Hierarchical Inferences Support Systematicity in the Lexicon

Matthias Hofer (mhofer@mit.edu)[1], Tessa Verhoef[2], & Roger Levy[1]
[1]Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences
43 Vassar Street, Cambridge, MA 02143 USA
[2]University of Leiden, Leiden Institute of Advanced Computer Science
Niels Bohrweg 1, 2333 CA Leiden, Netherlands

Abstract

Language exhibits striking systematicity in its form-meaning mappings: Similar meanings are assigned similar forms. Here we study how systematicity relates to another well-studied phenomenon, linguistic regularization, the removal of unpredictable variation in linguistic variants. Systematicity is ultimately a property of *classes* of form-meaning mappings, each member of which can be acted upon independently by linguistic regularization. Both are supported by a cognitive bias for simplicity, but this leaves open the question of how they interact to structure the lexicon. Using data from a recent artificial gesture learning experiment by Verhoef, Padden, and Kirby (2016), we formalize cognitive biases at the item level and the language level as inductive biases in a hierarchical Bayesian model. Simulated data from models that lack either one of those biases show how both are necessary to capture subjects' systematicity preferences. Our results bring conceptual clarity about the relationship between regularization and systematicity and promote a multi-level approach to cognitive biases in artificial language learning and language evolution. **Keywords:** systematicity; Bayesian modeling; regularization; sign language; artificial language learning

Introduction

One fundamental feature of language is that mappings between forms and meanings are systematic. A set of form-meaning relationships exhibits *systematicity* if signs for similar meanings share similar forms. While form-meaning associations are largely arbitrary in spoken language at or below the the level of the morpheme, systematicity is abundant at higher levels of linguistic organization. The phrases "the blue chair", "the broken chair", or "the inexpensive chair", for instance, all refer to propositions that include the meaning *chair* by virtue of sharing the form "chair". Similarly, forms that express categories of meanings such as *actions* often share common morphological features (e.g., English verbs that denote an ongoing action in progress share the suffix "-ing").

One widespread view in language evolution is that systematicity is the result of competing pressures for simplicity and informativeness in cultural transmission (Kirby, Griffiths, & Smith, 2014; Kirby, Tamariz, Cornish, & Smith, 2015). On this account, systematicity is preferred because it is a simple yet efficient way to organize a large lexicon.

In this paper, we present a modeling case study to investigate the relationship between systematicity, which is by definition a property of *classes* of form-meaning relationships, and linguistic *regularization*, a cognitive process that generates simplicity by removing unpredictable variation from the input (Smith & Wonnacott, 2010). While both phenomena are supported by cognitive biases for simplicity, their relationship is unclear because they are typically studied in isolation.

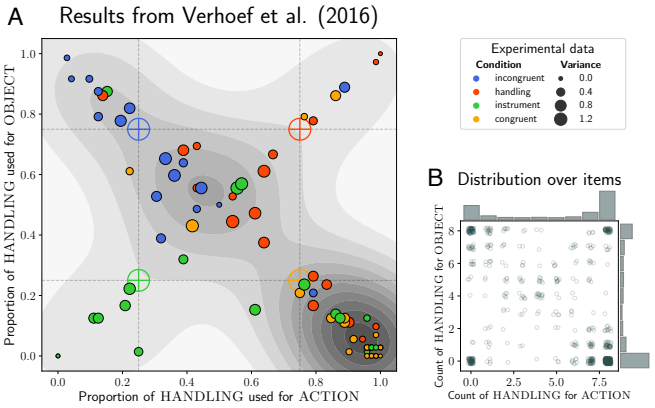

Figure 1: Results from Verhoef et al. (2016). **A.** Data from the test phase of the experiment, averaged by participant, displaying varying levels of systematicity in the output languages. **B.** The same data before averaging by participant shows high degrees of item-specific regularization.

