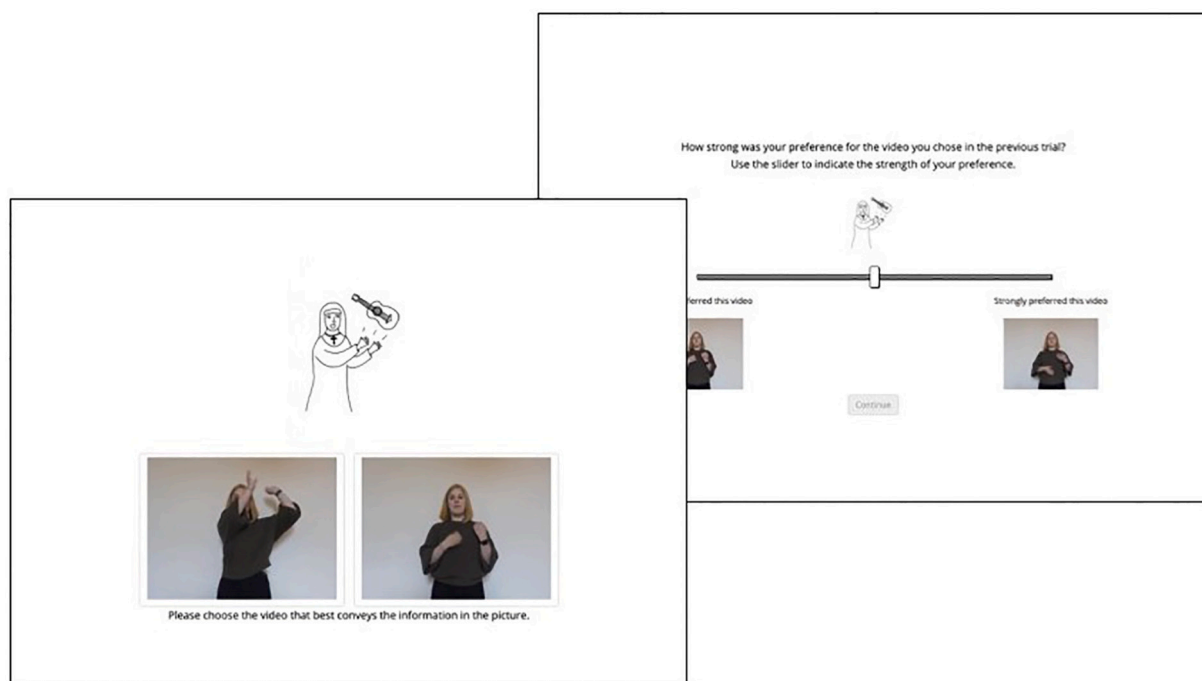


Fig. 2. Screenshots of the experiment, showing the forced-choice task (left) and the slider task (right).

beyond improvisation, using an artificial language learning experiment to test whether participants' preferences change as an effect of learning gestural descriptions with different frequencies of SOV and SVO.

#### 4. Experiment 2a: ordering preferences for events in a silent gesture learning task

##### 4.1. Methods

Participants ( $N = 200$ ) were recruited from Prolific to take part in an online gesture learning study in which they were first shown single gesture videos for different events in a training stage, and then asked to select gesture videos in a 2-alternative forced-choice task in a selection stage. The experiment took approximately 9 min to complete and participants were paid £1.31 for completion of the study. All participants were native speakers of English and had not taken part in any previous online gesture studies posted by the authors.

Materials used in experiment 2a are identical to those used in experiment 1a. Participants were shown line drawings of one extensional and one intensional event (as shown in Fig. 1), and two videos for

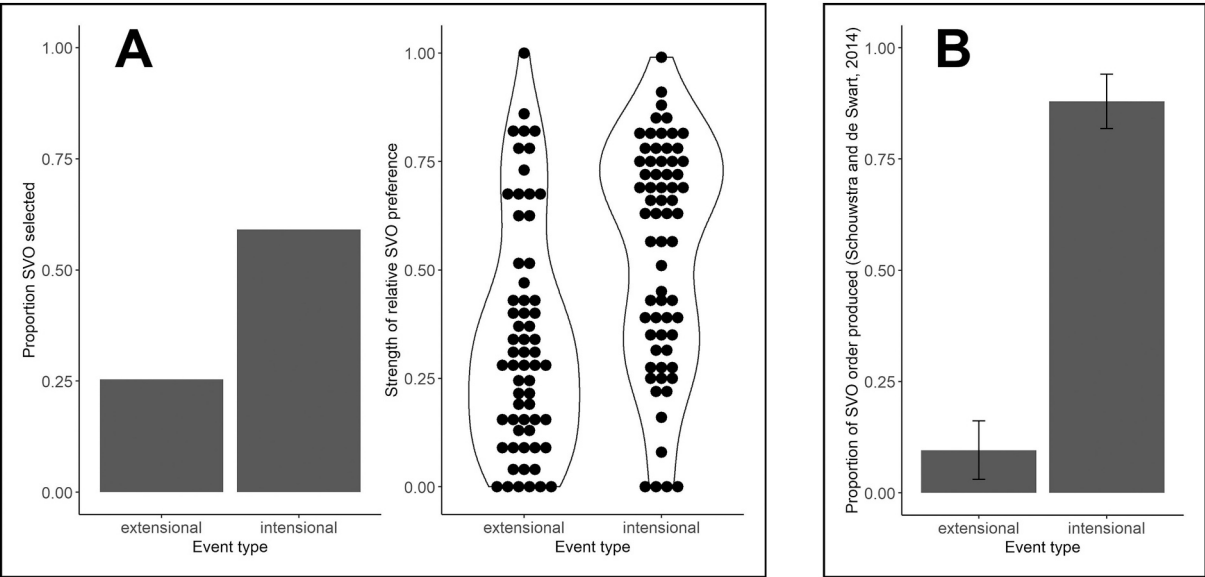
each event showing gesture sequences depicting the events in two orders, analogous to SVO-order and SOV-order.

Participants were randomly assigned to one of four conditions, which determined the input they received during the training stage of the experiment. During training, participants completed 20 trials in which an event image was shown with a single corresponding gesture video underneath, which could be SVO- or SOV-ordered. The frequency with which they saw each order was determined by the experimental condition (explained in detail below). For each event, each condition had a majority order in which participants saw the gesture sequence video in that order in 7 out of 10 trials.

##### 4.2. Procedure

The experiment consisted of 3 stages: a training stage, a selection stage and an estimation stage. In the **training stage**, participants completed 20 trials in which they saw either an extensional or an intensional event (10 trials for each event), with a gesture video shown on screen underneath the event image, and asked to watch the video carefully. Throughout training, participants saw gesture videos where

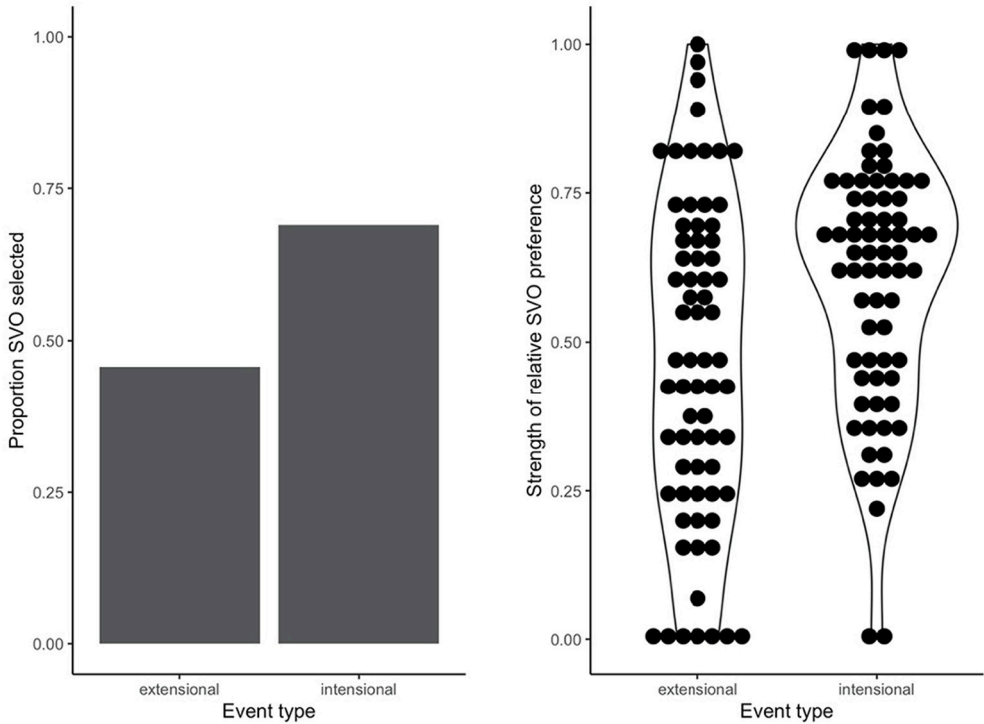




**Fig. 3.** (A) Proportion of participants who selected the SVO variant in experiment 1a, for each event type (A left), and the strength of preference for the SVO selection on the slider task (A right). (B) Proportion of SVO gesture sequences produced by participants in the study reported by Schouwstra and de Swart (2014) are added for comparison. All graphs show a main effect of event type (though the effect is strongest in the original experiment).

