# Effects of transmission perturbation in the cultural evolution of language

**Name Surname (mail@mail.com)**
Department, street & number
city, ZIP country

**Name Surname (mail@mail.com)**
Department, street & number
city, ZIP country

## 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 explain 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:** noise; cognitive biases; iterated learning; cultural 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 ways of behavior, but may also be influenced and selected for by a pressure for learnability. To put it extremely, 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, O'Connor 2015).[1] But there is good reason to expect that forces other than learning biases may systematically perturb the transmission of linguistic knowledge and so additionally contribute to the shaping of language by cultural evolution. [MF: could expand here on the type of noise we focus on?] The goal of this paper is to give a formalism in which to study the effects of some such perturbations and apply it to three case studies on (i) vagueness, (ii) meaning deflation, and (iii) underspecified lexical meaning.

[1] Depending on their formulation and the domain(s) they are proposed to apply to, biases may also interact. For instance, a domain-independent bias for simplicity may entail regularization but stand in conflict with mutual exclusivity.

## 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, may have produced a sequence of such observations. To do so, the learner considers the posterior probability of type $t$ 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 $s$. Models of iterated Bayesian learning usually assume that the learner adopts a type with a probability proportional to $P(t \mid d)^l$, where $l \geq 1$ is a parameter that regulates whether learners use probability matching ($l = 1$) or show tendency towards choosing a maximum of the posterior distribution ($l > 1$).

The set of possible data a learner may be exposed to is the set $D_k$ of all sequences $d$ with length $k$ of $\langle s,m \rangle$ pairs. The lower $k$, the less information the learner has to recover the true type of the teacher. Mutation matrix $Q$ captures the probability $Q_{ji}$ that a learner acquires type $i$ when learning from type $j$:

$$Q_{ji} \propto \sum_{d \in D_k} P(d \mid t_j) F(t_i \mid d)\,, \text{where}$$
$$F(t_i \mid d) \propto [P(t_i) P(d \mid t_i)]^l\,,$$

The prior $P(t)$ can be understood as a encoding a learning bias. 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 (c.f. Feldman 2000, Chater & Vitányi 2003, Kirby

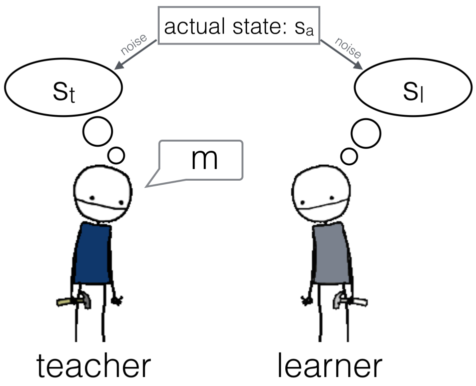



Figure 1: State-noise during observation of language use

et al. 2015). Crucially, even weak biases can magnify and have striking effects on an evolving linguistic system. Experimental and mathematical investigations in iterated learning have therefore argued that the linguistic structure evinced by the outcome of this process reflect learners' inductive biases (Kirby et al. 2007; 2014). The role of such biases can be viewed as that of introducing systematic perturbations in the transmission of linguistic knowledge, guiding learners to the convergence on particular evolutionary outcomes.

## 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 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). Agents may not always perceive a world state perfectly. If 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 a different state than the one witnessed by the parent who produced it. Imperfect perception of world states is nothing unnatural. For instance, when learning the meaning of a vague adjective such as *tall* from an utterance like "John is tall," agents may have diverging representations of how tall John actually is, even if he is in the shared perceptual environment. The main idea to be explored here is that regularities in state-misperceptions may have striking and possibly explanatory effects on language evolution.

Let $S$ be a set of states of affairs or meanings. 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:

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

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 that 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 $Q_{ji}$ that the learner adopts type $i$ given that the teacher is of type $j$, because it influences the probability of observing a sequence $d_l$:

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

where $F(t_i \mid d)$ is as before.

Noise free iterated Bayesian learning is obtained as a special case when the perceived state is always the actual state.

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. In particular, 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 biases 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 evinced in natural language. Case studies are ordered from more to less obvious examples in which state-noise may help explain phenomena of interest: (i) vagueness, (ii) meaning deflation, (iii) underspecification in the lexicon. No case study is meant to suggest that state-noise is the definite answer to the question of how these properties arose. Instead, we restrict our attention to minimal settings that deliberately abstract away from aspects not required for our present aim, which is to elucidate the role that transmission perturbations beyond inductive biases may play in shaping the cultural evolution of language.

