Sketch:
Transmission perturbations in the cultural evolution of language

January 13, 2017

1 Introduction

Language emergence and change are shaped by their use and transmission across generations. The latter process is often viewed as arising from a combination of general learning mechanisms and inductive cognitive biases (e.g. Griffiths and Kalish 2007, Kirby et al. 2014, Tamariz and Kirby 2016). Proposals of biases that shape language acquisition abound. Some prominent examples are mutual exclusivity (Merriman and Bowman 1989, Clark 2009), simplicity (Kirby et al. 2015), regularization (Hudson Kam and Newport 2005), and generalization (Smith 2011, O'Connor 2015).[1] Here, we show how environmental factors can produce evolutionary outcomes that look as if cognitive learning biases are present even if they are not. In doing so, we underline the pivotal role of systematic transmission perturbations of linguistic knowledge in language change while showing that such perturbations may stem from other sources of systematic noise, e.g., from errors in perception. This result highlights the frequently overlooked possibility that channel noise in evolutionary replication can mimic effects of inductive biases.

- Brief overview of the cases we consider

2 Model

- Introduction to iterated Bayesian learning. Highlighting predictions of the prior's influence (i) from a technical perspective with respect to sampling to MAP and (ii) conceptual perspective with respect to the claim that laboratory experiments can give us insights into learning priors [TB: For this we can draw directly from what we already had in the other draft]

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

- Noisy iterated learning. Following what we had before.
- Possibly: functional pressure.

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:

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

Finally, this gives us the probability that a teacher of type t produces a datum that is perceived by the listener as $d = \langle s_l, m\rangle$:

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

Generalize this to a sequence of perceived data d_l and write $P_N(d_l \mid t)$. Then, the noise-perturbed mutation matrix is defined as:

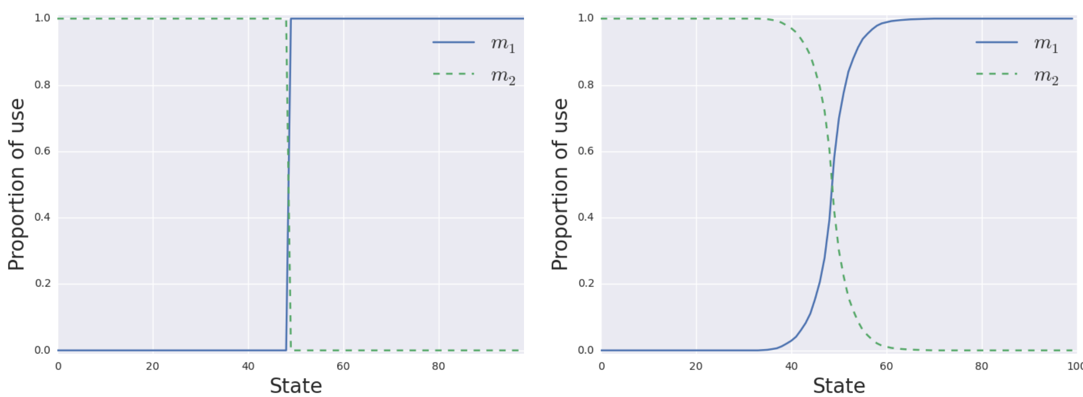
$$Q_{ij} \propto \sum_{d_l \in D} P(d_l \mid t_i) F(t_j, d_l)\,, \quad \text{where } F(t_j, d) \text{ is as before.}$$

In words, 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.

3 Applications

3.1 Vagueness

Main result. Noisy transmission perturbs initially crisp/clear linguistic distinctions, giving rise to vagueness. See Figure 1. Stabilization of the linguistic system around a particular threshold depends on functional considerations which are not modelled here but see Franke and Correia to appear.



(a) Initial "crisp" population

(b) Second "vague" generation

Figure 1: Noisy iterated learning with $\sigma = 0.4$, $k = 20$ and 100 sampled production sequences per parent (posterior sampling)

Setup.

- $S = [0, 99]$
- $|M| = 2$
- There is one signaling behavior per threshold θ and one threshold per state, i.e., 100 types.
- $P(m_1|s, t) = 1$ iff $s \geq \theta_t$, otherwise $P(m_2|s, t) = 1$.
- $P(s_{\text{perceived}}|s_{\text{actual}})$ is the probability density of getting $s_{\text{perceived}}$ from Normal$(s_{\text{actual}}, \sigma)$
- Data generated by teachers is sampled without noise to get a representative sample. But actual likelihoods of producing the data used to compute Q are subjected to noise as above (as specified above)
- Learners are not aware of noise (as specified above)
- No replication.

