Iterated Learning and the Cultural Ratchet

Aaron Beppu (abeppu@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology and Cognitive Science Program
University of California Berkeley, Berkeley, CA 94720 USA

Abstract

How does the behavior of individuals in a society influence whether knowledge accumulates over generations? We explore this question using a simple model of cultural evolution as a process of "iterated learning," where each agent in a sequence learns and passes on a piece of information. Using both mathematical analyses involving rational Bayesian agents and laboratory experiments with human participants, we vary whether agents observe data from the environment and what kind of information they receive from the previous agent. Our mathematical and empirical results both suggest that merely observing the behavior of other learners is not sufficient to produce cumulative cultural evolution, but that knowledge can be accumulated over generations when agents are able to communicate the plausibility of different hypotheses.

Introduction

What we as humans know about the world is largely informed by what we can learn from others. Our beliefs and behaviors are built up not only by our own interactions with the environment around us, but by observations of the behavior of others, and by receiving explicit statements of belief. Through this ability to build off the knowledge and behaviors of those around us, human societies are able to accumulate information, the so-called "ratchet effect" (Tomasello, Kruger, & Ratner, 1993). Large, lasting changes in our beliefs develop by accruing small, incremental changes over many generations. In this paper, we use a simple model to characterize the conditions necessary for this accrual, and compare the predictions of this model to experimental results. In particular, we suggest that these changes will accumulate if learners have access to the actual beliefs of their teachers, as opposed to merely observations of behavior informed by those beliefs.

Much of the previous discussion about cumulative cultural evolution has focused on the relationship between the teacher and the learner. For example, Tomasello (1993) largely claims that the ratchet effect is a product of sophisticated imitative learning, emphasizing the advantages of joint attention and a theory of mind. Gergely and Csibra (2005) suggest that "natural pedagogy" plays a role, focusing on how human teachers go out of their way to ease the learning process for children, providing highly tailored information and a greatly enriched learning environment. Both of these mechanisms increase the fidelity of cultural transmission, providing the opportunity for the ratchet effect occur.

Rather than focusing on the specific actions or abilities of teachers and learners, we consider the kind of information passed between generations. We model the process of cultural evolution using an "iterated learning" framework (Kirby, 2001), in which successive generations are modeled as a sequence of agents in which each agent learns from the previous agent. Iterated learning models have recently been used in simulations exploring the emergence of linguistic structures (Kirby, 2001) and in experiments with human participants revealing inductive biases in specific domains (Kalish, Griffiths, & Lewandowsky, 2007). We explore variants of this framework where each agent learns not only from the behavior of the previous generation, but also from observing the world and from theories summarizing the inferences of the previous generation. The consequences of each variant for the accumulation of knowledge are determined through a combination of mathematical analyses of sequences of Bayesian agents and an experiment in which cultural evolution is simulated in the laboratory with human participants.

Our mathematical and empirical analyses produce two results that help to identify the conditions under which cumulative cultural evolution can occur. First, we show that if the information passed from teacher to learner is limited to the learner's observations of the teacher's behavior, then cumulative cultural evolution will not take place. We then go on to show that cumulative cultural evolution can occur if teachers give learners information about which hypotheses they believe are plausible, rather than just behaviors consistent with the most plausible hypothesis. These results suggest that a capacity for communicating such beliefs may be necessary in order for the cultural ratchet to operate.

The plan of the paper is as follows. The next section introduces our formal framework, and summarizes our mathematical results. We then outline the motivation for our experiment with human learners, and describe the procedure and results of that experiment. The paper concludes with a discussion of the implications of these results for understanding the conditions under which the cultural ratchet can operate, and the kinds of cognitive capacities that teachers and learners need to possess in order to satisfy these conditions.

Analyzing Cultural Transmission

Our first step in exploring the conditions that result in cumulative cultural evolution is analyzing variants of iterated learning in which the learners are rational Bayesian agents. The assumption of rationality allows us to examine how ideal learners use the information transmitted between generations, and allows us to be explicit about the biases that influence learning. Specifically, we assume that all agents have a shared set of hypotheses H, and a prior distribution over these hypotheses $p(h)$ describing the degree to which each $h \in H$ is believed to be true in the absence of any evidence.