To this end, we model data from a recent artificial language learning experiment (Verhoef et al., 2016, Figure 1), which was conducted to study the role of gestural preferences and linguistic regularization in the emergence of systematicity in sign language lexicons (Padden, Meir, & Lepic, 2013). One compelling feature of this data is that participants display multi-level inferences that suggest that cognitive biases operate on at least two separate levels:

- **Item-specific inferences**: Learners make first-order generalizations about the distributions governing individual items. This suggests that learners are equipped with inductive biases that lead to *regularization* of these distributions.

- **Language-wide inferences**: Learners draw inferences about the rules governing groups of items based on their commonalities. This suggests that learners exhibit second-order biases for *systematicity*.

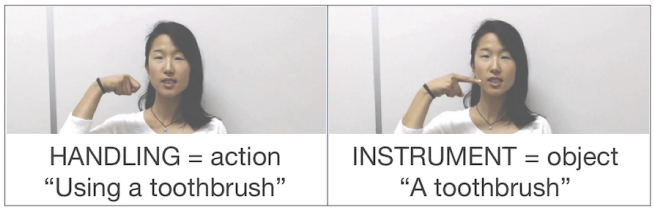

Figure 2: Two different gesture strategies, HANDLING and INSTRUMENT, used to refer to an action or an object. The figure depicts the congruent (iconic) mapping HANDLING to ACTION and INSTRUMENT to OBJECT.

To better understand the relationship between these two distinct levels, we formalize these biases in the context of a hierarchical Bayesian model (Kemp, Perfors, & Tenenbaum, 2007). Modeling results suggest that biases at each level can (to some extent) operate independently and must work in tandem to create systematicity at the language level. In particular, item-specific regularization alone is not sufficient to give rise to systematicity. Before describing the modeling in more detail, we will first present the artificial language learning experiment in the next section.

Gesture Learning Experiment

The data we model was obtained in an experiment studying the use of different gesture strategies for referring to hand-held tools (Padden et al., 2013; Padden, Hwang, Lepic, & Seegers, 2015; Ortega & Özyürek, 2016; Verhoef et al., 2016). Both in gesture and sign languages, two dominant strategies have been found: HANDLING (showing how you hold the tool) and INSTRUMENT (showing what the tool looks like), as shown in Figure 2. Sign languages differ in the relative frequencies of use of these strategies. Crucially, variation between forms is often conditioned on the intended meaning. American sign language (ASL) signers, for instance, prefer to map HANDLING forms to ACTION concepts and INSTRUMENT forms to OBJECT concepts (Padden et al., 2013). These same mapping preferences can be found in non-signing gesturers (Verhoef et al., 2016). The experiment explored the influence of such prior mapping preferences on systems created by participants in an artificial gesture learning experiment.

In the experiment, HANDLING and INSTRUMENT strategies were probabilistically paired with OBJECT or ACTION concepts for 9 different hand-held tools (e.g., hammer, toothbrush, mascara, etc.). 80 participants were evenly split across four conditions whose input languages varied in their mappings between gesture strategies and concepts:

- **Congruent.** Each concept in the training phase is presented 75% of the times with the preferred mapping (see Figure 2), e.g., "using a toothbrush" appeared 6 out of 8 times with a video depicting HANDLING and 2 times with a video depicting INSTRUMENT; "a toothbrush" appeared 6 times with INSTRUMENT and 2 times with HANDLING.
- **Incongruent.** The preferred mapping (see previous bullet point) is used 25% of the times.
- **Handling.** The HANDLING strategy is used 75% of the times for all items, independent of concept type.
- **Instrument.** The INSTRUMENT strategy is used 75% of the times for all items, independent of concept type.

