# Effects of transmission perturbation in the cultural evolution of language

**Thomas Brochhagen (mail@mail.com)** [MF: fill me]
Institute for Logic, Language & Computation, University of Amsterdam

**Michael Franke (mchfranke@gmail.com)**
Department of Linguistics, University of Tübingen

## Abstract

Two main factors seem to play a major role in the cultural evolution of language. On the on hand, there is functional pressure towards efficient transfer of relevant information. On the other hand, languages have to be learned repeatedly and will therefore show traces of systematic stochastic perturbations operating during the transmission of linguistic knowledge. While a lot of attention has been paid to the effects of cognitive learning biases on the transmission of language, there is reason to expect that the class of possibly relevant transmission biases is much larger. This paper therefore explores some potential effects of transmission noise due to errors in the observation of states of the world. We look at three case studies on (i) vagueness, (ii) meaning deflation, and (iii) underspecified lexical meaning. These case studies suggest that transmission perturbations other than learning biases might help xplain attested patterns in the cultural evolution of language and that transmission perturbations due to perceptual noise may even produce effects very similar to learning biases.

**Keywords:** cognitive biases; iterated learning; language evolution

## Introduction

Language is shaped by its use and transmission across generations. Linguistic properties therefore need not necessarily arise and stabilize solely due to functional pressure, such as the selection of more communicatively efficient behavior, but may also be influenced and selected for by a pressure for learnability. In the extreme, an unlearnable language will not make it to the next generation. The effects that (iterated) learning has on language are often seen as stemming from a combination of general learning mechanisms and inductive cognitive biases (e.g. Griffiths & Kalish 2007, Kirby et al. 2014, Tamariz & Kirby 2016). Proposals of biases that shape language acquisition abound. Some prominent examples are mutual exclusivity (Merriman & Bowman 1989, Clark 2009), simplicity (Kirby et al. 2015), regularization (Hudson Kam & Newport 2005), and generalization (Smith 2011). Forces other than learning biases may also systematically perturb the transmission of linguistic knowledge and therefore additionally contribute to the shaping of language by cultural evolution. In the following we focus on one particular source of transmission noise: agents' imperfect perception of the world. The overall goal of this paper is to give a formalism in which to study the effects of such perturbations and apply it to three case studies on (i) vagueness, (ii) meaning deflation, and (iii) underspecified lexical meaning.

## Iterated Bayesian learning

We model the transmission of linguistic knowledge as a process of iterated learning (for recent overviews see Kirby et al. 2014, Tamariz & Kirby 2016). More specifically, we focus on iterated Bayesian learning, in which a language learner must infer unobservables, such as the lexical meaning of a word, from the observable behavior of a single teacher, who is a proficient language user (e.g. Griffiths & Kalish 2007, Kirby et al. 2007). The learner observes instances $\langle s,m \rangle$ of overt language use in context, where $s$ is a world state and $m$ is the message that the teacher used in state $s$. The learner's task is to infer which latent type $t$ (e.g., which set of lexical meanings or which grammar) may have produced a sequence of such observations. To do so, the learner considers the posterior probability of $t$ given a data sequence $d$ of $\langle s,m \rangle$ pairs:

$$P(t \mid d) \propto P(t)\, P(d \mid t),$$

where $P(t)$ is the learner's prior for type $t$ and $P(d \mid t) = \prod_{\langle s,m \rangle \in d} P(m \mid s,t)$ is the likelihood of type $t$ producing the observed data $d$, with $P(m \mid s,t)$ the probability that a type $t$ produces message $m$ when in world state $s$. It is usually assumed that learners adopt a type with a probability $F(t \mid d) \propto P(t \mid d)^l$, where $l \geq 1$ regulates whether learners use probability matching ($l = 1$) or tend towards choosing a maximum of the posterior distribution ($l > 1$). If the set $D_k$ of possible data a learner may be exposed to is the set of all sequences with $k$ pairs $\langle s,m \rangle$, the probability that the learn a learner acquires type $t_i$ when learning from type $t_j$ is:

$$P(t_j \rightarrow t_i) \propto \sum_{d \in D_k} P(d \mid t_j) F(t_i \mid d).$$

If a population is a distribution over types, then iterated Bayesian learning predicts the most likely path of change in the population due to learning from finite observations.