Starting with these prior beliefs, agents observe the world and receive information from other agents, resulting in data,

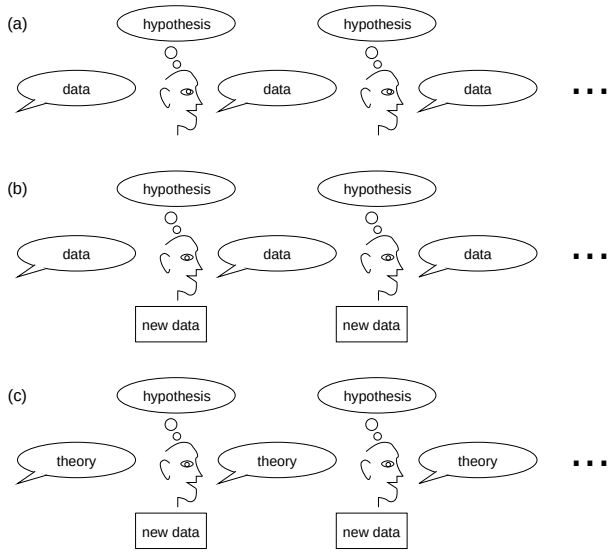

Figure 1: Three models of cultural evolution. (a) In the *pure iterated learning* model, each learner receives data from the previous learner and generates data provided to the next learner. (b) In the *mixed data* model, each learner also receives data generated by the external world. (c) In the *posterior passing* model, each learner passes information about his or her current posterior distribution over hypotheses to the next learner (here taking the form of a theory).

d. Using these data, each agent updates his or her beliefs by computing a posterior distribution $p(h|d)$ using Bayes' rule

$$p(h|d) = \frac{p(d|h)p(h)}{\sum_{h' \in H} p(d|h')p(h')} \quad (1)$$

where $p(d|h)$ is the likelihood, expressing how probable the data are given a particular hypothesis, and the sum in the denominator ensures that the result is a normalized probability distribution. The learner can now choose a hypothesis based on its posterior probability. Intuitively, Bayes' rule provides a way to combine prior beliefs with observations, with the probability assigned to each hypothesis being modified by the extent that it agrees or conflicts with the evidence.

In the remainder of this section we describe three models of cultural evolution, varying in the nature of the data provided to learners. These three models are illustrated schematically in Figure 1. We analyze how the hypotheses selected by the learners change over time in each of these models, allowing us to determine whether the assumptions behind the model are sufficient to support cumulative cultural evolution.

Pure Iterated Learning

The first model might be called the *pure iterated learning* model, as it represents the standard scenario assumed in iterated learning. In this scenario, each agent receives only data which are generated based on the hypothesis maintained by the previous agent. An intuitive example of this sort of process is the game "telephone", where each individual in a chain hears some message whispered by the previous person, and whispers what they think they heard to the next person.

More formally, all agents share a prior $p(h)$. There is some "true" hypothesis h^* which is used to initiate the process. The first agent receives data d^* sampled from the distribution associated with this hypotheses, $p(d^*|h^*)$. This agent's beliefs are updated by applying Bayes rule (Equation 1) to reach a posterior $p(h_1|d^*) \propto p(h_1)p(d^*|h_1)$, samples a hypothesis h_1 from this distribution, and produces data d_1 by sampling from $p(d_1|h_1)$.[1] These data are passed to the second agent, and each subsequent agent similarly receives and transmits data.

This process defines a Markov chain over hypotheses, with transition matrix $p(h_i|h_{i-1}) = \sum_{d_{i-1}} p(h_i|d_{i-1})p(d_{i-1}|h_{i-1})$. This chain has the prior $p(h)$ as its stationary distribution, meaning that the probability a learner chooses a hypothesis h asymptotically converges to $p(h)$ (Griffiths & Kalish, 2007). Laboratory experiments simulating this process of iterated learning with human participants produce results that are consistent with this prediction (Kalish et al., 2007).

An intuitive explanation for this convergence result can be provided by considering the "telephone" example. In this case, we expect that the original message tells us almost nothing about the message heard by a person far down the chain. This suggests that information is being lost over time. Loss of information allows the prior to influence the hypotheses selected by learners, who will use their a priori expectations to fill the gaps. Each generation, in reconciling the received data with their prior, thus passes less surprising data to the next generation. The prior itself is the only distribution over hypotheses that will be stable under this process.