Each tool $\times$ concept combination was presented 8 times and people received feedback about the correct gesture strategy. After training, participants were asked to select the correct gesture for each tool $\times$ concept combination another 8 times, this time without feedback. Figure 1A shows the experimental results and indicates how participants' output deviates from their respective input, indicated by one of the four target labels ($\bigoplus$). The axes show the proportion of HANDLING used for ACTION concepts (x-axis) and HANDLING used for OBJECT concepts (y-axis). Each data point shows the output language of a single participant averaged across the 9 items. To disambiguate between participants in the center of the plot that produced near-random responses for every item and participants that produced highly deterministic yet different responses for different items, the size of each data point shows the the variance in participants' use of different gesture strategies across items.

Figure 1A shows that only in the congruent condition a majority of subjects consistently extended the input pattern and produced languages that were more deterministic than the input, that is, regularization that lead to systematicity occurred. Subjects in the other conditions either regularized towards the direction of the input mapping, or towards other mappings, while they seem to be most strongly drawn in the direction of congruent mappings. This suggests that, while participants favored congruent mappings overall, they vary in their sensitivity to the input condition.

Participants also differ in their inferences about systematicity. Most participants produced systems in which tools are regularized in a single coherent way (data points close to the corners), but some participants didn't have a single preferred direction of regularization (high variance data points in the center region). While the latter similarly show strong item-level regularization behavior, they fail to show second order regularization (inferences about the kinds of gesture strategies used for different tools).[1] However, overall levels of regularization are high, as indicated by Figure 1B, which shows the distribution over test data before averaging by participant.

Bayesian Learning Model

To better understand how cognitive biases at multiple levels interact to give rise to systematicity, we formalize the experiment presented above as a computational model. Because we're interested in quantitatively assessing the effect of cognitive biases on behavior in a probabilistic setting, we follow

[1]These participants are importantly distinct from low variance participants in the center region, which fail to show either first or second order regularization.