Note also that constructing the set of learning data $D$ is computationally intractable for large $k$. We therefore approximate $D$ by sampling data from the production behavior of types. The values chosen correspond to experimentally determined amounts that minimize the effects that insufficient sampling may otherwise introduce. [MF: I would probably skip this. Technical details are backgrounded here anyway, and have to be.]

## Vagueness

Many if not most expressions in natural language are vague. Vagueness should be distinguished from imprecision and genuine ambiguity. The hallmark of a vagueness is susceptibility to a Sorites paradox (e.g. Williamson 1994): if a car for one million US$ is expensive and if a car that costs a single dolar less than an expensive car is still expensive, then also a car for one US$ is expensive. Clearly a fallacious reasoning pattern.

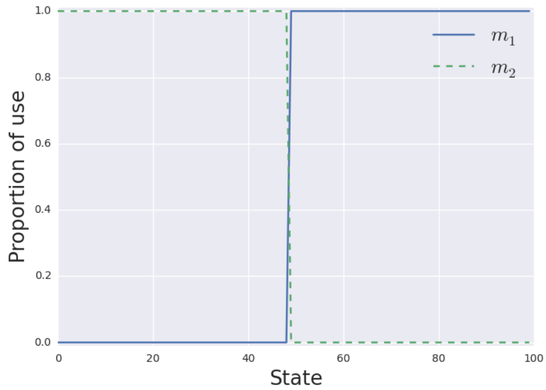
Vagueness also poses serious challenges to models of language evolution, since functional pressure towards maximal information transfer should, under fairly general conditions, work against vagueness (Lipman 2009). Many contributors have therefore argued that vagueness is intrinsically useful for communication (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, Lassiter & Goodman online first, O'Connor 2014). We follow the latter line of exploration here, arguing that vagueness arises natural under imperfect observability of states (see Franke & Correia (to appear) for an imitation-based dynamic based on the same idea).

[MF: rewrite up to here]

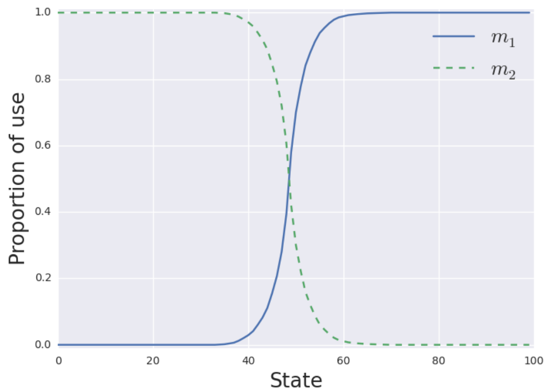
**Setup.** [MF: maybe better use $t$ instead of $\theta$? Just $t in [0;99]$] We analyze the effects a noisy perception has on the transmission of a simple language with 100 states, $s \in [0,99]$, and two messages $m \in \{m_1, m_2\}$. The probability of perceiving the actual state $s_a$ as $s_p$ is given by a normal distribution with the actual state as its mean and a standard deviation $\sigma$. That is, $P(s_p|s_a) \sim \text{Normal}(s_a, \sigma)$ with parameter $\sigma$ controlling the degree to which states are confused. [MF: this should be a truncated normal]

Signaling behavior is assumed to be uniform across speakers and to depend solely on a threshold $\theta_i$, $i \in [0,99]$. This threshold controls which message is used in a (perceived) state: If $s_j$ is the $j$-th state, then $P(m_1|s_j, \theta_i) = 1$ iff $j \geq i$. Otherwise, $P(m_2|s_j, \theta_i) = 1$. In words, if the state is (perceived to be) as large or larger than a type's threshold $\theta$, then message $m_1$ is used. Otherwise $m_2$ is used. Consequently there are 100 different types and learners will acquire a $\theta$ value based on the data they witness.

