Pre-reading 13

1. In their introduction Kirby et al talk about how language universals are typically interpreted, and in particular how they are taken relate to innate constraints on language learning. Which of the following statements captures the "standard" view on universals and innate constraints?

The answer I was looking for here was “The fact that language universals exist shows that there must be strong constraints on language learning, otherwise why would all languages share some features?”. I think that’s a reasonable interpretation of at least part of the literature, and also not a bad starting hypothesis (although I think it also turns out to be wrong). If we look at a bunch of languages and see that all of them (or nearly all of them) share a particular feature, it is tempting to conclude that that feature must be built in to language learners somehow, i.e. reflect some kind of fairly strong constraint on the types of languages people can represent or learn.

2. Imagine a model where there are two types of language - let's call them type A and type B. In a Bayesian model of language learning, how would you encode a strong innate constraint on language learning, favouring languages of type A over type B?

“In the prior, p(type-A) being close to 1.”

Bayesian inference involves the interplay between the data and the prior. A strong constraint on learning means that, unless the data is really really convincing, you will go with the language you think is a priori more probable; an absolute constraint says that you ignore the data entirely. We can build a strong constraint into one of these models by setting the prior for the preferred language type close to 1 - that means that the preferred language type will have high posterior probability (and therefore be likely to be selected by the learner) unless the data is really really strongly suggesting this is not the correct hypothesis (i.e. unless the strong prior is outweighed by an even stronger push in the other direction from the likelihood). You could encode an absolute constraint by setting the prior for language type A to 1, and therefore the prior for language type B to 0 - in that case, type B languages always have 0 probability in the posterior.

3. In the same model, how would you encode a weak innate constraint on language learning, favouring languages of type A over type B?

“In the prior, p(type-A) being a little above 0.5.”

If the prior for type A languages was exactly 0.5, the prior for type B languages would also be 0.5 (because there are only two language types, and the priors sum to 1) - this would be an unbiased learner, who is governed entirely by the data. We want a learner who a priori expects to learn a type A language, but is not forced to do so by the prior (i.e. doesn’t have a prior for type A close to 1). The way to do this is to make the prior for type A close-ish to 0.5, but a little bit off the exactly-neutral value of 0.5.

4. Kirby et al use a prior favoring regularity. Under their model, which of the following orderings of prior probability of the languages aaaa, aabb and abcd is correct?

“p(aaaa) > p(aabb) > p(abcd)”

aaaa is a fully regular language (same “signal class” for all meanings), aabb is partially regular, abcd is completely irregular - the prior prefers regular languages (i.e. assigns them higher probability), so this is the correct ranking. Note a couple of things. First, the parameter alpha determines the **strength** of the prior (i.e. how much higher the prior for regular languages is), but the **order** is always the same. Secondly, they are always ordered: although it looks on visual inspection of figure 3 in the paper, bottom panel, that the prior is flat, in fact it is very very subtly skewed in the usual direction; so the prior for aaaa is always higher than aabb even if the difference is only tiny.

5. Returning to Question 1: how do the results reported by Kirby et al. change the link between language universals and innate constraints?

This for me is the key point of the paper. They compare models with strong biases (top panel of figure 3) and weak biases (bottom panel of figure 3) and find that **the stationary distribution s are the same** - in other words, if you look at the distribution of languages produced by cultural evolution, you will see it is highly skewed in favour of some language types, and this can be true even if learners have only a tiny tiny bias in favour of those languages. So this means that you can’t necessarily infer strong biases from linguistic universals (or other skewed distributions over language types) - they could be underpinned by strong constraints on learning, or they could reflect far weaker biases in individuals, amplified by cultural evolution.