The prior $P(t)$ can be understood as encoding learning biases. For example, learners may have a preference for simpler languages over ones with a more complex grammar, larger or more marked inventories, or cognitively taxing components [MF: what are "inventories" and "taxing components"? More precisely?] (c.f. Feldman 2000, Chater & Vitányi 2003, Kirby et al. 2015). Crucially, even weak biases can magnify and have striking effects on an evolving linguistic system, especially if learning is fueled by only limited input (small $k$). Experimental and mathematical investigations in iterated learning have therefore argued that the linguistic structure evinced by the outcome of this process reflects learners' inductive biases (Kirby et al. 2007; 2014).

## Iterated Bayesian learning with state-noise

Other stochastic factors beyond learning biases in $P(t)$ can influence the adoption of a linguistic type $t$ based on the
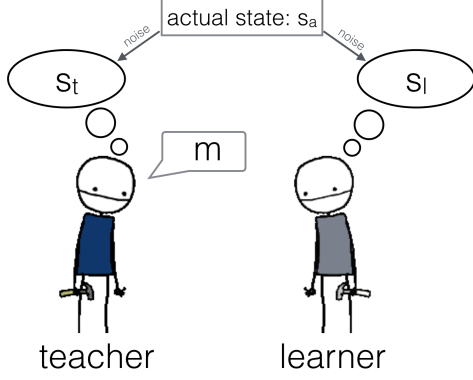
Figure 1: State-noise during observation of language use.

observation of $\langle s, m \rangle$ pairs. One further potential source of "transmission noise" are regular stochastic errors in the perception of world states (see Figure 1). Imperfect perception of world states may lead teachers to produce utterances that deviate from their production behavior had they witnessed the state correctly. Similarly, learners may mistake utterances as applying to different states than the ones witnessed by the teacher who produced them. For instance, when learning the meaning of a vague adjective such as *tall* from an utterance like "Jean is tall," agents may have diverging representations of how tall Jean actually is, even if she is in a shared perceptual environment. The main idea to be explored here is that regularities in misperceptions of states may have striking and possibly explanatory effects on language evolution.

Let $S$ be a set of world states. We denote the probability that the teacher (learner) observes state $s_t$ ($s_l$) when the actual state is $s_a$ as $P_N(s_t \mid s_a)$ ($P_N(s_l \mid s_a)$). The probability that $s_a$ is the actual state when the learner observes $s_l$ is therefore:

$$P_N(s_a \mid s_l) \propto P(s_a) \, P_N(s_l \mid s_a) \, .$$

Accordingly, the probability that the teacher observes $s_t$ when the learner observes $s_l$ is (assuming a finite state space):

$$P_N(s_t \mid s_l) = \sum_{s_a} P(s_a \mid s_l) \, P_N(s_t \mid s_a) \, .$$

The probability that a teacher of type $t$ produces data that is perceived by the learner as a sequence $d_l$ of $\langle s_l, m \rangle$ pairs is:

$$P_N(d_l \mid t) = \prod_{\langle s_l, m \rangle \in d_l} \sum_{s_t} P_N(s_t \mid s_l) \, P(m \mid s_t, t) \, .$$

It is natural to assume that learners, even if they (in tendency) perform rational Bayesian inference on the likely teacher type $t$ based on observation $\langle s_l, m \rangle$, do not also reason about state-noise perturbations. In this case the posterior probability of $t$ given the learner's perceived data sequence $d_l$ is as before: $P(t \mid d_l) \propto P(t) \, P(d_l \mid t)$. Still, state-noise affects the probability $P_N(t_j \to t_i)$ that the learner adopts $t_i$ given a teacher of type $t_j$, because state-noise influences the probability of observing

a sequence $d_l$ (with $F(t_i \mid d)$ as before):

$$P_N(t_j \to t_i) \propto \sum_{d \in D_k} P_N(d_l \mid t_j) F(t_i \mid d) \, .$$