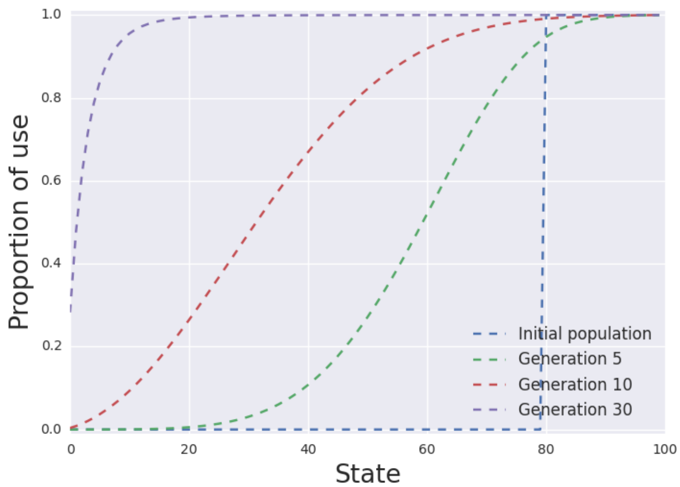


Figure 2: Noisy iterated learning with $\sigma = 0.4$, $k = 30$ and 300 sampled production sequences per parent (posterior sampling)

3.2 Deflation

Main result. Asymmetric and noisy perception can capture meaning deflation. See Figure 2.

- $S = [0, 99]$
- $|M| = 1$
- There is one type of signaling behavior per threshold θ and one threshold per state, i.e. 100 types.
- $P(m|s,t) = 1$ iff $s \geq \theta_t$, otherwise no message is sent. [TB: I'm pretty confident that adding some error-rate to this behavior wouldn't change the predictions. I left it deterministic for the time being]
- $P(s_{\text{perceived}}|s_{\text{actual}})$ is the probability density of getting $s_{\text{perceived}}$ from Normal$(s_{\text{actual}}, \sigma)$
- $P(\theta|d) \propto (\prod_{s \in d} P(m|s,\theta)) \times \text{Binom}(\text{successes} = k - |d|, \text{trials} = k, \text{succ.prob} = \sum_{s'=0}^{\theta-1} P(s'))$, where the latter is the probability of a type not reporting $k - |d|$ events for a total of k events.
- Data generated by teachers is sampled without noise to get a representative sample. But actual likelihoods of producing the data used to compute Q are subjected to noise as above (as specified above)

- Learners are not aware of noise (as specified above)
- No replication. [TB: Not sure how this would work anyway. The higher θ, the less a type communicates. If that's a communicative failure, then these types are even more dispreferred than with only learning. If it's not, then we have the same fitness for each type]

3.3 Quantifiers

Main result. Noisy perception of states can mimic cognitive biases. In this case, a bias towards simplicity (no upper-bounds) as analyzed in Brochhagen et al. (2016). Pragmatic inferences stabilize in population as byproduct of noise. [TB: Subfig(i) would show mean development of population over generations for all 4 types, Subfig(ii) would show heatmap of proportion of Gricean L_5 with x-axis ϵ and y-axis δ]

- $S = \{s_{\exists\neg\forall}, s_\forall\}$
- $|M| = 2$
- There are two lexica, one upper-bounded and one lacking upper-bound, and two signaling behaviors, literal and gricean, for a total of four lexica
- $P(m|s,t)$ is soft-maximizing literal or gricean behavior with α as exponent, using a type's lexicon – as in our other setup [TB: Alternatively, we could go for simple Boolean behavior to keep everything uniform]
- $P(s_{\exists\neg\forall}|s_\forall) = \delta$, $P(s_\forall|s_{\exists\neg\forall}) = \epsilon$)
- Learners are not aware of noise (as specified above)
- No replication

4 Discussion

[TB: To be specified. Parts can be taken from our previous draft]

5 Conclusion

References

Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31(3):441–480, 2007.

Simon Kirby, Tom Griffiths, and Kenny Smith. Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28:108–114, 2014. doi: 10.1016/j.conb.2014.07.014.

Monica Tamariz and Simon Kirby. The cultural evolution of language. *Current Opinion in Psychology*, 8:37–43, 2016.

William E. Merriman and Laura L. Bowman. The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, 54(3/4): i–129, 1989. doi: 10.2307/1166130.

Eve V. Clark. Lexical meaning. In Edith L. Bavin, editor, *The Cambridge Handbook of Child Language*, pages 283–300. Cambridge University Press, 2009. doi: 10.1017/cbo9780511576164.016.

Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102, 2015.

Carla L. Hudson Kam and Elissa Newport. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2):151–195, 2005. doi: 10.1207/s15473341lld0102_3.

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

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

Michael Franke and Jose Pedro Correia. Vagueness and imprecise imitation in signalling games. *British Journal for the Philosophy of Science*, to appear.

Thomas Brochhagen, Michael Franke, and Robert van Rooij. 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*, pages 2081–2086, Austin, TX, 2016. Cognitive Science Society.