English descriptions of events used in experiment 1b.	
Extensional	Intensional
Nun throws ukulele	Nun thinks about ukulele
Burglar cuts scarf	Burglar knits scarf
Chef pushes pram	Chef dreams about pram
Gnome eats banana	Gnome paints banana

the constituent gestures appeared in both SVO and SOV order, but the frequency with which they saw each order depended on the condition they were randomly assigned to. Each condition had a majority order for each event, shown in 7 out of the 10 training trials for that event. In the *natural* condition, the majority order reflected the natural semantically-conditioned ordering preference found in experiments 1a and 1b and in previous work (Schouwstra & de Swart, 2014; Schouwstra, de Swart, & Thompson, 2019), such that participants saw a majority order of SOV for the extensional event and SVO for the intensional event. In the *unnatural* condition, the majority order was the inverse of the natural condition,



**Fig. 4.** Proportion of participants who selected the SVO variant in experiment 1b, for each event type (left), and the strength of preference for the SVO selection on the slider task (right).

**Table 2**

Frequency of order variants shown in training for each event type and each condition in experiment 2a.

Condition	Extensional event		Intensional event	
	SOV	SVO	SOV	SVO
Natural	7	3	3	7
Unnatural	3	7	7	3
Majority SVO	3	7	3	7
Majority SOV	7	3	7	3

seeing a majority order of SVO for the extensional event and SOV for the intensional event.<sup>5</sup> In the two remaining conditions, *majority SVO* and *majority SOV*, the majority order was the same for both event types (SVO and SOV, respectively), modelling a situation with systematic rather than semantically-conditioned ordering. For each participant, we randomised the order of presentation of event-gesture order combinations. The ordering patterns in each condition are given in Table 2.

In the **selection stage**, participants completed trials similar to the selection tasks in experiments 1a and 1b. At each trial, participants saw either an extensional or an intensional event with both the SVO- and the SOV-ordered videos underneath the image in a forced-choice task. Participants were asked to select the video “like they saw in the first part of the experiment [the training stage]”. Participants completed 20 trials in total, 10 trials for the extensional and 10 trials for the intensional event. Trial order was randomised for each participant, as was the location of the SVO- and SOV-ordered videos in each trial. Finally, in the **estimation stage**, participants were shown both event images accompanied by both of their corresponding videos (see Fig. 5), next to a numerical scale arranged from 0 to 10. Participants were asked to estimate, for each event and each video, how many times they saw the video during the training stage. The order of each event image (top, bottom) and each corresponding video was randomised per participant.

The design of the estimation stage closely follows that reported by Ferdinand et al. (2019), which tested how participants learnt from variable input in a non-linguistic task and a linguistic task using written stimuli. If participants modify the input they receive in training, those modifications (and in particular, regularisation behaviour) could be due to difficulties encountered during encoding, or during retrieval, or both. Our selection stage models the production/retrieval stage, while the estimation stage models how participants encode the frequencies they saw in training. Testing whether the frequencies participants approximate are different in the selection and estimation stages will allow us to

ascertain whether the difference between input and output is driven by encoding (estimation stage) or retrieval (selection stage).

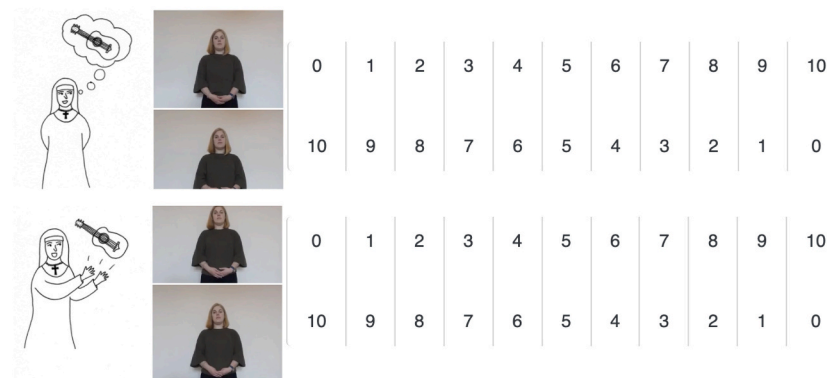
The design and analysis plans were pre-registered on the OSF framework prior to data collection; documentation relating the pre-registered study can be found at <https://osf.io/4wnjv>.

### 4.3. Results

An overview of participants’ output from the selection stage is shown in Fig. 6, showing the proportion of trials in which each participant selected the SVO variant for the intensional and extensional event. First, we aim to test whether the output participants produce in the selection task and the estimation task reflect the input frequencies seen in training. This allows us to determine whether participants learn word order in the gestural domain. We measure learning as the proportion of the output trials that match their input. Secondly, we are also interested in whether — as in other frequency learning paradigms such as those used by Culbertson et al. (2012) and Ferdinand et al. (2019) — participants will regularise by overproducing the majority order in their output. This would indicate that the regularisation process is modality agnostic. Finally, we want to know whether learning leads to systems that consistently use one word order across both events, or whether the preference for semantically-conditioned orders based on event type found in experiments 1a and 1b persists following learning. As such, our remaining analyses provide measures of *regularisation* (reduction of variation for a given event), *systematicity* (reduction of variation across events) and *naturalness* (semantic conditioning). Additionally, to test if participants learned the input languages they were given, we tested if their responses reflected the majority orders in these languages.

#### 4.3.1. Selection stage

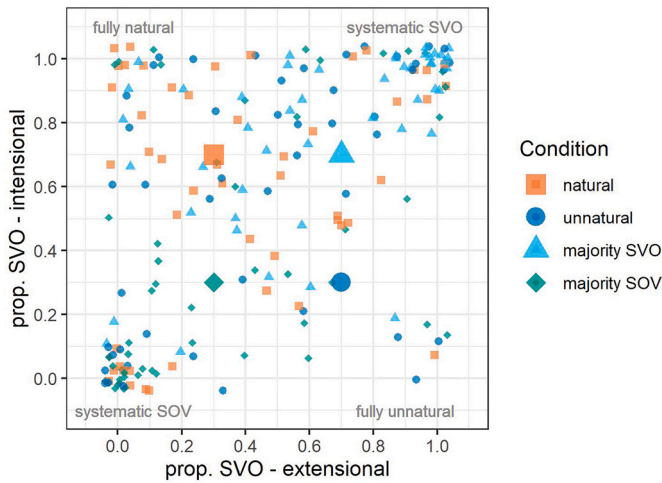
**4.3.1.1. Learning.** Firstly, we analysed whether participants learned from the input they received, measuring whether participants selected the majority order seen in training in each selection trial (Fig. 7A). We used a logistic mixed effects model<sup>6</sup> predicting reproduction of the majority order, including deviation-coded fixed effects of condition and event type (intensional/extensional) with a by-participant random intercept and a random slope of event type. The model including condition improved fit over the null model ( $\chi^2 = 29.61$ ,  $p < 0.001$ ); the inclusion of event type and the interaction between condition and event type did not improve model fit. The model intercept showed that, on average, participants select the majority order seen in training more



**Fig. 5.** Screenshot of the estimation stage of the experiment. Participants are shown each event with each gesture variant, and must estimate how often they saw each variant in training.

<sup>5</sup> We use the term ‘unnatural’ here to refer to the converse of the natural order in the context of these experiments. These terms are not used to evaluate relative ordering patterns in natural languages, nor should they be.

<sup>6</sup> Mixed-effects models here and throughout were implemented using *lme4* (Bates, Bolker, Machler, & Walker, 2015).