**Results.** The effect a single generational turnover under noisy transmission is depicted in Figure 2 for $\sigma = 0.4$. As illustrated in Figure 2a, the population initially consisted of a single type with $\theta_{50}$. As learners try to acquire this language, even small $\sigma$ will lead to the emergence of vagueness in the population. The same outcome is obtained for other values of $\sigma$ with straighter sigmoidal shapes in the use of the messages resulting from higher values, i.e., more borderline cases that do not clearly fall under either only $m_1$ or $m_2$. This is also true of the development of a population across multiple generations. In particular, neither message regains its once clearly delimited meaning. Instead, iterated transmission leads to type mixtures and, consequently, to convex areas of the state space that fall neither into the category of clear applications of $m_1$ nor $m_2$. The size of the state space devoted to borderline cases increases over generations with its growth being inversely related to $l$ and $k$. As is to be expected, if either the amount of samples or $k$ are too small to discern even strikingly different types from one another, then iterated learning under noisy perception leads to completely homogeneous populations with (almost) no state being exclusively associated with $m_1$ or $m_2$.




(a) Initial non-vague population




(b) Vague population after single generation

Figure 2: Noisy iterated learning with posterior sampling, $\sigma = 0.4$, $k = 20$ and 100 sampled production sequences per type.

**Discussion.** In a nutshell, transmission perturbations caused by the noisy perception of states reliably give rise to vagueness even if no borderline cases were initially part of a population's language. Of course, the stabilization of a linguistic system 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 such partition brings about for its users. That is, functional pressure may be necessary for borderline cases to be kept in check. Amongst others [TB: briefly mention other factors that have been argued to lead to vagueness with references]. In particular, Franke & Correia (to appear) have recently shown how noisy perception may lead to vagueness under functional pressure alone. Which of these factors or combination thereof plays a more central role for the emergence of vagueness is an empirical question we can 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 – be it in interaction under functional pressure or in acquisition under a pressure for learnability.

[MF: mention: vagueness only on the level of population?]

## Deflation

[TB: Short description of deflation and relevant work]

**Setup.** [MF: use $t$ instead of $\theta$?] We consider a similar setup to the one above. $S = [0,99]$, each type is associated with a threshold $\theta_i$ with $i \in [0,99]$, and the noise pattern is given by $P(s_p|s_a) \sim \text{Normal}(s_a, \sigma)$. However, we now trace the change of a single message $m$ coupled with linguistic behavior such that $P(m|s_j, \theta_i) = 1$ iff $s_j \geq \theta_t$, otherwise no message is sent. This behavior causes asymmetry in the production data as types with high $\theta$ will reserve their message only for a small subset of the state space and otherwise remain silent. Consequently, learning also needs to be modified to take such silent observations into account. For simplicity, we assume that learners are aware of $k$ and that $P(\theta|d) \propto (\prod_{s \in d} P(m|s, \theta)) \times \text{Binom}(\text{successes} = k - |d|, \text{trials} = k, \text{succ.prob} = \sum_{s'=0}^{\theta-1} P(s'))$. [MF: insert footnote to explain "Fixed $k$" idea and what the alternative should be] As before, the former factor corresponds to the likelihood of a type producing the witnessed data. In addition, 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 $\theta$, even if the few state-message pairs observed in the sequence may be equally likely to be produced by lower $\theta$.

**Results.** The development of a monomorphic population initially consisting only of $\theta_{80}$ is shown in Figure 3. In this setup even little noise will cause the message to gradually be applied to larger portions of the state space. As with the emergence of vagueness, 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 $\theta$ than present in the previous generation.

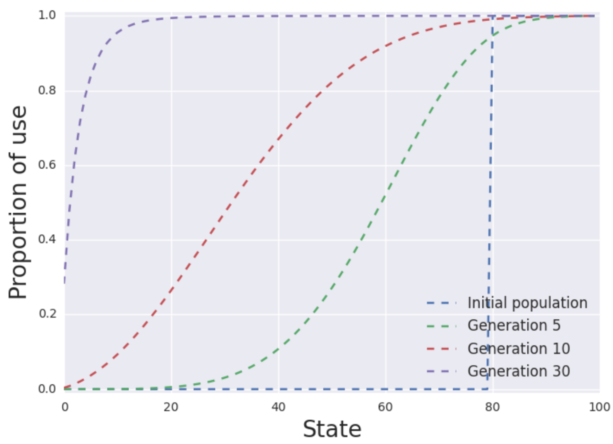



Figure 3: Noisy iterated learning with posterior sampling, $\sigma = 0.4$, $k = 30$ and 300 sampled production sequences per type.

