Effects of transmission perturbation in the cultural evolution of language

Thomas Brochhagen (t.s.brochhagen@uva.nl)
Institute for Logic, Language & Computation, University of Amsterdam

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

Abstract

Two factors seem to play a major role in the cultural evolution of language. On the one hand, there is functional pressure towards efficient transfer of information. On the other hand, languages have to be learned repeatedly and will therefore show traces of systematic stochastic disturbances of 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 perturbations 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 explain attested patterns in the cultural evolution of language and that 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 are therefore not necessarily solely due to functional pressure, such as the selection of more communicatively efficient behavior. They may also be effected 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, e.g.; mutual exclusivity (Merriman & Bowman 1989, Clark 2009), simplicity (Kirby et al. 2015), regularization (Hudson Kam & Newport 2005), and generalization (Smith 2011). But forces other than learning biases may also systematically perturb the transmission of linguistic knowledge and thereby contribute to the shaping of language by cultural evolution (cf. Perfors & Navarro 2014). In the following we focus on one particular source of transmission noise: agents' imperfect perception of the world. Our overall goal is to give a formalism with which to study the possible effects of such perturbations and to 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 (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). Concretely, 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 τ (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 τ given a data sequence d of $\langle s,m \rangle$ pairs:

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

where $P(\tau)$ is the learner's prior for type τ and $P(d \mid \tau) = \prod_{\langle s,m \rangle \in d} P(m \mid s, \tau)$ is the likelihood of type τ producing the observed data d, with $P(m \mid s, \tau)$ the probability that a type τ produces message m when in world state s. It is usually assumed that learners exposed to d adopt type τ with probability $F(\tau \mid d) \propto P(\tau \mid d)^{\gamma}$, where $\gamma \geq 1$ regulates whether learners probability match ($\gamma = 1$) or tend towards choosing a maximum of the posterior distribution ($\gamma > 1$). If the set D_k of data a learner may be exposed to is the set of all sequences with k pairs $\langle s,m \rangle$, the probability that a learner acquires type τ_i when learning from a teacher of type τ_j is:

$$P(\tau_j \to \tau_i) \propto \sum_{d \in D_k} P(d \mid \tau_j) F(\tau_i \mid d)\,.$$

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 $P(\tau)$ can be understood as encoding learning biases. For example, learners may have an a priori preference for simpler languages over ones with a more complex grammar, or over ones with larger or more marked lexical or phonemic inventories (cf. 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 explorations of iterated learning have consequently suggested 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 $P(\tau)$ can influence the adoption of a linguistic type τ based on the observation of $\langle s,m \rangle$ sequences. One further potential source

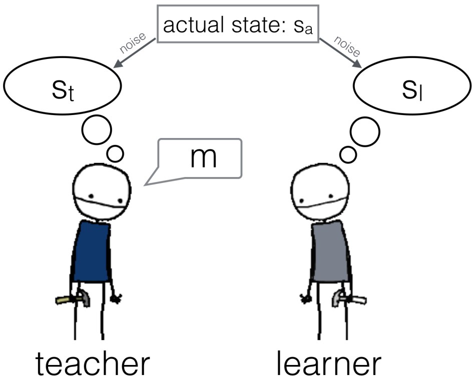

Figure 1: State-noise during observation of language use.

of "transmission noise" are regular stochastic errors in the perception of world states (see Figure 1). Imperfect perception 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 utterances 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.

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:

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

Assuming a finite state space for convenience, the probability that the teacher observes s_t when the learner observes s_l is:

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

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

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

It is natural to assume that learners, even if they (in tendency) perform rational Bayesian inference of the likely teacher type τ based on observation $\langle s_l, m \rangle$, do not also reason about state-noise perturbations. In contrast to, e.g., noisy-channel models that have agents reason over potential message corruption caused by noise (e.g. Bergen & Goodman 2015), our learners are not proficient language users that could leverage knowledge about the world and its linguistic codification to infer likely state misperception.[1] In this case the posterior probability of τ given the learner's perceived data sequence d_l is as before: $P(\tau \mid d_l) \propto P(\tau)\, P(d_l \mid \tau)$. Still, state-noise affects the probability $P_N(\tau_j \rightarrow \tau_i)$ that the learner adopts τ_i given a teacher of type τ_j, because it influences the probability of observing a sequence d_l (with $F(\tau_i \mid d)$ as before):

$$P_N(\tau_j \rightarrow \tau_i) \propto \sum_{d \in D_k} P_N(d_l \mid \tau_j) F(\tau_i \mid d)\,.$$