**Fig. 6.** Overview of selection stage results in experiment 2a, showing the proportions of SVO selections for both extensional (x-axis) and intensional (y-axis) events. Small shapes represent each individual participant, while large shapes represent the input values for each condition.

often than would be expected by chance ( $\beta = 0.83$ ,  $SE = 0.16$ ,  $z = 5.32$ ,  $p < 0.001$ ). Analysis of the fixed effects revealed that participants in the unnatural condition reproduced their majority order less frequently compared to the average across all conditions ( $\beta = -1.45$ ,  $SE = 0.27$ ,  $z = -5.32$ ,  $p < 0.001$ ), while participants in the SVO-majority condition reproduced their majority order more frequently on average ( $\beta = 0.94$ ,  $SE = 0.28$ ,  $z = 3.39$ ,  $p < 0.001$ ). We found no reliable differences for either the natural or SOV-majority conditions.

**4.3.1.2. Regularisation.** Here we follow Ferdinand et al. (2019) in defining regularisation as the reduction of variation in selected responses related to a given event. As such, we measure regularisation as the reduction in conditional entropy, which takes into account the probability of variants appearing in different contexts, given as:

$$H(V|C) = - \sum_{c_j \in C} p(c_j) \sum_{v_i \in V} p(v_i|c_j) \log_2 p(v_i|c_j)$$

where  $v$  is the set of variants (here SVO/SOV) and  $c$  the set of contexts they appear in (here the intensional or extensional event). We calculated the change in conditional entropy between participants' selection outputs and the input they received ( $H(V|C) = 0.88$  across all conditions; conditional entropy change shown in Fig. 7B). Inspection of the distribution of entropy change indicated that our data were non-normal and did not meet the assumptions for a linear modelling analysis. As such, we calculated 95% bootstrapped confidence intervals around the mean of each condition, as well as around the differences between condition means. To calculate our confidence intervals, we used the *boot* package in R (Canty & Ripley, 2021), generating 10,000 samples. We use the accelerated bias-corrected method as recommended by Puth, Neuhauser, and Ruxton (2015). Across conditions, the confidence intervals around the mean (Table 3) did not contain zero, indicating a drop in conditional entropy in each condition. Analysing differences across conditions, the confidence intervals given in Table 4 all contain zero, indicating that we do not find reliable differences in entropy change across conditions.

**4.3.1.3. Systematicity.** We analysed systematicity as the reduction in variation across both events in the system,<sup>7</sup> which we measure using Shannon entropy, where the entropy of a system is given as:

$$H(V) = - \sum_{v_i \in V} p(v_i) \log_2 p(v_i)$$

where  $V$  is the set of variants (here SOV and SVO orders). The most systematic ordering preference would use the same order for all event descriptions, giving an entropy value of 0; in contrast the natural and unnatural conditions have an input entropy of 1 because each order occurs in half of all trials. We calculated the entropy change between the input participants receive and the output they produce, illustrated in Fig. 7C. We also calculated 95% bootstrapped confidence intervals around the mean for each condition.

Table 5) and the difference between means across conditions (Table 6), following the same procedure as for the regularisation analysis. We find a reduction in overall entropy across conditions, such that the confidence intervals around the mean for each condition do not contain zero, but no reliable differences between conditions. That is, participants in all 4 conditions show evidence of systematisation.

**4.3.1.4. Naturalness.** Finally, we analysed the proportion of selection trials in which participants select the natural order based on the event (Fig. 7D), such that SVO is considered natural for intensional events and SOV for extensional events. We ran a logistic mixed effects model analysing whether selection order matched natural order, with a model structure identical to that used for our learning measure. Model comparison revealed that the full model with the interaction term represented the best fit in this case ( $\chi^2 = 34.6$ ,  $p < 0.001$ ). The model results showed a significant positive intercept ( $\beta = 0.62$ ,  $SE = 0.11$ ,  $z = 5.52$ ,  $p < 0.001$ ), indicating that, on average, participants selected the natural order more often than we would expect by chance (collapsed across conditions, natural order occurred in the input 50% of the time). The model revealed no significant main effects, but did show interactions between event type and the two majority order conditions, such that natural order is used more often on average for intensional events in the majority SVO condition ( $\beta = 3.64$ ,  $SE = 0.71$ ,  $z = 5.16$ ,  $p < 0.001$ ) and more for extensional events in the majority SOV condition ( $\beta = -2.99$ ,  $SE = 0.68$ ,  $z = -4.40$ ,  $p < 0.001$ ). That is, natural order is favoured in the systematic conditions when it is consistent with the majority input order.

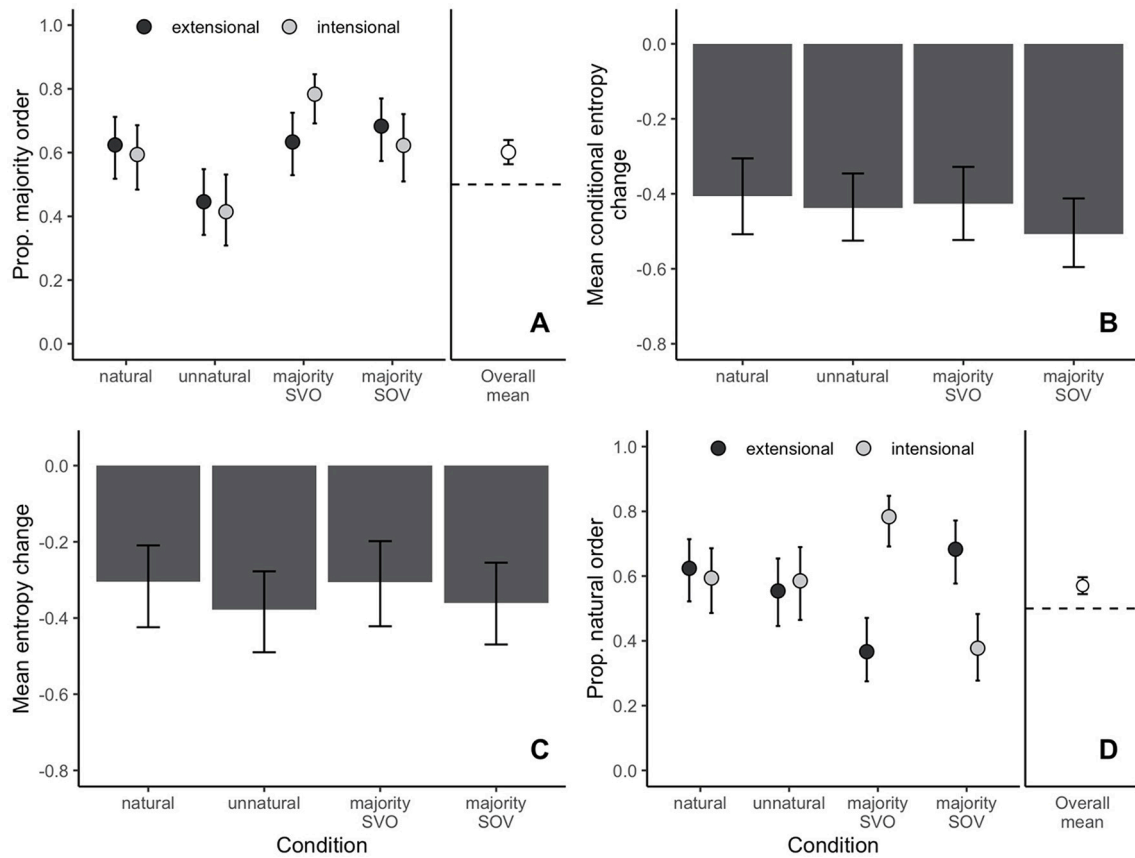
#### 4.3.2. Estimation stage

We compare output from the selection stage with participants' estimations of how often they saw each variant during training, to reveal the relative contributions of retrieval and encoding processes to changes made as a result of learning from variable input.

**4.3.2.1. Learning.** We used a logistic mixed effects model to analyse how well participants output reflected the frequency of the majority order seen in training (Fig. 8A). To ensure our output variables were comparable (as we do not have trial-by-trial binary values for the estimation stage), we took the total frequency of the SVO-variant selected, weighted by the number of trials (i.e. in the estimation trial, participants estimated how many times they saw the SVO variant out of the total number of trials). We included condition, event type and response type (selection/estimation) as deviation-coded fixed effects along with their interactions, with a by-participant random intercept and a random slope of event type. Model comparison indicated that the full model (with all

<sup>7</sup> Note that this measure differs from the planned measure described in our pre-registration, which operationalised harmonisation as the extent to which participants produced either of the two orders as a majority order. All analysis files, including the pre-registered analysis, can be found at <https://osf.io/rz7ea>.