In sum, it may be the case that learner and/or teacher do not perceive the actual state as what it is. They are not aware of this, and produce/learn as if what they observed was the actual state. The learner does not reason about noise when she tries to infer the speaker's type. She takes what she observes a state to be as the actual state that the teacher has seen as well and infers which type would have most likely generated the message to this state. This can lead to biases of inferring the "wrong" teacher type if the noise makes some types err in a way that resembles the noiseless behavior of other types. That is, such environmental factors can, in principle, induce transmission perturbations that look as if there was a cognitive bias in favor of a particular type, simply because that type better explains the noise.

## Case studies

In what follows we present three case studies that show how iterated learning under noisy perception of states can lead to the emergence of linguistic phenomena of interest. Case studies are ordered from more to less obvious examples in which state-noise may be influential and explanatory: (i) vagueness, (ii) meaning deflation, and (iii) underspecification in the lexicon. No case study is meant to suggest that state-noise is the definite and only explanation of the phenomenon in question. Our aim is to elucidate the role that transmission perturbations beyond inductive biases may play in shaping the cultural evolution of language and so we present minimal settings that isolate potential effects of state-noise in iterated learning.

### Vagueness

Many natural language expressions are notoriously vague and pose a challenge to logical analysis of meaning (e.g. Williamson 1994). Vagueness also challenges models of language evolution since functional pressure towards maximal information transfer should, under fairly general conditions, weed out vagueness (Lipman 2009). Many have therefore argued that vagueness is intrinsically useful for communication (e.g. van Deemter 2009, de Jaegher & van Rooij 2011, Blume & Board 2014). Others hold that vagueness arises naturally due to limits in perception, memory, or information processing (e.g. Franke et al. 2011, O'Connor 2014, Lassiter & Goodman 2015). We follow the latter line of exploration here, arguing that vagueness arises naturally under imperfect observability of states (see Franke & Correia (to appear) for a different evolutionary dynamic based on the same idea).

**Setup.** We analyze the effects of noisy perception on the transmission of a simple language with 100 states, $s \in [0, 99]$, and two messages, $m \in \{m_1, m_2\}$. The probability of perceiving the actual state $s_a$ as $s_p$ is given by a (discretized) normal distribution, truncated to $[0; 99]$, with the actual state as its

mean, a standard deviation $\sigma$. Linguistic behavior is determined by a type $t \in [0; 99]$ which encodes the threshold of applicability of $m_1$: $P(m_1 \mid s, t) = \delta_{s \geq t}$. In words, if a speaker observes a state that is as large or larger than its type (threshold for $m_1$), then message $m_1$, e.g., *tall*, is used; otherwise $m_2$, e.g., *small*.

**Results.** The effects of a single generational turnover under noisy transmission is depicted in Figure 2b. As shown in Figure 2a, this population initially consisted exclusively of type $t = 50$. [TB: Do you think it might be confusing to start speaking about populations here? We now don't have a paragraph mentioning what evolves explicitly] [MF: inserted talk of population (and definition) earlier. does this fix it?] As learners try to infer this type from observed language use, even small $\sigma$ will lead to the emergence of vagueness in the sense that there is no longer a crips and determinate cut-off point for message use in the population, but *borderline regions* in which $m_1$ and $m_2$ are used almost interchangeably. For larger $\sigma$, larger borderline regions ensue. The size of borderline regions further increases over generations with growth inversely related to $l$ and $k$. As is to be expected, if $k$ is too small to discern even strikingly different types, then iterated learning under noisy perception leads to heterogeneous populations with (almost) no state being (almost) exclusively associated with $m_1$ or $m_2$.