Mixed Data

We can take a step towards producing cumulative cultural evolution by imagining a case where each person both receives data from the previous person, and observes data from the external world. This is the *mixed data* case. Receiving data from the world is necessary for cumulative cultural evolution, as agents are actually trying trying to learn about something external to their society, rather than a purely cultural creation. As an example, we might imagine that a hunter observes locations where antelope can be found grazing, but more experienced hunters also recommend specific locations where they would expect to find antelope. By combining these specific locations with prior beliefs about where antelope are found, the hunter reaches a new set of beliefs about where antelope are likely to be, and can tell other hunters about specific locations where he expects to find antelope.

The formal description is a simple extension of the above model. All agents share some prior $p(h)$. Each agent i still receives data d_i from the previous agent. Additionally, the ith agent receives some data d_i^* from the world, generated from $p(d_i^*|h^*)$, where h^* is the true hypothesis that

[1] All results given in this section still hold if learners sample data directly from the posterior predictive distribution, summing over all hypotheses, rather than sampling a hypothesis first.

characterizes the state of the world. The agent then updates his or her beliefs by applying Bayes' rule as in Equation 1, with $p(h_i|d_i,d_i^*) \propto p(h_i)p(d_i|h_i)p(d_i^*|h_i)$, and generates data to pass on to the next agent, with $p(d_{i+1}|d_i,d_i^*) = \sum_{h_i} p(h_i|d_i,d_i^*)p(d_{i+1}|h_i)$. As a Markov chain over hypotheses, this would have transition matrix $p(h_{i+1}|h_i) = \sum_{d_i}\sum_{d_i^*} p(h_i|d_i,d_i^*)p(d_i|h_i)p(d_i^*|h^*)$.

An analysis similar to that for the case of pure iterated learning shows that this Markov chain will converge to the average of the posterior with respect to the distribution from which data from the world are generated, $\sum_{d^*} p(h|d^*)p(d^*|h^*)$. Thus, cultural transmission does not result in progress: asymptotically, learners will choose exactly the same hypotheses as if they had only seen data generated from $p(d^*|h^*)$, receiving no benefit from cultural transmission. Intuitively, it should be clear that this process will not converge to the prior, because of the data observed from the world, but it should also not converge to the truth, because each individual is still using the same prior and only has a small amount of data – the combination of d^* and d – from which to learn. In our hunting example, we might suppose that all hunters have a strong prior belief that antelope are likely to hide in thick brush, but in fact antelope are more likely to graze in open fields. Then even if very few of the observations of antelope are in thick brush, each hunter will nevertheless recommend thick brush locations.

Posterior Passing

When learning from others, we are usually not limited to merely observing their behavior or listening to their predictions. Often, we receive actual statements of belief. This is the last scenario we consider, which we call *posterior passing*. Each agent has access to the beliefs of the preceding person, as expressed in a posterior distribution over hypotheses rather than data reflecting those beliefs, and takes that posterior distribution as their own prior distribution. Agents also observe data from the world. A model along these lines was used by Boyd and Richerson (1993) to analyze balancing tradition with innovation in cultural evolution. As an example, we can return to our hunters, where this time experienced hunters help train novice hunters by laying out a set of hypotheses about the habits of antelope, and indicating the extent to which they think each of these hypotheses is likely.

This process converges to the truth – the beliefs of successive generations will converge on the hypothesis h for which $p(d^*|h)$ is closest to the distribution $p(d^*|h^*)$ that characterizes the state of the world h^* (where proximity is measured by the Kullback-Leibler divergence, Cover & Thomas, 1991). The intuitive explanation is that taking the teacher's posterior distribution as the learner's prior makes it possible to preserve the inferences of previous learners, in contrast to the *mixed data* case. Formally, if the first agent's posterior is $p(h_0|d_0^*)$, and the second agent takes this as a prior, then the resulting posterior is $p(h_1|d_0^*,d_1^*)$. If agents 0 through $(i-1)$ do this, then the posterior distribution of the ith agent is $p(h_i|d_0^*,\ldots,d_i^*)$. Thus, the whole chain acts like a single agent who has seen all of the data, and convergence to the truth follows from standard proofs of the asymptotic consistency of Bayesian inference (e.g., Robert, 1994).[2]

Summary

These three models describe fundamentally different kinds of processes. The *pure iterated learning* model is applicable in contexts where the thing being learned is purely about the society the agents live in – the structure of a language, or taboos and social norms. The *mixed data* case and the *posterior passing* case are both situations where the agents are actually trying to learn about the world, but are distinct in the kind of transmission that occurs. The crucial point is that only in the *posterior passing* case are the inferential contributions of all individuals preserved. In the *mixed data* case, the chain will converge, but not to the true distribution; the progress made by each individual upon observing data from the world is undone by the next participant's inferential process, and thus the cultural ratchet slips. This analysis makes a clear prediction about the conditions under which cumulative cultural evolution will take place, which we can now test in an experiment with human participants.