**Fig. 7.** Experiment 2a selection stage results. A) Proportion of trials in which participants select the majority order seen in training, for each condition and event type, as well as the overall mean in relation to chance performance (0.5). B) Regularisation results showing a reduction in conditional entropy in each condition. C) Proportion of trials in which participants select the natural order, for each condition and event type, as well as the overall mean in relation to chance performance (0.5). D) Systematicity results showing a reduction in entropy in each condition. All error bars represent bootstrapped 95% confidence intervals around the mean.

**Table 3**

Mean conditional entropy change in each condition, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean.

Condition	$\bar{x}$	Lower 95% CI	Upper 95% CI
Natural	-0.41	-0.50	-0.30
Unnatural	-0.44	-0.52	-0.34
Majority SVO	-0.43	-0.52	-0.33
Majority SOV	-0.51	-0.60	-0.42

**Table 4**

Mean difference in conditional entropy change between conditions, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean difference.

Conditions	$\bar{x}_a - \bar{x}_b$	Lower 95% CI	Upper 95% CI
Natural - unnatural	0.03	-0.10	0.17
Natural - SVO	0.02	-0.12	0.16
Natural - SOV	0.10	-0.04	0.23
Unnatural - SVO	-0.01	-0.15	0.12
Unnatural - SOV	0.07	-0.06	0.20
SVO - SOV	0.08	-0.05	0.21

**Table 5**

Mean entropy change in each condition, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean.

Condition	$\bar{x}$	Lower 95% CI	Upper 95% CI
Natural	-0.30	-0.42	-0.21
Unnatural	-0.38	-0.49	-0.28
Majority SVO	-0.31	-0.42	-0.20
Majority SOV	-0.36	-0.47	-0.26

**Table 6**

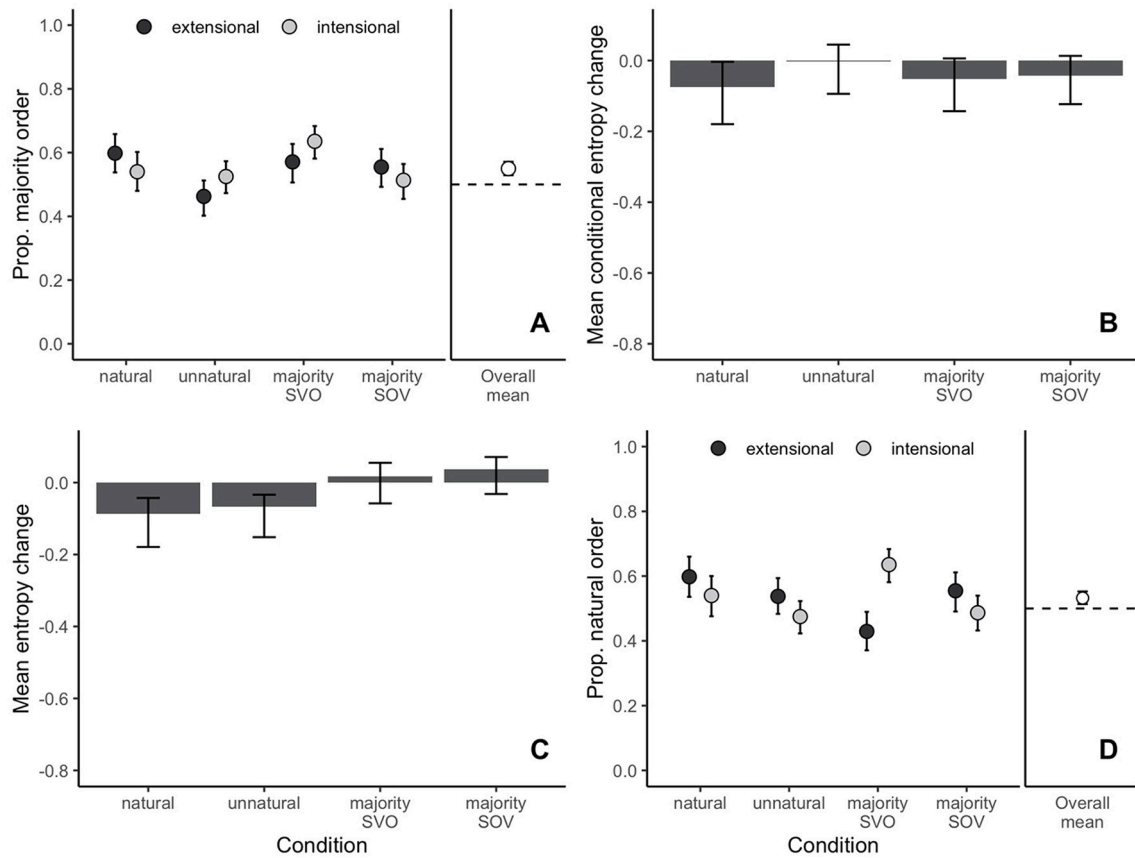
Mean difference in entropy change between conditions, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean difference.

Conditions	$\bar{x}_a - \bar{x}_b$	Lower 95% CI	Upper 95% CI
Natural - unnatural	0.07	-0.08	0.22
Natural - SVO	0.001	-0.15	0.16
Natural - SOV	0.06	-0.10	0.20
Unnatural - SVO	-0.07	-0.23	0.08
Unnatural - SOV	-0.02	-0.17	0.13
SVO - SOV	0.05	-0.10	0.21

interactions) improved fit over a reduced model without the three-way interaction ( $\chi^2 = 12.63$ ,  $p = 0.006$ ). Analysis of the model results<sup>8</sup> indicated that, overall, participants reproduced the frequencies seen in

training less often in the estimation stage than the production stage ( $\beta = -0.28$ ,  $SE = 0.05$ ,  $z = -5.44$ ,  $p < 0.001$ ), though responses for the estimation stage were higher than selection on average for the unnatural condition ( $\beta = 0.60$ ,  $SE = 0.09$ ,  $z = 6.78$ ,  $p < 0.001$ ), and lower for both the majority SVO ( $\beta = -0.32$ ,  $SE = 0.09$ ,  $z = -3.50$ ,  $p < 0.001$ ) and the majority SOV conditions ( $\beta = -0.35$ ,  $SE = 0.09$ ,  $z = -4.04$ ,  $p < 0.001$ ). Finally the three-way interaction term indicated that the difference between event types was larger in the estimation stage for the unnatural

<sup>8</sup> For clarity, we do not report all model results here. Full model results can be found at <https://osf.io/rz7ea>.



**Fig. 8.** Experiment 2a estimation stage results. A) Proportion of trials in which participants select the majority order seen in training, for each condition and event type, as well as the overall mean in relation to chance performance (0.5). B) Regularisation results showing a reduction in conditional entropy in each condition. C) Proportion of trials in which participants select the natural order, for each condition and event type, as well as the overall mean in relation to chance performance (0.5). D) Systematicity results showing a reduction in entropy in each condition. All error bars represent bootstrapped 95% confidence intervals around the mean.

condition ( $\beta = 0.49$ ,  $SE = 0.18$ ,  $z = 2.78$ ,  $p = 0.005$ ) and smaller for the majority SVO condition ( $\beta = -0.53$ ,  $SE = 0.18$ ,  $z = -2.87$ ,  $p = 0.004$ ).

**4.3.2.2. Regularisation.** We compared change in conditional entropy between the input participants received and their responses in the selection and estimation stages (Fig. 8B illustrates results for the estimation stage). We calculated 95% confidence intervals around the mean for each condition in the estimation stage (Table 7), as well as around the mean differences between selection and estimation for each condition (Table 8). Our analysis indicates a small reduction in conditional entropy in the natural condition, but no reliable change in the other conditions. Comparison with the selection stage indicates a significantly smaller reduction in conditional entropy across conditions in estimation than selection.

**4.3.2.3. Systematicity.** We compared change in overall entropy (i.e., not conditioned on event) between the input participants received and their responses in the selection and estimation stages (estimation stage results

**Table 8**

Mean difference in conditional entropy change between the selection and estimation stage for each condition, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean difference.