**Discussion.** Transmission perturbations caused by the noisy perception of states reliably give rise to vague language use even if the initial population state had a perfectly crisp and uniform convention. Of course, this is a specific picture of vagueness. As modeled here for simplicity, each speaker has a fixed and non-vague cut-off point $t$ is her lexicon. Still, the production behavior of a type-$t$ speaker in actual state $s_a$ is probabilistic and "vague", because of noisy perception:

$$P_N(m \mid s_a, t) = \sum_{s_p} P(s_P \mid s_a) P(m \mid s_p, t).$$

An extension towards types as distributions over thresholds is straightforward but the main point would remain: systematic state-noise perturbs a population towards vagueness.

Of course, the stabilization of a linguistic system or population on a particular vague/clear state partition may reasonably be expected to depend not only on the effects of learning, but also on the functional (dis)advantages that this partition brings about for its users. Functional pressure may therefore well be necessary for borderline regions to be kept in check, so to speak. Which factor or combination thereof plays a more central role for the emergence of vagueness is an empirical question we do not address here. Instead, we see these results as adding strength to the argument that one way in which vagueness may arise is as a byproduct of interactions between agents that may occasionally err in their perception of the environment.
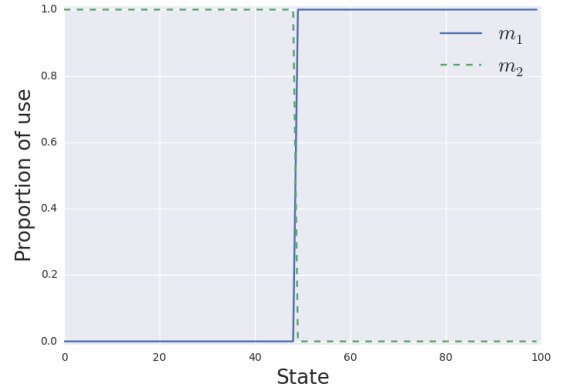


(a) Initial non-vague population.
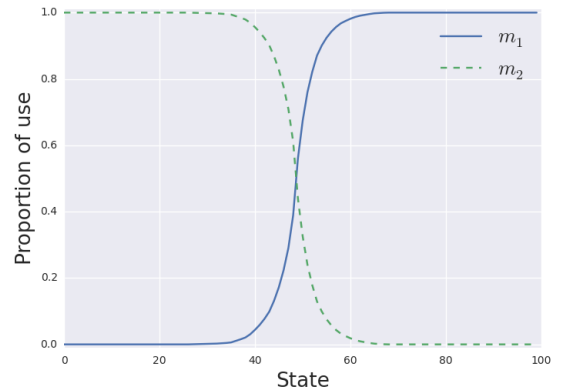


(b) Vague population after single generation.

Figure 2: Noisy iterated learning with posterior sampling, $\sigma = 0.4$, and $k = 20$. [MF: if we wanted to save space here, we could have just one plot with only $m_1$ once "initial state" and then "after one learning step"]

## Deflation

Meaning deflation is a diachronic process by which a form's once restricted range of applicability broadens. Perhaps the most prominent example is Jespersen's cycle (Dahl 1979), the process by which emphatic negation, such as French *ne ... pas*, broadens over time and instead becomes a marker for standard negation. As argued by Bolinger (1981), certain word classes are particularly prone to slight and unnoticed reinterpretation. Consequently, when retrieving their meaning from contextual cues, learners may continuously spread their meaning out. For instance, Bolinger discusses how the indefinite quantifier *several* has progressively shifted from meaning *a respectable number* to broader *a few* in American English. We follow this line of reasoning and show how state confusability may lead to meaning deflation. Other formal models of deflationary processes in language change have rather stressed the role of conflicting interests between interlocutors (Ahern & Clark 2014) or asymmetries in production frequencies during learning (Schaden 2012, Deo 2015).