**Discussion.** In contrast to the previous case study, the present one considers the effects of noisy perception under an asymmetry of data generation. Teachers only gave linguistic evidence when a state held true of the message according to their type. Otherwise no overt data was given to the learner. This differs from previous studies in which each state is assumed to elicit an explicit response from the teacher, even if erroneous. This setup can instead be likened to acquisition only from positive linguistic evidence in a world in which not every state is labeled (with the idealized assumption that learners are aware of the amount of silence "produced" by a parent).

The overall pattern discerned from this study is similar to that of the previous study. That is, noisy perception causes transmission perturbations that may relax once strict linguistic conventions. In contrast to the previous case study, if there are no alternative forms, e.g. *small* vs. *tall*, then asymmetry in production and noise will iteratively increase the state space that a message carves out, just as the overuse of a word may lead to the deflation of its meaning in natural language.

## Scalar expressions

Scalar expressions have been at the center of many studies on pragmatic inferences. Examples include quantifiers such as *some* and *most*, adjectives such as *cold* and *big*, as well as numerals such as *four* and *ten*. Their commonality lies in that their use often is taken to pragmatically convey an upper-bound that these expressions semantically lack (Horn 1972, Gazdar 1979). For instance, while *I ate some of the cookies* is truth-conditionally compatible with a state of affairs in which the speaker ate all of them, this utterance is usually reasoned to convey that the speaker ate *some but not all*. Otherwise, she would have used the stronger expression *all*. In this way, the meaning of a weak scalar expression is strengthened by speaker's and hearer's mutual reasoning about rational language use (Grice 1975).

To explain the selection of a lack of upper-bounds in weak scalar expressions Brochhagen et al. (2016) proposed a model combining functional pressure and iterated learning. Crucially, to explain this fact, this account requires the assumption of (at least a weak) prior that favors a lack of upper-bounds. Technically, this assumption is required to distinguish between a language that rules out the bound semantically and one that does so pragmatically. For brevity, let us call the former language $L_{\text{bound}}$ and the latter $L_{\text{lack}}$. To see the problem posed by $L_{\text{bound}}$, recall that learners need to infer unobservables such as linguistic behavior and a language from overt information. As a consequence, a user of $L_{\text{bound}}$ might therefore be hard or impossible to tease apart from one using $L_{\text{lack}}$ pragmatically, i.e., one that conveys the bound through pragmatic reasoning. In the following we focus on only these two languages to show under which conditions noisy perception may lead to the selection of $L_{\text{lack}}$ without a cognitive bias nor functional pressure.

**Setup.** We follow the setup of Brochhagen et al. (2016) but with a reduced type space by only considering $L_{\text{bound}}$ and $L_{\text{lack}}$. Both languages specify the truth-conditions for a fragment of two messages and two states. The former language partitions the state space such that $m_1$ is true of $s_1$ and $m_2$ of $s_2$. In $L_{\text{lack}}$ $m_2$ is also only true of $s_2$ but $m_1$ is true of both states, as it would be if the states were "I ate some of the cookies" and "I ate all of the cookies" and $m_1$ had the truth-conditions of *I ate some of the cookies*. For notational convenience we codify these truth-conditions in a Boolean matrix such that $L_{s_i,m_j} = 1$ iff $m_j$ is true of $s_i$, and otherwise 0.

[MF: exposition is not clear at all; what about this: use "some" and "all" as indices for messages and something similar for states; give the matrices with choice probabilities for speakers of the four types we care about for the fixed parameter values that the figure is also using; explain (superficially) how these are derived in our paper;]

There are two types of linguistic behavior; either literal or pragmatic. The production behavior of literal types is given by $P_{\text{literal}}(m|s;L) \propto \exp(\lambda L_{sm})$. Pragmatic behavior corresponds to $P_{\text{pragmatic}}(m|s;L) \propto \exp(P_{\text{literal}}(s|m;L))$, where $\lambda$ is a rationality parameter and $P_{\text{literal}}(s|m;L) \propto P(s)L_{sm}$. That is, pragmatic speakers reason about their addressees to refine their linguistic choices. This allows pragmatic users of $L_{\text{lack}}$ to convey an upper-bound with $m_1$ following the reasoning spelled-out above: If they wanted to convey the stronger state $s_2$, [MF: what's a stronger state?] they would have used stronger and unambiguous $m_2$ instead (c.f. Frank & Goodman 2012, Franke & Jäger 2014). Finally, the rationality parameter $\lambda$ controls linguistic choice. Intuitively, higher values increase the speaker's propensity to produce utterances that maximize communicative success, i.e., to use utterances that have the highest chance of being understood. For our purposes it suffices to fix $\lambda$ to be reasonably high so as to render speaker behavior (mostly) deterministic. Combining these two types of behavior with $L_{\text{bound}}$ and $L_{\text{lack}}$ gives a total of four different types.