Exploring the Cultural Ratchet through Function Learning

We test the predictions of our models by extending the function learning paradigm used by Kalish et al. (2007). In this paradigm, people learn a one-dimensional function relating two variables x and y. Each person receives data in the form of a set of (x,y) pairs, which are points generated by that function. After receiving these data each person is given a sequence of x values and asked to provide corresponding y values based on their beliefs about the function. The resulting (x,y) pairs indicate what the person has learned.

This paradigm gives us a convenient way to test the predictions of our models. Previous experiments using this paradigm have shown that people are strongly biased towards learning positive linear functions, finding these functions easy to learn, and have difficulty with non-monotonic functions (Busemeyer, Byun, DeLosh, & McDaniel, 1997). We interpret this inductive bias as a prior distribution placing high probability on positive linear functions. If we construct a scenario in which the true function people should be learning has low probability under this shared prior, such as a non-monotonic function, we should be able to determine whether a given scenario results in convergence towards the prior or towards the true function. In particular, if each person receives an amount of data that would normally be insufficient to allow them to learn a nonmonotonic function, cumulative cultural evolution should be the only way for people to learn the true function. Even if convergence is not directly observable in the lab, we can still evaluate whether a few genera-

[2] In fact, agents need not communicate the entire posterior distribution: Sampling hypotheses from the posterior is sufficient to converge to the truth, as in sequential Monte Carlo methods such as particle filtering (Doucet, Freitas, & Gordon, 2001).

tions of *posterior passing* produces results that are closer to the true function than those produced by *mixed data*.

Methods

Participants

We collected data from 208 participants drawn from the university community. Participants received either course credit or reimbursement of $10/hour for volunteering. These participants were arranged into chains of eight participants each, specifically eleven *posterior passing* chains, ten *mixed data* chains, and five *pure iterated learning* chains. Fewer *pure iterated learning* chains were used because this case is equivalent to an existing experiment by Kalish et al. (2007).

Stimuli and Apparatus

Participants completed the experiment in individual booths. All stimuli and responses were displayed on and recorded through an Apple iMac. In each trial, a stimulus value x was presented in the form of a horizontal blue bar, 0.75 cm wide and from 0.5 cm to 18.75 cm long (corresponding to $x = 0$ and $x = 1$ respectively). The stimulus bar was presented in the upper-left section of the screen, with its upper left corner 7 cm from the left edge and 3 cm from the top of the screen. Participants gave a response value y by using a slider to adjust the height of a vertical red bar of the same dimensions as the stimulus bar. The response bar was 0.5 cm from the bottom of the screen and 5.5 cm from the right edge of the screen. Participants were able to adjust the response bar using a slider until satisfied, and then submitted their responses by clicking a button approximately 4 cm to the right of the red bar. Feedback was given in the form of a yellow bar (of the same dimensions) approximately one cm to the left of the response bar. Additionally, in the "posterior-passing" condition, text was displayed in a box on screen at the same time as the stimulus, response and feedback bars.

Stimuli were taken from a pool of 100 values evenly spaced on $[0.005, 0.995]$. The true function that people were supposed to learn was $y = f(x) = 4(x-0.5)^2$. Given that we fix the range of stimulus and response values to [0, 1], this is a concave up parabola, with $f(0) = 1, f(0.5) = 0, f(1) = 1$.

Procedure

All conditions of the experiment followed the same general structure. During a training phase, a participant was shown a sequence of stimulus values. For each of these values, with the stimulus bar still on screen, participants chose a response by adjusting the slider, and submitted their response by clicking a button. The feedback bar was then displayed showing the correct response for the given stimulus. This was followed by a one second pause during which all three bars were displayed. Participant responses were considered correct if their error was less than 5% of the bar's range of values (slightly less than one cm). Participants with correct responses moved immediately onto the next trial. If a participant's response was outside this margin of error, a tone sounded. The participant then adjusted the response bar to match the feedback bar, and submitted this value. This was followed by a two second pause, and the onset of the next trial. During a test phase, participants performed the same task, but without feedback; given a presentation of the blue bar, they simply submitted a guess of the corresponding height of the response bar.