**Setup.** The setup for this case study is the same as for the previous, except that we now trace the change of a single message $m$, e.g., emphatic negation, without a fixed antonym which is send whenever $m$ does not apply. This leads to a situation in which learners observe positive examples of use $\langle s, m \rangle$ but do not positively observe situations in which $m$ did not apply to a particular state $s$. This causes asymmetry in the learning data as speakers with higher thresholds will reserve their message only for a small subset of the state space and otherwise remain silent. Learners take the absence of observations into account but cannot know what it is that they did not observe. For simplicity, we assume that learners are aware of $k$ so that:[1]

$$P(t|d_l) \propto \text{Binom}(\text{successes} = k - |d_l|, \text{trials} = k,$$

$$\text{succ.prob} = \sum_{i=0}^{t-1} P(s = i)) \prod_{s \in d_l} P(m|s, t).$$

As before, the former factor corresponds to the likelihood of a type producing the perceived data. The latter is the probability of a type not reporting $k - |d|$ events for a total of $k$ events. $P \in \Delta(S)$ is assumed to be uniform. In words, a long sequence of data consisting of mostly silence gives stronger evidence for the type producing it having a high index even if the few state-message pairs observed in the sequence may be equally likely to be produced by types with lower indices.

**Results.** The development of an initially monomorphic population consisting only of $t = 80$ is shown in Figure 3. In this setup even little noise will cause a message to gradually be applied to larger portions of the state space. As above, the speed by which meaning deflates is regulated by $\sigma$, $k$, and to lesser degree $l$. In general, more state confusion due to higher $\sigma$, shorter sequences, or less posterior maximization will lead to more learners inferring lower types than present in the previous generation.

**Discussion.** In contrast to the previous case study, we now considered the effects of noisy perception under asymmetric data generation where overt linguistic evidence is not always produced. This setup can be likened to acquisition only from positive linguistic evidence in a world in which not every state is equally likely to lead to an observable utterance.

The overall result is similar to that of the previous study. Noisy perception can cause transmission perturbations that relax once strict linguistic conventions. In contrast to the case of vagueness, if there are no competing forms, e.g., *small* vs. *tall*, asymmetry in production and noise will iteratively increase the state space that a form carves out. Just as the overuse of a word or difficulties in the retrieval of its meaning

---

[1]Knowing fixed $k$ allows learners to compute the likelihood of a type not reporting $k - |d_l|$ state observations. A more involved but better justified alternative is to specify a prior over $k$ and for learners to perform a joint inference on $k$ and the teacher's type given a partial observation of only overtly produced data. For simplicity, we opt for the former, albeit admittedly artificial, assumption.
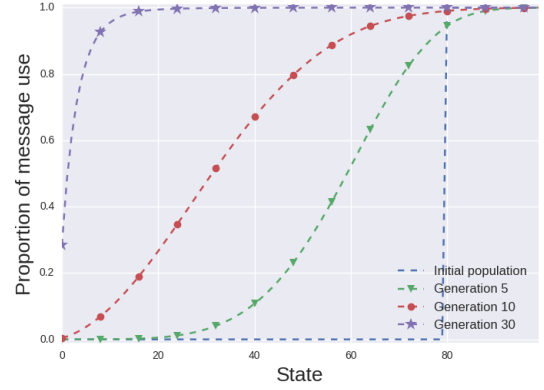


Figure 3: Noisy iterated learning with posterior sampling, $\sigma = 0.4$, and $k = 30$.

from contextual cues may lead to the deflation of its meaning in natural language. [MF: last sentence not well formed and not clear to me what it should mean]

### Scalar expressions

Scalar expressions have been at the center of many studies on pragmatic inference. Examples include quantifiers such as *some* and *most*, adjectives such as *cold* and *big*, and numerals such as *four* and *ten*. Commonly, their use is taken to pragmatically convey an upper-bound which is not hard-coded in the lexical semantics (Horn 1972, Gazdar 1979). For instance, while "Bo ate some of the cookies" is semantically compatible with a state in which Bo ate all of them, this utterance may convey that Bo ate *some but not all*, as otherwise the speaker would have used *all*. A semantically weak meaning is thus pragmatically strengthened by interlocutors' mutual reasoning about rational language use (Grice 1975). Why does such pragmatic strengthening not lead to wide-spread lexicalization of upper-bounded meanings? Brochhagen et al. (2016) explore an evolutionary model that combines functional pressure and iterated learning. Crucially, this account requires the assumption of a prior that favors a lack of upper-bounds. Here, we demonstrate that state-noise can mimic the effects of such a cognitive learning bias.

**Setup.** The simplest possible model distinguishes two kinds of lexica and two pragmatic usage patterns, a pair of which constitutes a type. Both lexica specify the truth-conditions of two messages in either of two states. Let us mnemonically label them $m_{\text{some}}$, $m_{\text{all}}$, $s_{\exists\neg\forall}$ and $s_\forall$, where the former state is one in which natural language *some but not all* holds, and the latter one where *all* holds. In lexicon $L_{\text{bound}}$ message $m_{\text{some}}$ is only true of $s_{\exists\neg\forall}$ and $m_{\text{all}}$ only of $s_\forall$. In English-like lexicon $L_{\text{lack}}$, message $m_{\text{all}}$ is also only true of $s_\forall$, but the meaning of $m_{\text{some}}$ is underspecified and lexically holds in both states. Speakers may use either lexicon either literally or pragmatically. The former selects any of several true messages with equal probability, the latter prefers to send the most informa-

tive messages from those that are true in the observed state (Grice 1975). This gives rise to probabilistic speaker behavior $P(m \mid s, t = \langle \text{lexicon}, \text{use} \rangle)$ which approximates the following:[2]

$$
\begin{array}{c}
\underline{L_{\text{bound}}} \\
\begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array}
\end{array}
\qquad
\begin{array}{c}
\underline{L_{\text{lack}}} \\
\begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array}
\end{array}
$$

$$
\text{Literal} \quad
\begin{array}{c} s_{\forall} \\ s_{\exists \neg \forall} \end{array}
\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
\qquad
\begin{array}{c} s_{\forall} \\ s_{\exists \neg \forall} \end{array}
\begin{pmatrix} 0.5 & 0.5 \\ 1 & 0 \end{pmatrix}
$$

$$
\begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array}
\qquad\qquad
\begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array}
$$

$$
\text{Pragmatic} \quad
\begin{array}{c} s_{\forall} \\ s_{\exists \neg \forall} \end{array}
\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
\qquad
\begin{array}{c} s_{\forall} \\ s_{\exists \neg \forall} \end{array}
\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
$$

As pragmatic users of $L_{\text{lack}}$ are (almost) indistinguishable from types with $L_{\text{bound}}$ in terms of their behavior, the emergence of predominance of $L_{\text{lack}}$ in a repeatedly learning population must come from transmission biases. A learning bias in favor of $L_{\text{lack}}$ in the learners' priors will select for $L_{\text{lack}}$ (Brochhagen et al. 2016), but here we assume no such cognitive bias. Rather we assume state-noise in the form of parameters $\varepsilon$ and $\delta$. The former corresponds to the probability of perceiving the actual state $s_{\exists \neg \forall}$ as $s_{\forall}$, $P(s_{\forall}|s_{\exists \neg \forall}) = \varepsilon$, and $P(s_{\exists \neg \forall}|s_{\forall}) = \delta$. [MF: let's try to motivate this as best as we can in one sentence?]

**Results.** To quantify the effects of the dynamics we ran a fine-grained parameter sweep over $\varepsilon$ and $\delta$ with 50 independent simulations per parameter configuration. Each simulation started with a random initial population distribution over types and applied iterated Bayesian learning with state-noise for 20 generations. The mean proportion of resulting pragmatic users of $L_{\text{lack}}$ under different noise signatures is shown in Figure 4. These results show that when $\delta$ is small and $\varepsilon$ is high, iterated noisy transmission can lead to populations consisting of mostly, if not exclusively, a type that does not lexicalize an upper-bound for *some*-like expressions but conveys it pragmatically. Similar results are obtained for increments in $k, l$.