Condition	$\bar{x}_{est} - \bar{x}_{select}$	Lower 95% CI	Upper 95% CI
Natural	-0.33	-0.45	-0.20
Unnatural	-0.43	-0.54	-0.32
Majority SVO	-0.37	-0.49	-0.24
Majority SOV	-0.47	-0.57	-0.34

shown in Fig. 8C). We calculated 95% confidence intervals around the mean for each condition in the estimation stage (Table 9), as well as around the mean differences between selection and estimation for each condition (Table 10). Our analysis indicates a small reduction in entropy in the natural and unnatural conditions, but no reliable change in the two majority order conditions. Comparison with the selection stage indicates a significantly smaller reduction in entropy across conditions in estimation than selection.

**Table 7**

Mean conditional entropy change in each condition in the estimation stage, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean.

Condition	$\bar{x}$	Lower 95% CI	Upper 95% CI
Natural	-0.08	-0.18	-0.004
Unnatural	-0.003	-0.09	0.04
Majority SVO	-0.05	-0.15	0.004
Majority SOV	-0.04	-0.12	0.01

**Table 9**

Mean entropy change in each condition in the estimation stage, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean.

Condition	$\bar{x}$	Lower 95% CI	Upper 95% CI
Natural	-0.09	-0.17	-0.04
Unnatural	-0.07	-0.15	-0.03
Majority SVO	0.02	-0.06	0.05
Majority SOV	0.04	-0.03	0.07

**Table 10**

Mean difference in entropy change between the selection and estimation stage for each condition, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean difference.

Condition	$\bar{x}_{\text{est}} - \bar{x}_{\text{select}}$	Lower 95% CI	Upper 95% CI
Natural	-0.22	-0.34	-0.10
Unnatural	-0.31	-0.43	-0.19
Majority SVO	-0.32	-0.45	-0.20
Majority SOV	-0.40	-0.51	-0.28

**4.3.2.4. Naturalness.** We use the same procedure as our learning measure to analyse to what extent participants' responses in the selection and estimation stages indicate a preference for natural ordering patterns (Fig. 8D shows results from the estimation stage). Model structure was identical to that described for our learning measure. The full model with the three-way interaction term demonstrated a significantly better fit over a reduced model ( $\chi^2 = 75.75$ ,  $p < 0.001$ ). Overall, we found that participants exhibit less preference for natural order in the estimation stage than in the selection stage ( $\beta = -0.22$ ,  $SE = 0.05$ ,  $z = -4.23$ ,  $p < 0.001$ ). In addition, inspection of the three-way interaction indicates that the difference between event types was smaller in the estimation stage than in selection for both the majority SVO condition ( $\beta = -1.07$ ,  $SE = 0.18$ ,  $z = -5.77$ ,  $p < 0.001$ ) and the majority SOV condition ( $\beta = 1.41$ ,  $SE = 0.18$ ,  $z = 8.00$ ,  $p < 0.001$ ).

In summary, we have found that participants trained on SOV and SVO gesture variants communicating a single extensional and a single intensional event regularise the variable input they receive in training, in two ways. Participants show both a preference for naturalness, such that they produce the natural ordering pattern more often than would be expected by chance, and also systematising behaviour, where one order is used more frequently to describe both events. Moreover, the change between input and output is primarily driven by the retrieval biases tested in the selection task, rather than the encoding bias tested in the estimation task. In experiment 2b, we ask whether these behaviours hold when participants are trained on gestures for multiple events that can be grouped into extensional and intensional events.

## 5. Experiment 2b: ordering preferences for event categories in a silent gesture learning task

Participants ( $N = 200$ ) were recruited from Prolific to take part in an online experiment almost identical in procedure to experiment 2a, where they were first trained on gesture sequences for events and then asked to select gesture sequences for events in the selection stage. Our findings from study 2a suggested that differences between the input and output were driven primarily by selection/retrieval and not estimation/encoding; as such, we did not include the estimation stage in study 2b. All participants were native speakers of English and had not taken part in any previous online gesture studies posted by the authors. While in experiment 2a the task used only 1 intensional-extensional pair of events, here we used 4 intensional-extensional event pairs and corresponding gesture sequences, identical to those used in experiment 1b. This allows us to test whether the same patterns hold where conditioning is on event type, rather than specific events.

The design of experiment 2b, including assignment to the 4 conditions was identical to that of experiment 2a, with one exception. In the present experiment, participants saw a randomised set of 3 out of the 4 event pairs (to reduce the total duration of the experiment). As such, they completed 8 trials for each event in training and selection (48 trials in total in each stage) and, in training, saw the majority order in 6/8 trials (75% majority).

### 5.1. Results

Fig. 9 gives an overview of participants' output in experiment 2b. We

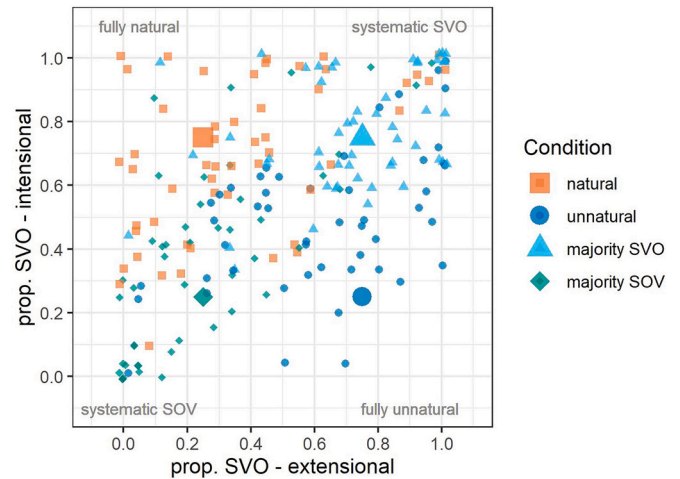


Fig. 9. Overview of selection task results in experiment 2b.

apply the same measures here as in experiment 2a, measuring i) learning ii) regularisation, iii) systematicity and iv) the preference for natural order. All analysis procedures, including model structure, are identical to those used for experiment 2a. Pre-registration, data files and analysis scripts can be found at <https://osf.io/smvwp>.

#### 5.1.1. Learning

Learning results are shown in Fig. 10A. Model comparison indicated that the model with the interaction term improved fit over a reduced model ( $\chi^2 = 10.60$ ,  $p = 0.01$ ). Analysis of the model revealed a significant positive intercept (grand mean;  $\beta = 1.02$ ,  $SE = 0.09$ ,  $z = 11.36$ ,  $p < 0.001$ ), indicating that participants selected the majority order seen in training more often than we would expect by chance. As in experiment 2a, participants in the unnatural condition selected the majority order seen in training less often than average ( $\beta = -0.60$ ,  $SE = 0.15$ ,  $z = -3.90$ ,  $p < 0.001$ ), while participants in the majority SVO condition selected the training majority more often than average ( $\beta = 0.49$ ,  $SE = 0.15$ ,  $z = 3.17$ ,  $p = 0.002$ ). In addition, we found an interaction between condition and event type, such that for the majority SOV condition, participants demonstrated lower reproduction of the input majority for intensional compared to extensional events ( $\beta = -0.60$ ,  $SE = 0.29$ ,  $z = -2.04$ ,  $p = 0.04$ ).