In sum, it may be that learner and/or teacher do not perceive the actual state as what it is. If they are not aware of this, they 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 teacher's type. She takes what she observes as the actual state that the teacher has seen as well, and infers which linguistic type (e.g. which set of lexical meanings or grammar) would have most likely generated the message to this state. This can lead to biases of inferring the "wrong" teacher type if 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 can lead to the emergence of linguistic phenomena. The 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. Instead, our aim is to elucidate the role that transmission perturbations beyond inductive biases may play in shaping the cultural evolution of language. We therefore 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 of perception, memory, or information processing (e.g. Franke et al. 2011, O'Connor 2014, Lassiter

[1] To do so, agents would have to infer or come equipped with knowledge about $P_N(\cdot|s_a)$, which could itself be subject to updates. We stick to the simpler case of ignorance about noise here, but as long as the actual state is not always recoverable our general results also hold for agents that reason about noise.

& Goodman 2015). We follow the latter line of exploration here, showing that vagueness can naturally arise 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 that agents perceive actual state s_a as perceived s_t/s_l is given by a (discretized) normal distribution, truncated to $[0;99]$, with s_a as mean and standard deviation σ. Linguistic behavior is fixed by a type $\tau \in [0;99]$ which is the threshold of applicability of m_1: $P(m_1 \mid s,\tau) = \delta_{s \geq \tau} = (1 - P(m_2 \mid s,\tau))$. In words, if a speaker observes a state that is as large or larger than its type, then message m_1 is used (*tall*), otherwise m_2 (*small*).

Results. The effects of a single generational turnover under noisy transmission of a population that initially consisted exclusively of type $\tau = 50$ is depicted in Figure 2. As learners try to infer this type from observed language use, even small σ will lead to the emergence of vagueness in the sense that there is no longer a crisp and determinate cut-off point for message use in the population. Instead, *borderline regions* in which m_1 and m_2 are used almost interchangeably emerge. For larger σ, larger borderline regions ensue. The size of such regions further increases over generations with growth inversely related to γ 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.

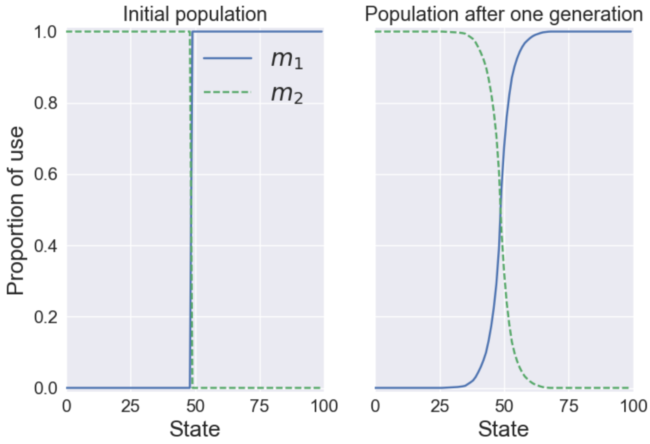

Figure 2: Noisy iterated learning with posterior sampling ($\gamma = 1$), $\sigma = 0.4$, and $k = 20$.

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

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

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, convergence on any particular population state will also depend on the functional (dis)advantages of particular patterns of language use. 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. If state perception is systematically noisy and learners are not aware of this, some amount of vagueness may be the natural result.