**Discussion.** The main goal of this case study was to show that noisy perception may mimic effects of learning biases. In the case of Brochhagen et al. the assumed bias was one for simplicity. Accordingly, learners had an a priori preference for not codifying an upper-bound lexically. This increases the learners' propensity to infer pragmatic $L_{\text{lack}}$ over $L_{\text{bound}}$ even if the data witnessed could not tease them apart. Here, we assumed no such bias but nevertheless arrived at an evolutionary outcome that is comparable to the one predicted if the

---

[2]Concretely, our results are based on probabilistic speaker behavior following the definitions of Brochhagen et al. (2016) under parameter values that approximate the above very closely. As nothing essential to our main argument and simulation results hinges on these details we background them here for ease of exposition. [MF: So, yes, I'm radically throwing out $\lambda$ as well. I hope that's okay and I'm not missing anything. But the text before was pretty hard to digest.]
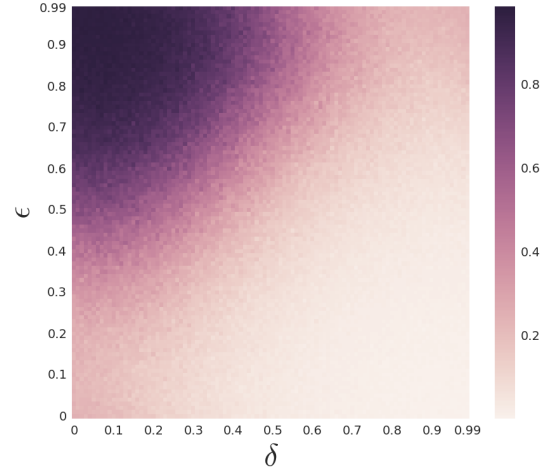


Figure 4: Mean proportion of pragmatic $L_{\text{lack}}$ users after 20 generations with posterior sampling, $k = 5$, and $\lambda = 20$.

bias were present. Note however that this outcome strongly depends on the types involved. Whether a type thrives under a particular noise signature depends on the proportion of types confused with it during transmission. The addition or extraction of a single type may therefore lead to different results.

At present, it is unclear what role noisy perception should play in the selection of underspecified meaning. These results should therefore be taken as suggestive but not indicative of a relationship between the two. A possible way to explore this relation may lie in their connection to empirical work on the verification of quantified statements (see Szymanik 2016 for a recent overview). The idea being that some states are easier to verify, e.g., $s_{\forall}$, and therefore less confusable with other states than others, e.g., $s_{\exists \neg \forall}$.

## General discussion

We proposed a model of iterated Bayesian learning that integrates systematic noise in agents' perception of world states, giving rise to stochastic perturbations that may influence and (potentially, partially) explain language change. We investigated the model's predictions in three case studies that show that iterated noisy transmission can lead to outcomes akin to those found in natural language. As stressed before, these results are not meant to suggest noisy perception to be the sole or main determinant of these phenomena. Instead, our aim was mainly conceptual and technical in nature.

Beyond technical aspects, we foregrounded two intertwined issues in the cultural evolution of language. First, the fact that noise signatures may mimic the effects of cognitive biases has consequences for the interpretation of the outcomes of acquisition processes. Care must therefore be exercised in reading off the influence of possible cognitive learning biases from data obtained "in the wild" or the laboratory. Second, and more importantly, these results may be seen as complementing and stressing the pivotal role of sys-

tematic transmission perturbations as explanatory and predictive devices of language change – independent of the perturbation's source. They thereby strengthen and widen the scope of research on iterated learning by bringing attention to forces beyond inductive biases.