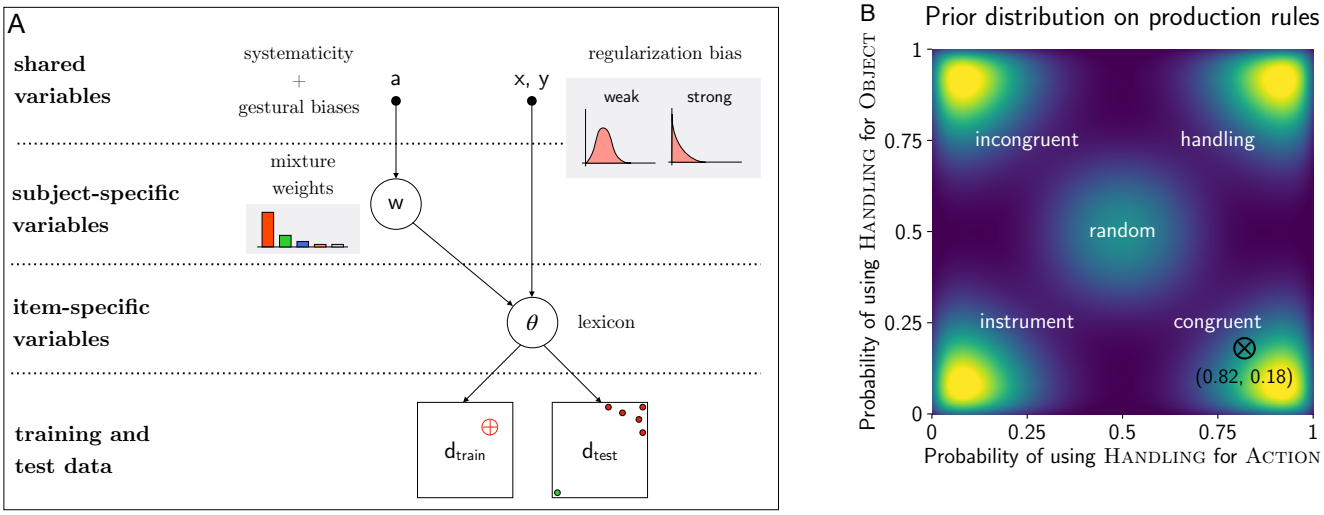

Figure 3: **A**. The hierarchical Bayesian model and its various levels of shared and independent structure. Two model variations are derived from the full model depicted here: In the no item-specific inferences model, the lexicon θ is no longer item-specific and thus moved up next to the mixture weights w. In the no language-wide inferences model, the mixture weights w node is shared across all subjects instead of subject-specific and thus moved up next to parameters a, x and y. **B**. The hypothesis space for a single production rule θ of the lexicon. Here we depict a prior distribution over θ with equal mixture weights $w = (0.2, 0.2, 0.2, 0.2, 0.2)$ and with regularization strengths $x = 11, y = 2$ and random component $r = 8$. Brighter colors correspond to higher probability regions.

prior work and treat language learning as inference in a statistical model (Kirby et al., 2015; Culbertson & Smolensky, 2012). This model, describing inferences performed from the perspective of the *subject*, must be distinguished from the model used by the *scientist* to quantify uncertainty about the parameters involved in the subject's inference. From the perspective of the subject, learning consists of observing data from one of the four conditions under the influence of prior assumptions about how the data was generated (inductive biases). After learning, test phase data is generated from the subject's updated model. The scientist observes both training and test data and infers which prior biases are most likely responsible for the observed patterns. The model we will present has important connections to the model of word order preferences developed by Culbertson and Smolensky (2012), which can be understood as a special case of our model.

Informal description of the model

Both training and test data are modeled as counts drawn from a 2-dimensional Binomial distribution with parameter $\boldsymbol{\theta} = (\theta_a, \theta_o)$, the probability of generating a HANDLING form for the ACTION meaning and for the OBJECT meaning, respectively. We will refer to a set of these rules (one for each item) as a *lexicon*. Rules can be visualized as a point in a 2-dimensional space (see Figure 3B). Learning for the subject consists of observing training counts $\boldsymbol{d} = (d_a, d_o)$ for each tool and computing a posterior distribution over lexicons $P(\boldsymbol{\theta}|\boldsymbol{d})$. According to Bayes' rule, this distribution is proportional to the product of the prior probability of a lexicon $P(\boldsymbol{\theta})$ and the (Binomial) likelihood of the data under that lexicon $P(\boldsymbol{d}|\boldsymbol{\theta})$. After learning, subjects produce test phase data by sampling a lexicon from their posterior distribution and subsequently generating data using that lexicon.

The prior distribution is conditioned on cognitively-interpretable parameters w_s, a, x, y, which represent different *cognitive biases* that act on learning: w_s represents a subject's individual preference for certain form-meaning mappings (e.g., congruent, handling, etc.) over others, a represents subjects' overall bias for systematicity across form-meaning mappings, and parameters x and y encode the strength of subjects' item-level regularization biases. From the perspective of the researcher, placing prior distributions on these conditioned variables allows us to infer subjects' inductive biases by way of how different biases would manifests in behavior. This yields the full probabilistic model depicted in Figure 3A.

Item-level (first-order) inferences Following previous work (Culbertson & Smolensky, 2012) we assume that knowledge about the lexicon is expressed in terms of a mixture distribution (see Figure 3B). We use four symmetric components plus one random component, each corresponding to a canonical mapping strategy as described in Table 1. Components are modeled as Beta distributions with different permutations of the model parameters x and y, whose values govern the strength of the *regularization bias* in the model.