Participants were divided into chains of eight participants in each of the three conditions. In all chains, participants were trained using some set of stimulus/response values, and then tested using 50 distinct stimulus values. Conditions differed in the data that were used to train participants. Across all conditions, the first participant in any chain was trained with 50 distinct data points from the true function $f(x)$. In the *pure iterated learning* condition, following participants were trained using the 50 stimulus/response pairs generated by the previous participant during the test phase. In the *mixed data* condition, subsequent participants were trained with these 50 recycled values, but in addition, received 50 data points generated from the true function. In the *posterior passing* condition, all participants were trained using only data from the true function (again, 50 data points). In addition, each participant received a description written by the previous participant describing what the previous participant thought the function was. This description was displayed on screen during both training and test phases. Each participant wrote a corresponding description at the end of the test phase.

Results

Example chains from each condition are shown in Figure 2. The *pure iterated learning* condition quickly converged to a positive linear relationship for all but one chain, which converged to a negative linear relationship. This is consistent with the results of Kalish et al. (2007), in which 28 of 32 chains converged to a positive linear relationship and the remainder converged to a negative linear relationship. Chains from the *mixed data* conditions generally looked like noise, sometimes with small areas where the participant latched onto something correct. Sometimes a participant in these chains would induce a linear relationship, though this generally disappeared from the chain within a few generations. Importantly, the responses of later participants in a chain resembled the responses of the first participant, consistent with the prediction that receiving data from the previous generation should ultimately not improve performance. Lastly, the *posterior passing* chains did tend to approach the true function more closely than the other groups. However, this was not necessarily actual convergence. In a few cases, a chain would approach the true function, but a participant late in the chain would infer the wrong relationship, derailing convergence.

We performed a more quantitative analysis of the data by computing the root mean square error (RMSE) with respect to the true function, $\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(f(x_i)-y_i)^2}{n}}$, for each participant. Figure 3(a) shows this error for each participant, grouped into chains. Though the raw data is messy, we can still glean quite a lot from it. The *pure iterated learning*

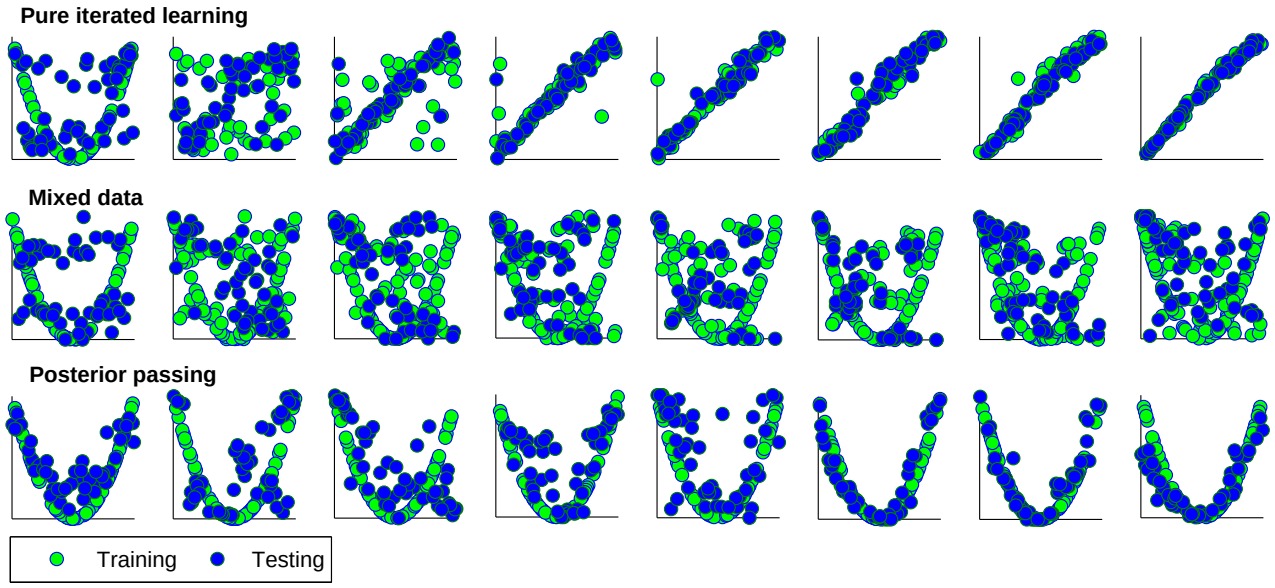

Figure 2: Example chains from each condition. