ITERATED LEARNING WITH SELECTION: CONVERGENCE TO SATURATION

MICHAEL KALISH

Institute for Cognitive Science, University of Louisiana at Lafayette
Lafayette, LA 70504-3772 USA

A formal approach to language evolution requires specification of the properties of variation and selection. Variation is plausibly the result of replication; errors in intergenerational learning produce variability in each generation (Griffiths & Kalish, 2007). A mechanism for selection is less transparent, and this may explain a bias toward selection-free evolutionary accounts of iterated learning as intergenerational transmission. Learning has interesting properties as a source of variation since its variability is not purely random, but rather depends on the data available for learning and the inductive biases of the learners. Exploring the role of inductive biases in iterated learning has resulted in clear results concerning the dynamic and asymptotic properties of the process. However, if we assume that a single set of linguistic universals dominate human languages these results leave a puzzle, since they suggest that there should be a distribution of universals equivalent to the prior bias (that is, learnability) of these priors (Dowman, Kirby & Griffiths, 2006). One might ask, are universals homogeneous or is there some stability in their spatial heterogeneity?

Under the assumption that learners are Bayesian (that is, that they update their knowledge according to their experience), the iterated transmission of information results in the convergence of a population of independent learners to their common inductive priors (Griffiths & Kalish, 2007). To date, however, iterated learning has only been examined in the limit case of a large population of well mixed individuals, reproducing without constraint by fitness. The research presented here is a first empirical step in broadening this focus to spatially distributed populations of fixed size in which fitness plays a role in replication.

I examined two different processes that both included selection based on communicative fitness and mutation based on Bayesian learning. (1) A birth-first (Moran-like) process where only one agent in the space, chosen with a probability proportional to its relative fitness, reproduces on each cycle. The spawn then replaces a randomly chosen agent within the parent's neighborhood, possibly including the parent. (2) a deterministic (cellularautomaton-like) process where every agent is replaced by the spawn of the fittest agent in the neighborhood. Agents were defined as Bayesian learners, equipped with just two hypotheses (A and B) which they induced through exposure to samples drawn from four possible signals (see Griffiths & Kalish, 2007 for details of the 'two language' example). Agents were placed on a taurus and associated in Moore neighborhoods. I varied the number of samples (controlling stability of transmission) and the prior bias of hypothesis B (which controls the stationary distribution in the absence of selection). Fitness was symmetric between pairs of agents, reflecting their probability of mutual understanding, as in Nowak, Plotkin & Krakaur (1997).

Similar to Nowak's (2006) analytic results for arbitrary mutation, the stability of intergenerational transmission largely determined the outcome of the simulations for the deterministic process. At high stability initial conditions dominated; whatever hypothesis was most prevalent initially increased fitness for agents operating with that hypothesis and thus the transmission probability of it. At low stability, as predicted by iterated learning, bias dominated as each agent was unlikely to shift from their prior due to the noisy data. At middle levels of stability the space was likely to saturate at one of the two hypotheses, with probability determined by both stability and prior bias. Spaces in which both hypotheses were maintained indefinitely decreased with increasing stability, but only stochastically. The spatial distributions of hypotheses in these spaces were not entirely random, but self-maintaining structures did not occur. The Moran process, in contrast, converged to the prior bias regardless of initial conditions, with convergence rate decreasing nonlinearly with the number of samples seen during learning.

Either linguistic universals are homogeneous, or they are not because either (1) our space is in transition or (2) more complex processes govern the space of learners. Distinguishing these three possibilities remains a target for this research.

1.1. *References and Citations*

Dowman, M., Kirby, S., and Griffiths, T. L. (2006). Innateness and culture in the evolution of language. In Cangelosi, A., Smith, A., and Smith, K., editors, *The Evolution of Language: Proceedings of the 6th International Conference on the Evolution of Language*. World Scientific Press.

Griffiths, T. & Kalish, M. (in press). Iterated learning with Bayesian agents. *Cognitive Science*.

Nowak, M. A. (2006). *Evolutionary Dynamics*. Harvard U. Press: Cambridge

Nowak, M. A., Plotkin, J. B., and Krakauer, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162.