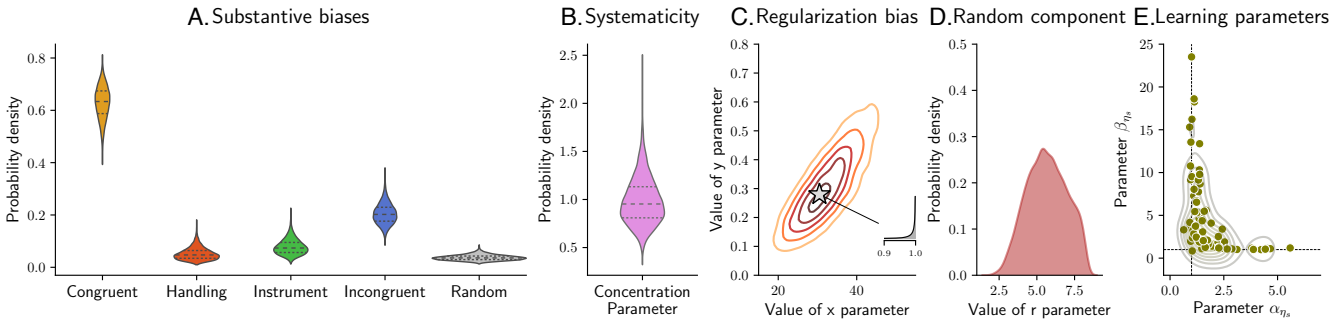

Figure 4: Results for the full model (see text, from left to right): The distribution over a, the prior parameter for the Dirichlet distribution, is split into a base vector and a concentration parameter. **A**. The base vector shows the direction of substantive biases. **B**. The concentration parameter shows that the model has learned a sparse prior over mixture weights. **C**. Estimates for the regularization bias show very sparse Beta distributions, i.e., strong prior biases for regularization. **D**. Strength of random component r. **E**. Best fitting values for subjects' learning parameters.

Table 1: Parameters of the mixture components. Each component is the product of two independent beta distributions of the form: $\text{Beta}(\theta_a|\alpha_a,\beta_a)\text{Beta}(\theta_o|\alpha_o\beta_o)$. Our model also contains the identifiability constraints $x > y$, as well as the constraint that $x \geq 1$.

Mixture component	α_a	β_a	α_o	β_o
Congruent	x	y	y	x
Handling	x	y	x	y
Instrument	y	x	y	x
Incongruent	y	x	x	y
Random	r	r	r	r

Language-wide (second-order) inferences Lexicons (i.e., a set of production rules, one rule per item) are sampled from this mixture distribution according to a probability vector w_s, which is unique to each subject s. While this mixture weight parameter encodes information about statistical tendencies in a subject's lexicon, placing a distribution on w_s allows us to expresses second-order generalizations about lexicons across subjects. Since w_s is a probability vector, this prior takes the form of a Dirichlet distribution (a multivariate generalization of the Beta distribution), parametrized using a vector a. Analogous to the parameters of a Beta distribution, a is able to express information about the expected composition of a lexicon, as well its systematicity (i.e., whether sampled mixture weights tend to be more uniform or more extreme). To this end, a can be decomposed into a scalar concentration parameter $\gamma = \sum_{i=1}^{N} a_i$, governing *systematicity* and a vector with entries $b_i = a_i / \sum_{i=j}^{N} a_j$, which encodes an overall *mapping preferences* for different mixture components (substantive biases).

Learning from the subject's perspective

Before learning about subject's inductive biases from the scientist's perspective, we must first obtain the subject's posterior distribution over lexicons given the observed training data $P(\boldsymbol{\theta}|\boldsymbol{d}) \propto P(\boldsymbol{d}|\boldsymbol{\theta}) \times P(\boldsymbol{\theta})$. Inference is tractable because the product of the prior $P(\boldsymbol{\theta})$, a mixture of Beta distributions, and the likelihood $P(\boldsymbol{d}|\boldsymbol{\theta})$, a Binomial distribution, can be expressed in closed form, which allows us to perform the update analytically. Informally, we update each Beta distribution of the mixture as if it had generated all the observed data. The mixture weight for each component is then updated in proportion to how well it predicted the data relative to the other components.