#### 5.1.2. Regularisation

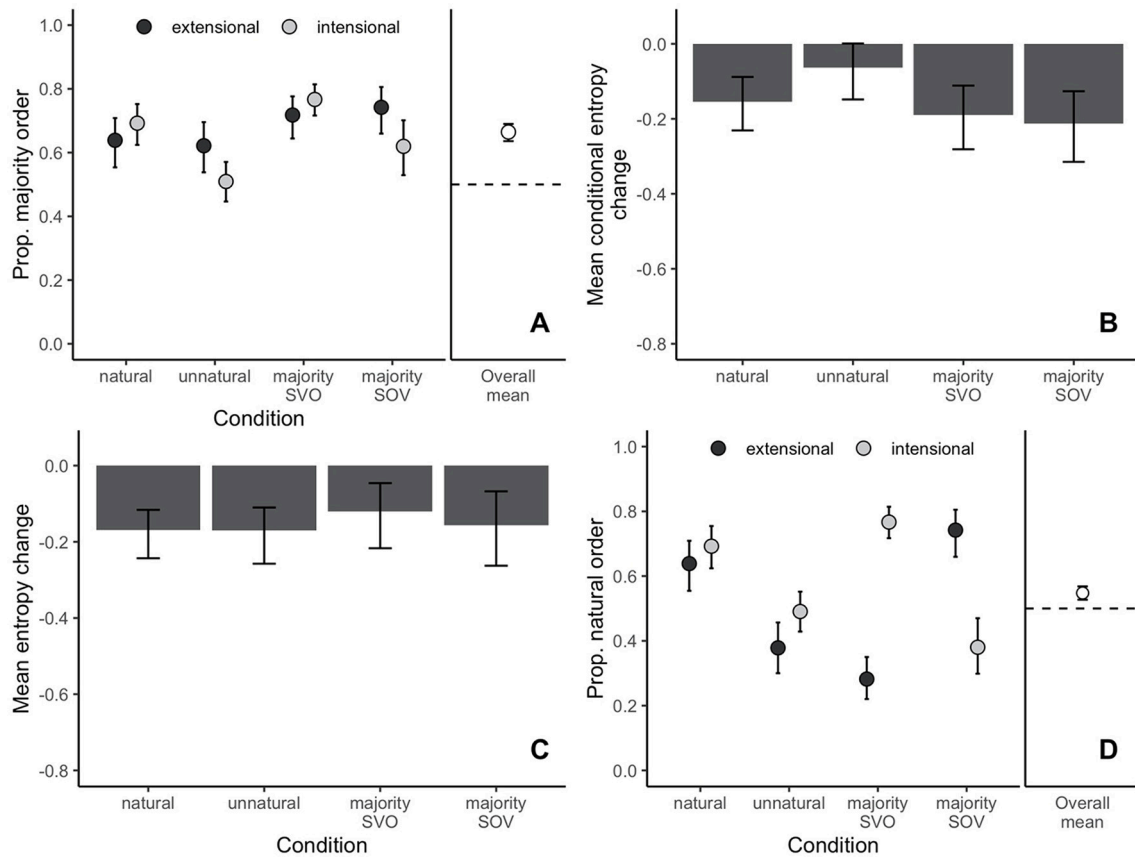
The reduction in conditional entropy in each condition is illustrated in Fig. 10B. Mean values and 95% confidence intervals around the mean for each condition are given in Table 11; mean differences between conditions and 95% confidence intervals are given in Table 12. We find a reduction in conditional entropy in all conditions except for the unnatural condition, though overall this reduction is lower on average than for experiment 2a. Inspection of the differences between conditions suggest that the two majority order conditions show a higher average reduction in conditional entropy than the unnatural condition.

#### 5.1.3. Systematicity

Fig. 10D illustrates the reduction in Shannon entropy in each condition, with mean values and 95% confidence intervals given in Table 13. Participants in each condition demonstrate a reduction in entropy, though lower on average compared to that found in experiment 2a ( $x = -0.15$ ), with no reliable differences between conditions (see Table 14).

#### 5.1.4. Naturalness

The proportion of selection trials in which participants selected the natural order is shown in Fig. 9C. A model analysing natural order



**Fig. 10.** Experiment 2b results. A) Proportion of trials in which participants select the majority order seen in training, for each condition and event type, as well as the overall mean in relation to chance performance (0.5). B) Regularisation results showing a reduction in conditional entropy in each condition. C) Proportion of trials in which participants select the natural order, for each condition and event type, as well as the overall mean in relation to chance performance (0.5). D) Systematicity results showing a reduction in entropy in each condition. All error bars represent bootstrapped 95% confidence intervals around the mean.

**Table 11**

Mean conditional entropy change in each condition, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean.

Condition	$\bar{x}$	Lower 95% CI	Upper 95% CI
Natural	-0.15	-0.23	-0.09
Unnatural	-0.06	-0.15	0.002
Majority SVO	-0.19	-0.28	-0.11
Majority SOV	-0.21	-0.31	-0.13

**Table 12**

Mean difference in conditional entropy change between conditions, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean difference.

Conditions	$\bar{x}_a - \bar{x}_b$	Lower 95% CI	Upper 95% CI
Natural - unnatural	-0.09	-0.19	0.02
Natural - SVO	0.04	-0.07	0.15
Natural - SOV	0.06	-0.06	0.18
Unnatural - SVO	0.13	0.02	0.24
Unnatural - SOV	0.15	0.03	0.27
SVO - SOV	0.02	-0.10	0.15

selection including the full fixed effects structure (including the interaction between condition and event type) improved fit over the model without the interaction term ( $\chi^2 = 64.94$ ,  $p < 0.001$ ). Inspection of the model results revealed a significant positive intercept ( $\beta = 0.28$ ,  $SE = 0.06$ ,  $z = 4.96$ ,  $p < 0.001$ ), indicating an overall preference for the natural ordering pattern. Model results suggest that participants in the unnatural condition selected the natural order less often than on average

**Table 13**

Mean entropy change in each condition, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean.

Condition	$\bar{x}$	Lower 95% CI	Upper 95% CI
Natural	-0.17	-0.25	-0.12
Unnatural	-0.17	-0.26	-0.11
Majority SVO	-0.12	-0.21	-0.05
Majority SOV	-0.16	-0.26	-0.07

**Table 14**

Mean difference in entropy change between conditions, with lower and upper bounds for bootstrapped 95% confidence intervals around the mean difference.

Conditions	$\bar{x}_a - \bar{x}_b$	Lower 95% CI	Upper 95% CI
Natural - unnatural	0.001	-0.09	0.10
Natural - SVO	-0.05	-0.15	0.06
Natural - SOV	-0.01	-0.13	0.11
Unnatural - SVO	-0.05	-0.16	0.06
Unnatural - SOV	-0.01	-0.13	0.11
SVO - SOV	0.04	-0.10	0.17

( $\beta = -0.69$ ,  $SE = 0.10$ ,  $z = -7.12$ ,  $p < 0.001$ ), while the interaction terms suggested, as in experiment 2a, that the majority order conditions demonstrated a greater preference for the natural order when it was consistent with the training majority (majority SVO:  $\beta = 2.57$ ,  $SE = 0.37$ ,  $z = 6.88$ ,  $p < 0.001$ , majority SOV:  $\beta = -2.82$ ,  $SE = 0.39$ ,  $z = -7.29$ ,  $p < 0.001$ ).

## 5.2. Discussion

In experiment 2a and 2b, participants reduced the variability in the word order patterns that they learned, both within event type, and across event types. In other words, they both *regularise* and *systematise* the input. At the same time, however, participants show an overall preference for *naturalness*: the selected responses were more likely to be natural than unnatural. This preference interacted with the rules of the input language, such that participants were more likely to choose a natural order if that was the majority order in the language they had been trained on. However, while these preferences occur after a single period of learning, languages are transmitted from one generation of learners to the next, for many generations. In the following section, we use our experimental results to model this process of cultural evolution, to understand how these preferences might evolve over longer timescales.

## 6. Predicting the evolution of ordering preferences

As languages are transmitted from one generation to the next, the learning preferences of each consecutive generation has an effect on the data produced, so over generations the language is shaped by accumulating learning preferences (Kirby, Griffiths, & Smith, 2014). The data from experiments 2a and 2b describe one generation of observing, processing and producing SOV and SVO word orders for multiple events (where we take the choice of two gesture videos as a stand-in for full production that we might see in a lab-based artificial sign language learning experiment). However, these results do not inform us directly about how these processes shape word-order after multiple generations. To investigate this, we use an iterated learning model of cultural transmission in which the produced data of one learner serves as the training data for another learner. Griffiths and Kalish (2007) have shown that iterated learning is equivalent to a finite Markov chain, which is a discrete-time random process over a sequence of variants ( $v_{t=1}, v_{t=2}, \dots, v_{t=n}$ ), in which only the previous value ( $v_{t-1}$ ), has an influence on the current value ( $v_t$ ):

$$P(v_{t=1}, v_{t=2}, \dots, v_{t=n}) = P(v_t | v_{t-1})$$

A Markov process is specified by a transition matrix, which defines the probabilities of each possible observation state ( $S_i$ ) to transition to each possible production state ( $S_j$ ) after one step of time ( $t$ ). In the Markov process of experiments 2a and 2b, we make the simplifying assumption that there are only four possible states corresponding to each of the four experimental conditions: natural, unnatural, majority SOV, majority SVO.<sup>9</sup> The transition probabilities of each state were determined by the productions of participants in the respective condition. So, for example, if a participant produced majority SOV, irrespective of what proportion they actually produced, they would be treated as producing an output state corresponding to the majority SOV state in the transition matrix. A word order was considered the majority order for an event type when it was produced in >50% of the trials. Data from participants without a majority order for either event type (i.e., who produced equal proportions of each word order) were excluded from the transition matrices (experiment 2a:  $n = 24$ ; experiment 2b:  $n = 9$ ). Transition matrices for are shown in the top panels of Fig. 11 for experiment 2a (A) and experiment 2b (B).