## Conclusion

Acquisition is a central force shaping linguistic structure. The consideration of the (imperfect) means by which such knowledge is transmitted is therefore crucial to our understanding of the cultural evolution of language. Here, we focused on one factor that may give rise to systematic stochastic perturbation in learning —agents' noisy perception of the world— and analyzed its effects in three case studies on (i) vagueness, (ii) meaning deflation, and (iii) underspecified lexical meaning. Our results suggest that the class of relevant perturbation sources reaches beyond the well-studied effects of inductive learning biases. In particular, that some linguistic properties, such as (i), (ii) and more tentatively (iii), may emerge as a byproduct of constraints on agents' perception of the world.

## Acknowledgments

## References

Ahern, C., & Clark, R. (2014). Diachronic processes in language as signaling under conflicting interests. In E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *The evolution of language: Proceedings of 110textsuperscriptth international conference (evolang10)* (pp. 25–32). Singapore: World Scientific Press.

Blume, A., & Board, O. (2014). Intentional vagueness. *Erkenntnis*, *79*(4), 855-899.

Bolinger, D. (1981). The deflation of several. *Journal of English Linguistics*, *15*(1), 1–3.

Brochhagen, T., Franke, M., & van Rooij, R. (2016). Learning biases may prevent lexicalization of pragmatic inferences: a case study combining iterated (bayesian) learning and functional selection. In *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2081–2086). Austin, TX: Cognitive Science Society.

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

Clark, E. V. (2009). Lexical meaning. In E. L. Bavin (Ed.), *The cambridge handbook of child language* (pp. 283–300). Cambridge University Press.

Dahl, Ö. (1979). Typology of sentence negation. *Linguistics*, *17*(1-2).

Deo, A. (2015). The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics & Pragmatics*, *8*(1), 1–52. doi: 10.3765/sp.8.14

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Franke, M., & Correia, J. P. (to appear). Vagueness and imprecise imitation in signalling games. *British Journal for the Philosophy of Science*.

Franke, M., & Jäger, G. (2014). Pragmatic back-and-forth reasoning. *Semantics, Pragmatics and the Case of Scalar Implicatures.*, 170–200.

Franke, M., Jäger, G., & van Rooij, R. (2011). Vagueness, signaling & bounded rationality. In T. Onoda, D. Bekki, & E. McCready (Eds.), *Jsai-isai 2010* (pp. 45–59). Springer.

Gazdar, G. (1979). *Pragmatics, implicature, presupposition and logical form*. New York: Academic Press.

Grice, P. (1975). Logic and conversation. In *Studies in the ways of words* (pp. 22–40). Cambridge, MA: Harvard University Press.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, *31*(3), 441–480.

Horn, L. R. (1972). *On the semantic properties of logical operators in english*. Bloomington, IN: Indiana University Linguistics Club.

Hudson Kam, C. L., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151–195.

de Jaegher, K., & van Rooij, R. (2011). Strategic vagueness, and appropriate contexts. In *Language, games, and evolution* (pp. 40–59). Berlin, Heidelberg: Springer.

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.

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.

Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a bayesian model of interpretation. *Synthese*.

Lipman, B. L. (2009). *Why is language vague?* (Manuscript, Boston University)

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3/4), i-129.

O'Connor, C. (2014). The evolution of vagueness. *Erkenntnis*, *79*(4), 707–727.

Schaden, G. (2012). Modelling the "aoristic drift of the present perfect" as inflation: An essay in historical pragmatics. *International Review of Pragmatics*, *4*, 261–292. doi: 10.1163/18773109-00040207

Smith, K. (2011). Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Human Biology*, *83*(2), 261–278.

Szymanik, J. (2016). *Quantifiers and cognition: Logical and computational perspectives*. Springer International Publishing.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

van Deemter, K. (2009). Utility and language generation: The case of vagueness. *Journal of Philosophical Logic*, *38*(6), 607–632.

Williamson, T. (1994). *Vagueness*. London and New York: Routledge.