One important addition to the model is the use of a subject-specific learning parameter η_s, which allows the model to show graded regularization behavior in response to data (Meylan, Frank, & Levy, 2013). This learning parameter effectively operates as a discounting factor on the data. If $\eta_s = 0.5$, for instance, instead of updating with the observed counts $k = 2, n = 8$, the model is updated using $k = 1, n = 4$. No update of the prior occurs with $\eta_s = 0$ while $\eta_s = 1$ corresponds to full Bayesian updating.

Learning from the researcher's perspective

To infer subjects' inductive biases concerning language-wide inferences (parameter a) and item-level inferences (parameters x, y, r), as well as the aforementioned learning rates η_s, uninformative prior distributions were placed on these parameters. The model is conditioned on the remaining, subject-produced test data $\boldsymbol{d}$. For this model and model variants described below, we computed sampling-based approximations of the posterior distribution using a NUTS sampler in the Python-based probabilistic programming language PyMC3 (Salvatier, Wiecki, & Fonnesbeck, 2016).

Item-level vs language-wide inferences

To better understand the distinct contribution of each level of inference and how they might jointly give rise to systematicity, we explore two variants of the model outlined so far (also see caption in Figure 3). The no-item-level-inferences model uses a single production rule for all items, thereby pre-

venting the model to generalize on an item-by-item basis. The no-language-wide-inferences model uses a single mixture proportion w, shared across subjects and thus precludes the model from learning and expressing generalizations about different languages. In this respect, the model developed by Culbertson and Smolensky (2012), which lacks both of these components, is an important baseline model. The models are evaluated in terms of how well simulated data from the models explains the patterns of systematic inferences implicit in the experimental data.

Modeling results

We first report results for the inductive biases that were estimated using the full model, after which we discuss how it compares with models that lack the ability to make item-specific or language-wide inferences.

Inductive biases of the full model

Figure 4 shows posterior estimates for the inductive biases acquired by the full model.

Mapping biases Both mapping biases and the concentration parameter, which we interpret as a systematicity bias, are derived from the posterior distribution over a, the parameter of the Dirichlet distribution generating mixture proportions. As Figure 4A indicates, subjects are strongly biased towards congruent gesture-meaning mappings, i.e., they overall prefer the iconic mapping INSTRUMENT to OBJECT and HANDLING to ACTION (see Figure 2). Moreover, subjects have an additional, slightly weaker, preference for incongruent mappings over handling and instrument mappings. Since the congruent and incongruent mappings are the only strategies that allow the system to express distinctions among objects and actions, this indicates that subjects have a strong overall bias towards informative systems. This is surprising since subjects' weren't explicitly penalized for collapsing distinctions, which is often necessary to prevent a loss of expressivity in artificial language learning experiments. Handling and instrument mappings were in turn weakly preferred over random mappings, suggesting an additional preference for structured and simple mappings. Mapping biases followed the order *congruent* > *incongruent* > *handling* or *instrument* > *random* in 94.5% of the samples from the posterior trace.

Systematicity The concentration parameter expresses second order regularization biases over the structure of the lexicon. Figure 4B shows that the model acquired a low concentration parameter that encodes an inductive bias for sparse lexicons (i.e., lexicons that are primarily composed of a single mapping strategy).

Item-level regularization bias To assess the strength of the regularization bias, we examine the posterior distribution over x and y, the prior parameters of the first four mixture components, where values were constrained such that $x > y$. Figure 4C suggests that the region of highest posterior probability corresponds to extremely sparse Beta distributions, resulting in a strong preference for near-deterministic production rules. To demonstrate this, we depict the regularization bias that corresponds to values of the mode of the posterior distribution, which shows a Beta distribution where most of the probability mass is located within the (0.99, 1.0) interval. The strength of this prior regularization bias is qualitatively similar to the bias reported in Culbertson and Smolensky (2012) ($x = 16.5, y = 0.001$).

