Note on a noisy variant of the RMD

Michael Franke

Motivation

Our present model computes the mutation probability that a teacher type t_i has offspring of type t_j as:

$$Q_{ij} \propto \sum_{d \in D} P(d|t_i) F(t_j, d)\,, \quad \text{where} \quad F(t_j, d) \propto P(t_j|d)^l \quad \text{and} \quad P(t_j|d) \propto P(t_j) P(d|t_j)$$

What makes for the evolution of a natural lower-bound-only semantics for *some* is the **cognitive bias** encoded in the prior $P(t_j)$. While it is not unnatural or indefensible to assume such a cognitive bias, we also do not have a strong argument in support of it. We also do not want to argue that *this* is the right explanation, do we? Our main achievement is rather a perspicuous alignment of conceptual bits and pieces into a working formal model. The main ingredients of this model are (i) a combination of learning biases and selective pressure towards efficient communication, and (ii) the distinction between semantic meaning and pragmatic use, both of which are shaped and honed by evolution at once. This is great, but maybe we might want to have more.

Let's focus on (i) for a moment. In a recent paper, Pedro Correia and I have explored an evolutionary dynamic that ensues from "noise-perturbed imitation" (Franke and Correia, to appear). The idea is that agents update their strategies by occasionally imitating choices with a probability proportional to how good these observed choices are. Crucially, however, agents may not always perceive a state correctly. We applied this to vagueness, where this is most natural: some actual state s_a has a chance $P_N(s_p \mid s_a)$ of being observed as s_p, e.g., maybe $P_N(s_p \mid s_a) = \text{Gaussian}(s_p, \mu = s_a, \sigma = \text{something})$ is a normal distribution around the actual state. What we showed in this paper is that noisy perception of states not only leads to vagueness in signal meaning, but also speeds up the evolutionary process by unifying and regularizing agents' strategies. We interpreted this as a sort of emergent "*as-if* generalization": at the population level it seems as if agents would extrapolate and generalize what they have learned about one state to nearby similar states. But this is not, of course, what actually happens. The mere presence of systematic noise in the transmission of strategies can introduce regularization that looks as if the agents have a learning bias. But, most importantly, this disturbance of transmission fidelity is due to perceptual noise and properties of the environment, not learning biases. In other words, learning biases are clearly not the only transmission biases that can shape evolution alongside functional pressure. Environmental and perceptual noise can play a role too. This also ties in with a bunch of work on work that tries to explain some features of language as being optimal adaptations to a "noisy-channel" model (work by Ted Gibson, Steve Piantadosi etc.). However, in this work the

noise is rather in the signal m not the state perception t. In general, then, I believe that there is a certain lacuna here: paying attention to the effects of systematic distortions of state perceptions on the cultural evolution of language.

So, I asked myself whether noisy perception of states could also be included into our model and whether this could have a similar effect to a cognitive bias in favor of lower-bound-only semantics, without actually making use of such a cognitive bias. The answers in short are: "yes, yes". However, perhaps unsurprisingly, not all conceivable kinds of perception error lead to this outcome. So, if we wanted to defend that such-and-such a model is the right explanation we'd be hard pressed to defend a particular noise structure. I don't think that we can do this. The main message, in my view, should rather be: look we have two possible lines of explanation for lower-bound semantics, one with cognitive biases, the other with perceptual errors. Our main contribution is showing how two things can interact in cultural evolution, namely (i) transmission biases and (ii) functional pressure, and that, crucially, transmission biases need not only be cognitive/learning biases but can also ensue from perception and the environment. We then have a case study that puts everything into place and demonstrates the latter claim. That's the story line that we could adopt if we like the following model extension and whatever comes with it.

RMD with noisy perception

This model is implemented in file `3rmd-singlescalar_withNoise.py` in the new repository `rmd-bounds2`.

We have two states: $s_\exists$ (some but not all) and $s_\forall$ (all). Let's assume that the probabilities for observing the state in the column when the actual state is the state in the row are given by this table:

	$s_\exists$	$s_\forall$
$s_\exists$	$1-\epsilon$	ϵ
$s_\forall$	δ	$1-\delta$

So, ϵ is the error probability when the true state is $s_\exists$ and δ is the error probability when the true state is $s_\forall$. How does this affect our model of Bayesian learning?