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 becomes a marker for standard negation. As argued by Bolinger (1981), certain word classes are particularly prone to slight and unnoticed reinterpretation. When retrieving their meaning from contextual cues, learners may consequently 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 is the same as that of the previous case study, except that we now trace the change of a single message m, e.g., emphatic negation, without a fixed antonym being sent whenever m does not apply. This is a crude way of modeling use of markers of emphasis or high relevance for which no corresponding "irrelevance marker" exists. Learners accordingly 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. This causes asymmetry in the learning data because some types 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. We assume that learners are aware of k so that:[2]

$$P(\tau|d_l) \propto \text{Binom}(\text{successes} = k - |d_l|, \text{trials} = k, \\ \text{succ.prob} = \sum_{i=0}^{\tau-1} P(s = i)) \prod_{s \in d_l} P(m|s,\tau)\,.$$

As before, the second factor corresponds to the likelihood of a type producing the perceived data. The first 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 threshold of applicability even if the few state-message pairs observed may be equally likely to be produced by types with lower thresholds.

Results. The development of an initially monomorphic population consisting only of $\tau = 80$ is shown in Figure 3. Even little noise causes a message to gradually be applied to larger portions of the state space. The speed of meaning deflation is regulated by σ, k, and to lesser degree γ. In general, more state confusion due to higher σ, shorter sequences, or less posterior maximization will lead to more learners inferring lower types than present in the previous generation.

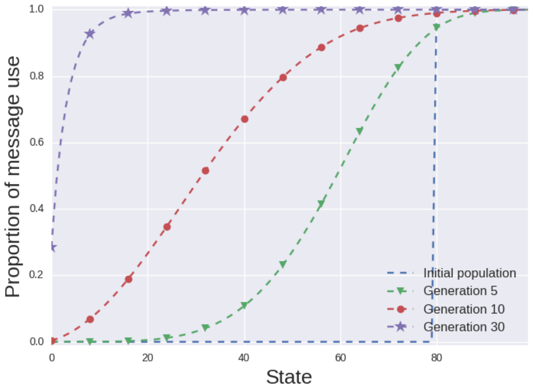

Figure 3: Noisy iterated learning with posterior sampling ($\gamma = 1$), $\sigma = 0.4$, and $k = 30$.

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, i.e., acquisition in a world in which not every state is equally likely to lead to an observable utterance.

The outcome is similar to that of the previous study. Noisy perception can cause transmission perturbations that gradually relax formerly strict linguistic conventions. In contrast to the case of vagueness, if there are no relevant competing forms, e.g., *small* vs. *tall*, asymmetry in production and noise will iteratively increase the state space that a form carves out.

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 present in their 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 is often taken to convey that Bo ate *some but not all*, as otherwise the speaker would have said *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 widespread lexicalization of upper-bounded meanings? To address this question, Brochhagen et al. (2016) explore an evolutionary model that combines functional pressure and iterated learning. This account assumes 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 behavioral strategies to use them, 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_{some}, m_{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_{bound}, which lexicalizes an upper-bound for *some*-like expressions, message m_{some} is only true of $s_{\exists\neg\forall}$ and m_{all} only of $s_{\forall}$. In the English-like lexicon L_{lack}, message m_{all} is also only true of $s_{\forall}$, but the meaning of m_{some} is underspecified and lexically holds in both states. Speakers follow one of two strategies of language use: literal or pragmatic. The former select a random true message, whereas the latter prefer to send the most informative messages from those that are true in the observed state (Grice 1975). This gives rise to probabilistic speaker behavior $P(m \mid s, \tau = \langle \text{lexicon}, \text{use} \rangle)$ which approximates the following choice probabilities:[3]

$$\begin{array}{cccc} & & \underline{L_{\text{bound}}} & \underline{L_{\text{lack}}} \\ & & \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array} & \begin{array}{cc} m_{\text{some}} & m_{\text{all}} \end{array} \\ \underline{\text{Literal}} & \begin{array}{c} s_{\forall} \\ s_{\exists\neg\forall} \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{array}{c} s_{\forall} \\ s_{\exists\neg\forall} \end{array} & \begin{pmatrix} 0.5 & 0.5 \\ 1 & 0 \end{pmatrix} \end{array}$$

[2] Knowing k allows learners to compute the likelihood of a type not reporting $k - |d_l|$ state observations. A better but more complex alternative is to specify a prior over k with learners performing a joint inference on k and the teacher's type. For simplicity, we opt for the former, albeit admittedly artificial, assumption.