Random component The posterior on r, the random component, is depicted in Figure 4D. Values correspond to a moderately peaked Beta distribution. The low mixture weight placed on the random component (see Figure 4A) suggests that this component is not instrumental in explaining subjects' inferences.

Learning parameters A learning parameter was included in our model to account for the observation that subjects generalized in ways not consistent with full updating based on the input. For each subject, we therefore estimated a learning parameter $\eta_s \sim \text{Beta}(\alpha_{\eta_s}, \beta_{\eta_s})$ and Figure 4E shows maximum a posteriori estimates (MAP) for each subject's α_{η_s} and β_{η_s}. Most subjects lie along the $\alpha_{\eta_s} = 1$ line, corresponding to varying degrees of strong learners, or the $\beta_{\eta_s} = 1$ line, corresponding to degrees of weaker learners, respectively.

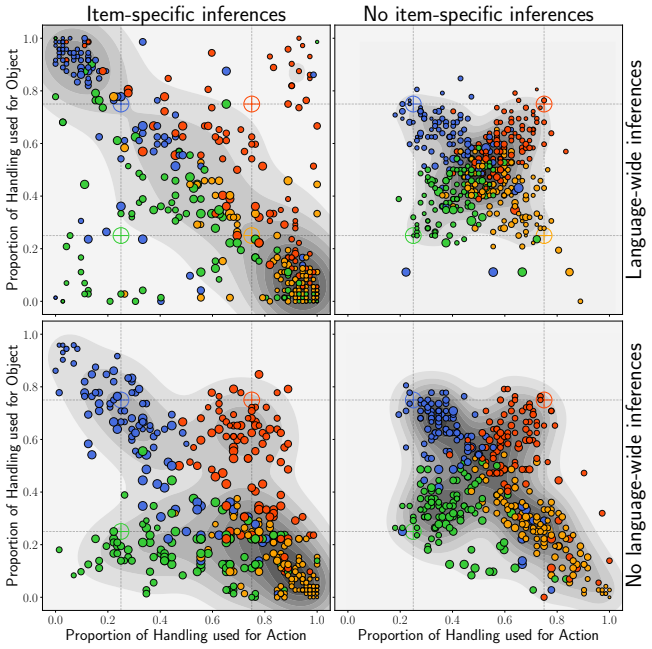

Figure 5: Simulated data from the full model, the no-item-specific-inferences model, the no-language-wide-inferences model, and a model lacking both components.

Contrasting item-specific vs language-wide inferences

The results so far demonstrate that our model acquired strong inductive biases for systematicity. To investigate which role item-specific and language-wide inferences play in acquiring

such biases, we compare *simulated data* from four different models: the full model, the no-item-specific-inferences model, the no-language-wide-inferences model and a model that lacks both inferential capabilities (see Figure 5). When asking whether a model acquires a bias for systematicity, it is important to note that all of the models are *in principle* capable of learning and expressing such a bias. Here we are instead interested in whether the models will *actually* acquire such a bias when conditioned on the experimental data, that is, when they are tasked with reproducing patterns that exist in the data under the structural constraints that their respective model architectures place on them.

Data was simulated by fixing the models' a, x, y, and r parameters to their respective MAP estimates given the experimental data and by randomly sampling learning parameters η_s and mixture weights w_s for 400 simulated subjects, 100 each per experimental condition. Simulated subjects observed the same training stimuli as subjects in the original experiment.

Figure 5 shows the distribution of simulated data for each of the four models. The full model (upper left corner) reproduces all key aspects of the data: regularization at the level of items and systematic generalizations at the language level commensurable with the experimental data (Figure 1). The model also reproduces overall preference for congruent form-meaning mappings observed in the experiment.

The **no-item-specific-inferences model** uses a single production rule for all items. Models that lack this component (right hand side of Figure 5) fail to produce systematic languages. As generalizations can differ on a per item basis but must here nevertheless be explained by a single rule, test data are "pulled" towards the center of the space. Coercing the model to fit to our data, which contains multi-modality at the item level, results in model failure. The no-item-specific inference constraint leads to mode collapse and, subsequently, to a failure to produce the kinds of systematic inferences observed in data.

