Comparison between iterated learning and fitness+learning

June 6, 2016

For comparability, we follow the setup analyzed by Griffiths and Kalish (2007, §6) on the emergence of compositionality. $S = \{00, 01, 10, 11\} = M$. $\mathcal{L}$ is the set of all possible mappings from S to M, i.e. $4^4 = 256$ possible signaling systems.[1] Four of which are compositional in that meaning and form components agree position-wise.[2] Syntax, i.e. component order, is assumed to be given. $P(s) = {}^1\!/\!_{|S|}$. The simulations involve a (hierarchical) prior parameter $\alpha \in [.01, 0.5]$, an error parameter $\epsilon \in [.01, .05]$ and data length $n \in [1, 10]$.

Hierarchical prior:

$$P(L) = \begin{cases} \frac{\alpha}{4} & \text{if } L \text{ is compositional} \\ \frac{1-\alpha}{252} & \text{otherwise} \end{cases} \tag{1}$$

Production:

$$P(m|s, L) = \begin{cases} 1 - \epsilon & \text{if } m \text{ is true of } s \text{ in } L \\ \frac{\epsilon}{3} & \text{otherwise} \end{cases} \tag{2}$$

[1] Note that Griffiths and Kalish (2007) only take one-to-one mappings into consideration.

[2] Note that Griffiths and Kalish (2007) take 260 languages into consideration. 256 possible (holistic) languages and 4 compositional ones. However, the former set includes 4 'holistic' languages identical to the 4 compositional ones. In contrast, we consider 252 holistic and 4 compositional languages. (N.B.: They include 4 languages holistic languages indistinguishable from their compositional counterparts to show that the learning bias drives selection. We already know that, so it's not necessary.)

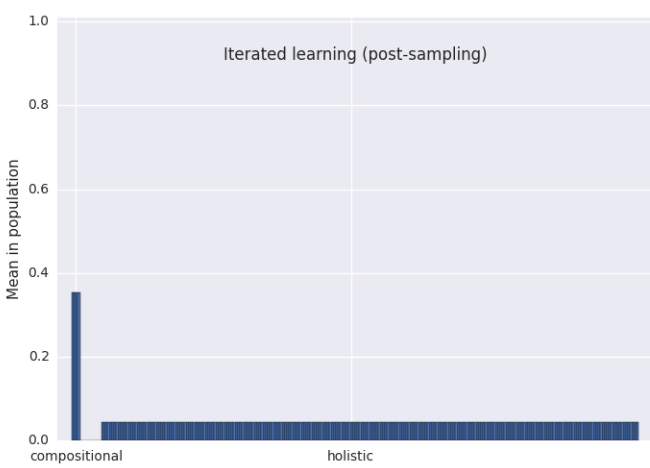
Interpretation (not covered by Griffiths and Kalish 2007):

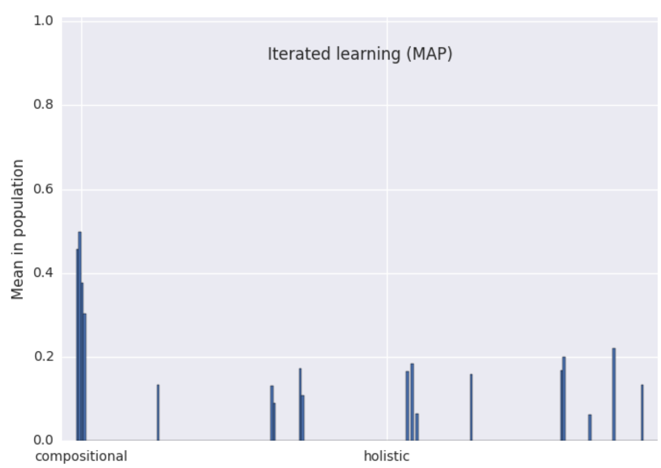
$$P(s|m, L) \propto P(s)P(m|s, L) \tag{3}$$

Results. The following plots exemplify the main differences between iterated learning and evolutionary game theory with Bayesian learning. Each dynamic was run with learning as sampling from the posterior (*post-sampling*) as well as with learning as maximum posterior likelihood (*MAP*). Furthermore, we include results for both parental learning (standard mutation) and communal learning (weighted mutation) for the EGT variant. All quantities were subjected to a square root transformation to make the differences between these variants more apparent– in particular for the quantities of holistic languages that remain under each variant. The parameter values correspond to those reported by Griffiths and Kalish 2007, 466 (rightmost plots), therefore, the results from iterated learning with posterior sampling as well as the prior double as a replication of their setup.

To ensure comparability, we also follow Griffiths & Kalish in running all simulations for 10000 generations and taking an average of 1000 such games for each parameter value and variant of dynamics/learning.

Simulation 1. $\alpha = 0.5, \epsilon = 0.05, n = 1$. Observations:





Simulation 2. $\alpha = 0.5, \epsilon = 0.05, n = 1$

Simulation 3. $\alpha = 0.5, \epsilon = 0.05, n = 1$

Simulation 4. $\alpha = 0.5, \epsilon = 0.05, n = 1$

References

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