[3] Concretely, results are obtained for probabilistic speaker behavior following the definitions of Brochhagen et al. (2016). Nothing essential to our main argument and simulation results hinges on these details, so we background them here for ease of exposition.

$$\underline{\text{Pragmatic}} \quad \left(\begin{array}{ccc} & m_{\text{some}} & m_{\text{all}} \\ s_{\forall} & 0 & 1 \\ s_{\exists\neg\forall} & 1 & 0 \end{array}\right) \quad \left(\begin{array}{ccc} & m_{\text{some}} & m_{\text{all}} \\ s_{\forall} & 0 & 1 \\ s_{\exists\neg\forall} & 1 & 0 \end{array}\right),$$

where $P(m|s,\tau) = M_{[s,m]}$ with M being type τ's choice matrix.

As pragmatic users of L_{lack} are (almost) indistinguishable from types with L_{bound}, the emergence of a predominance of L_{lack} in a repeatedly learning population must come from transmission biases. A learning bias in favor of L_{lack} in the learners' priors will select for it (Brochhagen et al. 2016), but here we assume no such cognitive bias. Rather we assume state-noise in the form of parameters ϵ and δ. The former is the probability of perceiving actual state $s_{\exists\neg\forall}$ as $s_{\forall}$, $P(s_{\forall}|s_{\exists\neg\forall}) = \epsilon$, and $P(s_{\exists\neg\forall}|s_{\forall}) = \delta$. For instance, states may be perceived differently because different numbers of objects must be perceived (e.g., quantifiers and numerals) or they may be more or less hard to accurately retrieve from sensory information (e.g., adjectives).

Results. To quantify the effects of the dynamics we ran a fine-grained parameter sweep over ϵ and δ with 50 independent simulations per parameter configuration. Each simulation started with a random initial population distribution over types and applied iterated learning with state-noise for 20 generations, after which no noteworthy change was registered. Mean proportions of resulting pragmatic users of L_{lack} under different noise signatures are shown in Figure 4. These results suggest that when δ is small and ϵ high, iterated noisy transmission can lead to populations consisting of almost exclusively English-like lexica with pragmatic language use. Similar results are obtained for larger k or γ.

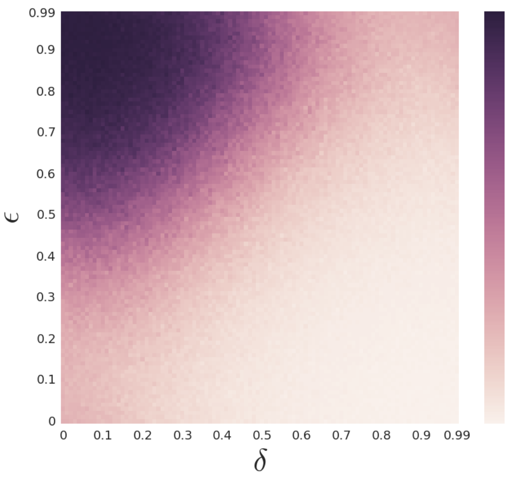

Figure 4: Mean proportion of pragmatic L_{lack} users after 20 generations with posterior sampling ($\gamma = 1$) and $k = 5$.

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; learners had an a priori preference for not codifying upper-bounds lexically, which increased their propensity to infer pragmatic L_{lack} over L_{bound} even if the witnessed data could not tease them apart. We assumed no such bias but nevertheless arrived at an evolutionary outcome that is comparable to the one predicted if the bias were present. However, 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. In the case of quantifiers, 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 general 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 outcomes of acquisition processes. Care must therefore be exercised in reading off the influence of possible learning biases from data obtained "in the wild" or the laboratory. Noisy perception can instead be regarded as a neutral model of cultural evolution as it does not appeal to functional competition nor differential learnability among types (Reali & Griffiths 2009). Second, and more importantly, these results may be seen as complementing and stressing the pivotal role of systematic 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 (cf. Perfors & Navarro 2014).