The **no-language-wide-inferences model** (lower left corner), on the other hand, seems to exhibit some amount of systematicity. On closer examination, however, it becomes apparent that it predominantly reproduces patterns that are present in the input, while adding very little systematicity of its own by means of its inductive biases. While models with weak inductive biases are able to reproduce patterns that already exist in the input if given enough data, these patterns would not be sustainable in our current model and eventually disappear. More generally, experimental data can vary in the richness of patterns of systematicity that it exhibits. The fact that the model variants explored here are too constrained to capture patterns in the data suggests that the data exhibits high degrees of systematicity, which requires multi-level inferences such as exhibited by the full model.

Discussion

We modeled data from a recent artificial language learning experiment by Verhoef and colleagues (2016) to clarify the relationship between item-specific *regularization* and language-wide *systematicity* in domains where learners draw hierarchical inferences. We developed a Bayesian learning model that renders explicit the nature and interplay of cognitive biases that operate at these different levels.

Simulations from structurally different versions of the model showed that both item-specific and language-wide inferences are necessary to capture participants' behavior. While models that are prevented from forming item-specific (first-order) generalizations fail to capture regularization, models that don't allow language-wide (second-order) inferences fail to show systematicity. This supports the hypothesis that learners' inductive biases about form-meaning mappings are structured hierarchically.

One open question is whether cognitive biases at these two levels are an instance of a more general, low-level preference for simplicity (Chater & Vitányi, 2003) that manifests itself differently at each level, or whether they are biases with fundamentally distinct origin.

In future work, we intend to explore extensions of our model to other domains in the context of generalizations about items vs generalizations across items, such as Cornish, Smith, and Kirby (2013) and Cuskley (2019), or Smith and Wonnacott (2010). Hierarchical probabilistic models are a powerful tool for cognitive science because they allows us to flexibly express learners' inferences at multiple, interrelated levels (Kemp et al., 2007). We hope that some of the wider implications of this work will be promoting a multi-level approach to the study of inductive biases in artificial language learning and language evolution more broadly.

References

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in cognitive sciences*, *7*(1), 19–22.

Cornish, H., Smith, K., & Kirby, S. (2013). Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Culbertson, J., & Smolensky, P. (2012, November). A Bayesian Model of Biases in Artificial Language Learning: The Case of a Word-Order Universal. *Cognitive Science*, *36*(8), 1468–1498. doi: 10.1111/j.1551-6709.2012.01264.x

Cuskley, C. (2019). Alien forms for alien language: investigating novel form spaces in cultural evolution. *Palgrave Communications*, *5*(1), 1–15.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical bayesian models. *Developmental science*, *10*(3), 307–321.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, *28*, 108–114.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Meylan, S., Frank, M., & Levy, R. (2013). Modeling the development of determiner productivity in children's early speech. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Ortega, G., & Özyürek, A. (2016). Generalisable patterns of gesture distinguish semantic categories in communication without language. In *Proceedings of the 38th annual meeting of the cognitive science society* (p. 1182-1187). Austin, TX: Cognitive Science Society.

Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015, January). Tools for Language: Patterned Iconicity in Sign Language Nouns and Verbs. *Topics in Cognitive Science*, *7*(1), 81–94. doi: 10.1111/tops.12121

Padden, C., Meir, I., & Lepic, R. (2013, July). Patterned iconicity in sign language lexicons. *Gesture*, *13*(3), 287–308. doi: 10.1075/gest.13.3.03pad

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016, apr). Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, *2*, e55. doi: 10.7717/peerj-cs.55

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449.

Verhoef, T., Padden, C., & Kirby, S. (2016). Iconicity, naturalness and systematicity in the emergence of sign language structure. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference (evolangx11).*