Lastly, and differently from Brochhagen et al.'s noise-free model, noise is introduced by two parameters $\varepsilon$ and $\delta$. The former corresponds to the probability of perceiving an actual state $s_1$ as $s_2$, $P(s_1|s_2) = \varepsilon$, and conversely $P(s_2|s_1) = \delta$.

**Results.** To quantify the effects of the dynamics we ran 50 independent simulations per parameter configuration. Each population was initialized with an arbitrary distribution over types. The mean proportion of 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, types that lexicalize no upper-bounds for their weak scalar expressions provided language users are pragmatic. Similar results are obtained for increments of $k$ or $l$.

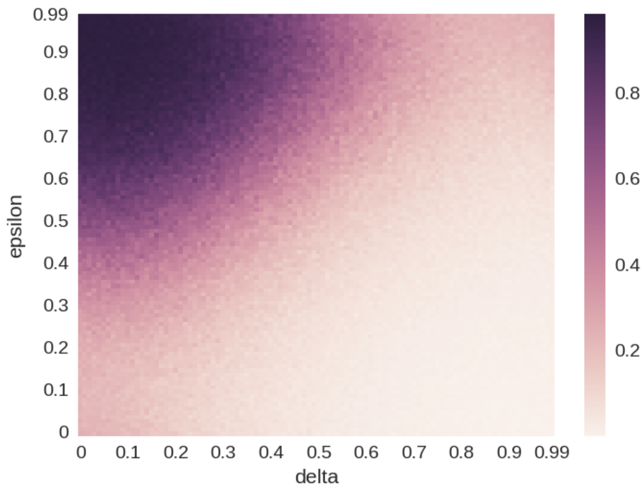



Figure 4: Mean proportion of pragmatic $L_{\text{lack}}$ users after 20 generations with posterior sampling, $k = 5, \lambda = 20$ and 10 sampled production sequences per parent type. [TB: Axes should be $\delta$ and $\varepsilon$.]

**Discussion.** The main purpose of this case study is to show that noisy perception can mimic the effect of cognitive 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 over codifying it. As noted above, this influenced the propensity of learners to infer pragmatic $L_{\text{lack}}$ over $L_{\text{bound}}$ even if the evidence provided by the data 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 it were present. However, this outcome strongly depends on the types involved. Whether a type thrives under a particular noise pattern 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. [TB: Maybe mention explicitly that it doesn't work with the full space. Add some discussion on the relation between noise and quantifiers/scalar expressions]

## Discussion

[TB: TO DO]

- We present three case studies that show how transmission perturbations can lead to the emergence of vagueness, meaning deflation, and a lack of upper-bounds in weak scalar expressions in populations of language users. These results are not meant to suggest noisy perception to be the sole or main determinant of these phenomena. Instead, this investigation's main contribution is conceptual and technical in nature in that it aims to clarify the role of systematic transmission perturbations of linguistic knowledge in language change while showing that such perturbations may stem from other sources, e.g., from learners' noisy perception.

## Conclusion

[TB: TO DO]

## Acknowledgments

[TB: TO DO]

## References

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

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. doi: 10.1016/s1364-6613(02)00005-0

Clark, E. V. (2009). Lexical meaning. In E. L. Bavin (Ed.), *The cambridge handbook of child language* (pp. 283–300). Cambridge University Press. doi: 10.1017/cbo9780511576164.016

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. doi: 10.1207/s15473341lld0102_3

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. doi: 10.1073/pnas.0608222104

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. doi: 10.1016/j.conb.2014.07.014

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. (online first). 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. doi: 10.2307/1166130

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

O'Connor, C. (2015). Evolving to generalize: Trading precision for speed. *The British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axv038

Smith, K. (2011). Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Human Biology*, *83*(2), 261–278. doi: 10.3378/027.083.0207

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.