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

TB acknowledges funding from the European Community's Seventh Framework Programme under grant agreement no. 567652 /ESSENCE/. MF gratefully acknowledges financial support by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63).

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.), *Proceedings of EVOLANG 11* (pp. 25–32). Singapore: World Scientific Press.

Bergen, L., & Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, *7*(2), 336–350.

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.

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., & 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, presuposition 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).

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

Perfors, A., & Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive Science*, *38*(4), 775–793.

Reali, F., & Griffiths, T. L. (2009). Words as alleles: connecting language evolution with bayesian learners to models of genetic drift. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1680), 429–436.

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.

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.

Reviewer comments

MR+RV1-3 Why don't teachers/learners reason about noise? (cf. noisy channel models, e.g., Gibson et al 2013, Bergen & Goodman 2015). What happens if they do?
[TB: The main difference to (many?) noisy-channel models is that here states rather than messages are perceived noisily. Additionally, noisy-channel models (usually?) consider rational language use between agents that know a language, whereas our learners need to acquire one. Teachers/Learners could be modelled as having access (and correcting for) state noise signatures, but as long as this correction is not perfect we would still see the main effects we are after. I find reasoning about noise tricky conceptually as well; would teachers produce utterances correcting for what they may have misperceived (e.g. say *small* even if you think you saw *big*)? Do learners come equipped with full knowledge about noise signatures? Why? I now briefly mention some of this in the section introducing noisy IL.]

MR Do results presented correspond to stable states? Case study 1 only shows 1 step, case study 2 shows 30 and case study 3 shows 20. Do these correspond to stable outcomes?
[TB: Clear in the discussion of each case study (but admittedly not very prominent)]

RV1 Explicit comparison to other neutral models (e.g. Reali & Griffiths 2010)
[TB: Similarities: No direct competition between types in the dynamics (no selection nor selective mutation). Differences: R+G still have a learning prior that modulates selection of hypotheses.]

RV1 Clarify what a type is early on
[TB: Done inasmuch as I could given variation across studies and space limitations]

RV2 Clarify differences (or similarity) between case study 1 and 2
[TB: I think this is clear enough from the discussion on deflation]

RV2 Notational clarifications:

- k is used to designate numbers instead of n
  [TB: Minor. I'd leave it as is.]
- t is both *teacher* and *type*
  [TB: True τ for *type*]
- l is both *learner* and *posterior parameter*
  [TB: True γ for *posterior parameter*]
- a to index actual state is counterintuitive
  [TB: Disagree. I'd leave it as is.]
- Switch from s_l to s_p in case studies
  [TB: This is intended as it encompasses teacher and learner perception. I clarified this]
- Consistent distinction between P_N and P (e.g. first two equations on p.2)
  [TB: Done]

- $P(t|d_l)$: s is not defined within the binomial; is the sum in *succ_prob* really over t (not d)?
  [TB: Everything seems correct to me. With more space it would help to clarify how summing up to $\tau - 1$ in *succ.prob* relates to the speaker behavior of teacher of type τ, which is the confusing bit in this term. Then again, we at least have an intuitive explanation of the formula so I hope we're fine with only 1/3 of the reviewers being confused by this.]
- write out the binomial pmf, not its parameterisation on p.4
  [TB: I find our formulation easier to parse, so I left it. But if you think it's non-standard we can replace it by the pmf]
- scalar d vs sequence/vector d
  [TB: Haven't caught where d is allegedly used as a scalar]

RV3 Perhaps the graph in figure 2 could be changed so that the before and after populations were side by side?
[TB: Done]

RV3 Perfors & Navarro (2014) paper should be mentioned - at least to say how this approach relates to theirs
[TB: P&N consider the effects $P(s)$ has on evolving languages. In a nutshell: If k is small we see more effects of learning biases, if k is large, $P(s)$ plays a role. They contend that this relativizes K&G's convergence to the prior and, in this sense, the paper is similar to ours. Otherwise, there's no connection. I added some brief mentions of this paper but not explicit comparison.]