We can take a distribution of different language types represented as a vector and multiply this by the transition matrix to get the distribution of different language types expected after one generation of learning. Repeating this process  $n$  times models the change in distribution after  $n$

generations. For most cases of language evolution that can be modelled this way, there will be some distribution of language types that remain unchanged if they are multiplied by the transition matrix, and furthermore this distribution will eventually be reached by the process of transmission given enough generations. This is termed the *stationary distribution* and can be thought of as the probability of different languages over time after the influence of the starting state has been washed out by sufficient generations of language change. In the context of experiment 2a and 2b, the stationary distribution is a distribution over the four states of the system, where each probability corresponds to the proportion of time the system will spend in each production state. The stationary distribution of a transition matrix is proportional to its first eigenvector. Stationary distributions are shown in the bottom panels of Fig. 11, for experiment 2a (A) and 2b (B).

These results suggest that, while all output states (i.e. language types) are possible, they vary in their probability in the stationary distribution, with the majority orders being overall most common (comprising around 75% output languages), but with natural systems also likely to occur (in approximately 20% output languages). In contrast, unnatural systems are expected to be very unlikely to persist over time (< 5% of cases).

## 7. Discussion

Previous experimental work has shown that, when asked to produce gestures without speech, the order in which participants produce gestures is conditioned on the semantics of the events, with extensional events tending to appear in SOV sequences and intensional events appearing in SVO sequences (Schouwstra & de Swart, 2014). Here, we replicate and extend that finding. In experiments 1a and 1b, we find a similar preference for natural semantically-conditioned orders in a forced choice task, demonstrating that the preference for natural ordering is not solely a production bias, but a preference that operates in the absence of existing conventions.

In experiments 2a and 2b, we asked whether this naturalness preference persists following learning, or whether learning would lead to a shift towards a preference for systematic orders, where a single order is used regardless of event semantics. We found that participants do show increasing systematisation following learning, with entropy reducing across both extensional and intensional event categories. This possibly reflects a general bias for simplicity (Culbertson & Kirby, 2016), where the simplest system is one in which all events are expressed with a single order. However, we also find that the preference for natural systems still persists and appears to exist in competition with a preference for systematic ordering patterns, contrary to our predictions.

Our learning task also replicates previous findings from regularisation experiments in a new modality, finding that participants are able to learn from gestural input in an online artificial language learning experiment and that participants regularise variable input in the manual modality to a similar extent as written or spoken stimuli (Culbertson et al., 2012; Ferdinand et al., 2019). In particular, the reduction in conditional entropy we find across conditions is comparable to that found by Ferdinand et al. (2019) for their linguistic task (i.e. with a more extreme reduction in entropy than the non-linguistic task). Moreover, our results from experiment 2a showing greater change between input and output in the production task compared to the estimation task, suggest, in line with Ferdinand et al. (2019) and Saldana et al. (2018), that the changes we see between input and output are better explained as operating during retrieval than during encoding.

Our finding that naturalness can persist beyond learning suggests that the preference for semantically-conditioned ordering patterns continues even once conventions are established. Importantly, this does not appear to be the case in the unnatural condition, where the input frequencies were not learned well by participants, suggesting that the preference for natural ordering (where extensional events are preferred in SOV order and intensional events in SVO order) is about the specific

<sup>9</sup> Note that, in the context of the current experiment, this is a substantial simplification of the possible input - output space, which, in reality, would contain  $2^n$  states for all possible sequences of  $n$  responses (here 10 per event in experiment 2a, 8 per event in experiment 2b).



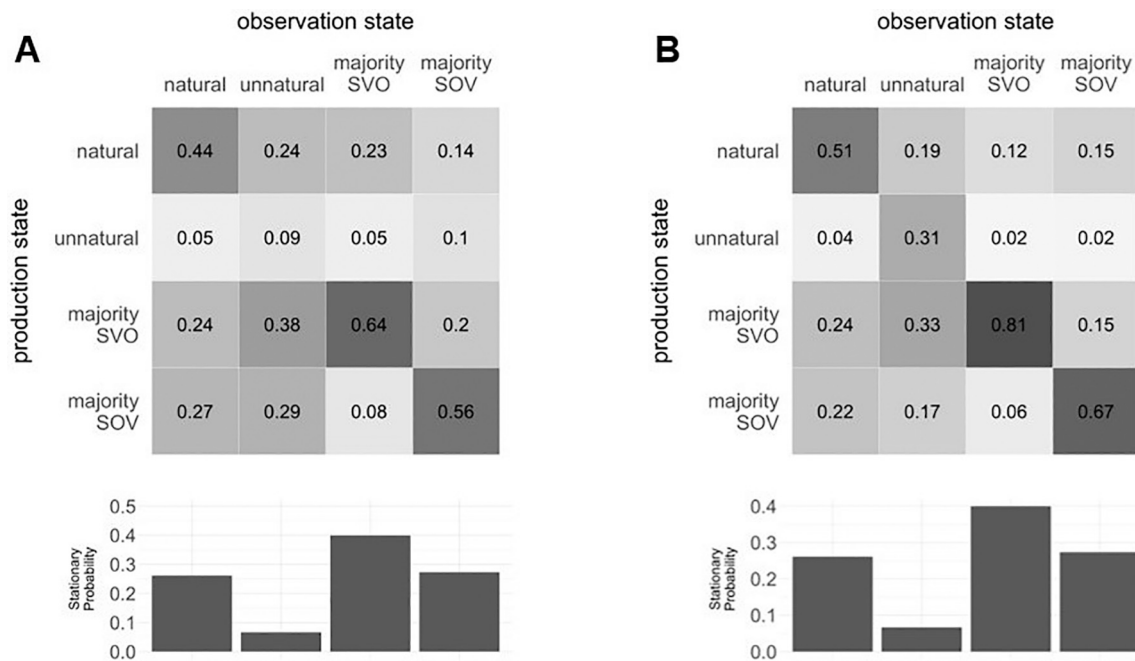


Fig. 11. Transition matrices (top) and stationary distribution (bottom) for experiments 2a (A) and 2b (B).

mapping between event and order rather than a general preference for consistent conditioned variation. Previous explanations for the natural preference have relied on iconicity; for extensional events, subject and object must be co-present before the event takes place, but for intensional events the direct object is a product of the event itself (i.e. the cake does not exist until you bake it). That is, the preferred orders reflect the temporal or conceptual structure of the events themselves. In this way, the unnatural mapping is dispreferred because it is *anti-iconic* and runs counter to participants' expectations. In addition, we find little difference in participants' behaviours in experiment 2a, where they see gestures for only 2 events and experiment 2b, where they see multiple instances of extensional and intensional events. Therefore, we assert that the natural ordering preference can be regular, such that it applies across whole categories and not just to individual items, as we found in experiments 1a and 1b.

We also used the data from our learning experiments to model the cultural transmission of word order preferences over time. Previous work has shown that weak bias can be amplified through cultural transmission (Griffiths & Kalish, 2007; Kirby, Dowman, & Griffiths, 2007; Real & Griffiths, 2009; Thompson, Kirby, & Smith, 2016), so learning at a single time point may not be sufficient to explain the types of structures we see in natural languages. Our work suggests that both preferences for systematic and natural ordering patterns will be preserved over time, but that languages with unnatural orders will be strongly dispreferred relative to the other biases at play.

Our finding that naturalness does persist beyond learning runs counter to our expectations that a preference for natural ordering would give way to a preference for systematic languages, given that most of the world's languages are characterised as having one dominant word order. Instead, our findings suggest that languages should be able to evolve regular ordering patterns that are either natural or systematic, and we should expect to see both types of languages in the real world. Recent data from two sign languages, NSL (Flaherty et al., 2018) and Libras (Napoli et al., 2017) have offered evidence of the natural ordering pattern, in line with this prediction. This demonstrates that, while it may be rare in natural languages, it is nevertheless possible. Furthermore, it is possible that natural ordering patterns cross-linguistically, though they may not be most frequent for a given language, may be present at some level if a gradient view of word order is taken, such that different

ordering preferences may be evident in different contexts. For example, Levshina et al. (2021) strongly advocate for a gradient approach as the default approach to word order, with (for example) animacy, information structure and dependency length affecting processing and production of word orders in natural languages. Indeed, previous silent gesture research has uncovered similarly varied factors affecting the ordering of gesture sequences in the lab (Gibson et al., 2013; Hall et al., 2013; Kirton et al., 2021; Meir et al., 2014). At present, more work is needed to understand the contexts that shape and shift ordering preferences over time. Future work should investigate how different interacting factors may lead to the types of gradient ordering systems we see in languages today. Also, it should be noted that word order is by no means the only strategy for conveying who does what to whom; alternative strategies include case marking, and (particularly in signed languages) usage of space. Another important line of work is collecting evidence (from corpora as well as experiments) to investigate the tradeoff between word order and other strategies (e.g., Bjerva & Börstell, 2016; Fedzechkina, Newport, & Jaeger, 2017; Hörberg, 2018; Levshina, 2021; Tal & Arnon, 2022).