Let's denote the probability that the teacher (learner) observes state s_t (s_l) when the actual state is s_a as $P_N(s_t \mid s_a)$ ($P_N(s_l \mid s_a)$). This is given by the table above. We can then derive the following probabilities. Firstly, the probability that s_a is the actual state when the learner observes s_l is just Bayes rule, combining prior of s_a with the noise from above:

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

Secondly, 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 $P_N(d_l \mid t)$. We then define the noise-perturbed mutation matrix as:

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

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 he tries to infer the speaker's type. He takes what he observes a state to be as the actual state that the teacher has seen as well and computes 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. I.e., this could, in principle, induce transmission biases that look as is there was a cognitive bias in favor of a particular type, simply because that type better explains the noise.

On top of changing the mutation matrix in this way, we also need to adapt the calculation of expected utilities, taking into consideration that states are perceived noisily. So, where before we had:

$$U_S(t_i, t_j) = \sum_s P(s) \sum_m S_n(m|s; L) \sum_{s'} R_o(s'|m; L)\delta(s, s')\,,$$

we now have:

$$U_S(t_i, t_j) = \sum_{s_a} P(s_a) \sum_{s_t} P_N(s_t \mid s_a) \sum_m S_n(m|s_t; L) \sum_{s'} R_o(s'|m; L)\delta(s, s')\,.$$

Results

In a nutshel: this works for some parameter settings but not for others. I have not fully explored the whole space. I would be content with a paper that says: there are parameter values that produce behavior as if there are cognitive biases without cognitive biases. I believe that this would be an additional and very interesting contribution, even if there are noise structures that have other effects.

One parameter setting that gives us the desired result (most prevalent type is t_{11}) is:

```
alpha = 5 # rate to control difference between semantic and pragmatic violations
cost = 0 # cost for LOT-concept with upper bound
lam = 20 # soft-max parameter
k = 3  # length of observation sequences
sample_amount = 50 #amount of k-length samples for each production type
epsilon = 0.3 # probability of perceiving S-all, when true state is S-sbna
```

```
delta = 0.1 # probability of perceiving S-sbna, when true state is S-all
gens = 20 #number of generations per simulation run
runs = 50 #number of independent simulation runs
states = 2 #number of states
messages = 2 #number of messages
learning_parameter = 10 #prob-matching = 1, increments approach MAP
```

One sample result is these parameter values is:

```
[ 0.01 ,  0.01 ,  0.01 ,  0.012,  0.028,  0.024,  0.01 ,  0.01 ,
        0.01 ,  0.301,  0.471,  0.102]
```

Importantly, we don't need costs ($c = 0$)! Unfortunately, we do need to tinker with α (for defensible, techical reasons) and we need to assume that $\epsilon > \delta$, otherwise either t_{10} or even t_{12} may dominate.

That means that if we wanted to also argue that this is a plausible explanation for lower-bounds semantics, we'd need to argue why it is natural that $\epsilon > \delta$. Maybe such a story can be given, but I presently do not see why/how any such story is obviously better than any other story I could come up with for any other relation of ϵ and δ.

One reasoning for $\epsilon > \delta$ could be this. When the true state is s_a, we would actually need negative evidence that some item/person/whatever does not have a property. Negative evidence of this kind might be relatively infrequent, especially if untrue. On the other hand, perceiving s_a when the true state is s_e could be more frequent, because it would, for example, require wrong information about the domain size. It could also be a cause of overemphasizing: speakers tend to exaggerate and want to claim *all*, when in fact that is not true (e.g. Schaden, 2012, for a model that has the speakers' tendency to overemphasize as a motor of language change).

References

Franke, Michael and José Pedro Correia (to appear). "Vagueness and Imprecise Imitation in Signaling Games". In: *The British Journal for the Philosophy of Science*.

Schaden, Gerhard (2012). "Modelling the "Aoristic Drift of the Present Perfect" as Inflation: An Essay in Historical Pragmatics". In: *International Review of Pragmatics* 4, pp. 261–292.