## 8. Conclusion

In the absence of existing conventions, participants prefer orders that are semantically conditioned on the semantics of the event — SOV order for extensional events and SVO order for intensional events. In a learning task, participants regularised the variable input they received, suggesting that linguistic regularisation operates similarly across modalities. However, the output participants produced demonstrated preferences for both the natural semantically-conditioned order, as well as systematic ordering that used the same word order across events. We show that natural ordering patterns can persist beyond communication without conventions, appearing in competition with a bias for systematicity. The observations from NSL and Libras have shown us what these languages can look like, and an important direction for future research is to investigate how other natural languages combine natural and systematic ordering in one and the same system.

## Author contributions

**YM:** Conceptualisation, Design and methodology, Data processing and analysis, Writing, Writing – review and edits, **LW:** Conceptualisation, Design and methodology, Investigation, Data processing and analysis, Writing, Writing – review and edits, **DN:** Conceptualisation, Design and methodology, Investigation, Data processing and analysis, **SK:** Funding acquisition, Conceptualisation, Design and methodology, Writing – review and edits, **MS:** Funding acquisition, Conceptualisation, Design and methodology, Data processing and analysis, Writing – review and edits

## Acknowledgements

This work was funded by the Economic and Social Research council (ES/R011869/1), awarded to MS and SK. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

## References

- Bates, D., Bolker, B., Machler, M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bjerva, J., & Börstell, C. (2016). Morphological complexity influences verb-object order in Swedish sign language. In D. Brunato, F. Dell'Oretta, G. Venturi, T. Francois, & P. Blache (Eds.), *Proceedings of the 1st Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)* (pp. 137–141). International Committee on Computational Linguistics.
- Canty, A., & Ripley, B. D. (2021). *boot: Bootstrap R (S-Plus) Functions (1.3-25)* [Computer software].
- 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. <https://doi.org/10.1016/j.cognition.2015.09.004>
- Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, 6(January), 1–11. <https://doi.org/10.3389/fpsyg.2015.01964>
- Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122(3), 306–329. <https://doi.org/10.1016/j.cognition.2011.10.017>
- D'Ambrosio, J. (2019). Semantic verbs are intensional transitives. *Mind*, 128(509), 213–248. <https://doi.org/10.1093/mind/fxz029>
- Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer, & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/81>
- Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science*, 41(2), 416–446. <https://doi.org/10.1111/cogs.12346>
- Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, 184, 53–68. <https://doi.org/10.1016/j.cognition.2018.12.002>
- Flaherty, M., Schouwstra, M., & Goldin-Meadow, S. (2018). Do we see word order patterns from silent gesture studies in a new natural language? In C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 12th international conference (EVOLANGXII)*. NCU Press. <https://doi.org/10.12775/3991-1.030>
- Forbes, G. (2020). Intensional transitive verbs. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (winter 2020). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2020/entries/intensional-trans-verbs/>
- Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman, & S. Levinson (Eds.), *Language acquisition and conceptual development* (1st ed., pp. 215–256). Cambridge University Press. <https://doi.org/10.1017/CBO9780511620669.010>
- 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*, 24(7), 1079–1088. <https://doi.org/10.1177/0956797612463705>
- Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, 105(27), 9163–9168. <https://doi.org/10.1073/pnas.0710060105>
- Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31(3), 441–480. <https://doi.org/10.1080/15326900701326576>
- 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. <https://doi.org/10.1016/j.jml.2014.12.003>
- Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, 129(1), 1–17. <https://doi.org/10.1016/j.cognition.2013.05.004>
- Hörberg, T. (2018). Functional motivations behind direct object fronting in written Swedish: A corpus-distributional account. *Glossa: A Journal of General Linguistics*, 3(1), Article 1. <https://doi.org/10.5334/gjgl.502>
- Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12), 5241–5245. <https://doi.org/10.1073/pnas.0608222104>
- Kirby, S., Griffiths, T. L., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28C, 108–114. <https://doi.org/10.1016/j.conb.2014.07.014>
- Kirton, F., Kirby, S., Smith, K., Culbertson, J., & Schouwstra, M. (2021). Constituent order in silent gesture reflects the perspective of the producer. *Journal of Language Evolution*. <https://doi.org/10.1093/jole/lzaa010>
- Kline Struhl, M., Salinas, M. A., Lim, E., Fedorenko, E., & Gibson, E. (2017). Word order patterns in gesture are sensitive to modality-specific production constraints. *PsyArXiv*. <https://doi.org/10.31234/osf.io/62puh>
- Levshina, N. (2021). Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.648200>
- Levshina, N., Namboodiripad, S., Allasonnière-Tang, M., Kramer, M. A., Talamo, L., Verkerk, A., ... Stoyanova, N. (2021). Why we need a gradient approach to word order. *PsyArXiv*. <https://doi.org/10.31234/osf.io/yg9bf>
- Meir, I., Aronoff, M., Börstell, C., Hwang, S., Ilkbasaran, D., Kastner, I., Lifshitz, A., Basat, B., Padden, C., & Sandler, W. (2014). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, 158(April), 1–40. <https://doi.org/10.1016/j.cognition.2016.10.011>
- Moltmann, F. (2020). Truthmaker semantics for natural language: Attitude verbs, modals, and intensional transitive verbs. *Theoretical Linguistics*, 46(3–4), 159–200. <https://doi.org/10.1515/tl-2020-0010>
- Napoli, D. J., Spence, R. S., & de Quadros, R. M. (2017). Influence of predicate sense on word order in sign languages: Intensional and extensional verbs. *Language*, 93(3), 641–670. <https://doi.org/10.1353/lan.2017.0039>
- Napoli, D. J., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology*, 5, 1–18. <https://doi.org/10.3389/fpsyg.2014.00376>
- Oomen, M. (2017). Iconicity in argument structure. *Sign Language & Linguistics*, 20(1), 55–108. <https://doi.org/10.1075/sll.20.1.03oom>
- Puth, M.-T., Neuhauser, M., & Ruxton, G. D. (2015). On the variety of methods for calculating confidence intervals by bootstrapping. *Journal of Animal Ecology*, 84(4), 892–897. <https://doi.org/10.1111/1365-2656.12382>
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Real, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–328. <https://doi.org/10.1016/j.cognition.2009.02.012>
- Saldana, C., Smith, K., Kirby, S., & Culbertson, J. (2018). Is regularisation uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order. *PsyArXiv*. <https://doi.org/10.31234/osf.io/zk59y>
- Saul, J. M. (2002). *II—Jennifer Saul: What are Intensional Transitives?* *Aristotelian Society Supplementary*, 76(1), 101–119. <https://doi.org/10.1111/1467-8349.00091>
- Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131(3), 431–436. <https://doi.org/10.1016/j.cognition.2014.03.004>
- Schouwstra, M., de Swart, H., & Thompson, B. (2019). Interpreting silent gesture: Cognitive biases and rational inference in emerging language systems. *Cognitive Science*, 43(7), Article e12732. <https://doi.org/10.1111/cogs.12732>
- Schouwstra, M., Smith, K., & Kirby, S. (2020). *The emergence of word order conventions: Improvisation, interaction and transmission*. <https://doi.org/10.31234/osf.io/wdfu2>
- Schwarz, F. (2020). Intensional transitive verbs. In *The Wiley Blackwell companion to semantics* (pp. 1–33). American Cancer Society. <https://doi.org/10.1002/9781118788516.sem113>
- Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use, and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B*, 372, 20160051. <https://doi.org/10.1098/rstb.2016.0051>
- Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3), 444–449. <https://doi.org/10.1016/j.cognition.2010.06.004>
- Tal, S., & Arnon, I. (2022). Redundancy can benefit learning: Evidence from word order and case marking. *Cognition*, 224, Article 105055. <https://doi.org/10.1016/j.cognition.2022.105055>
- Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, 113(16). <https://doi.org/10.1073/pnas.1523631113>, 201523631.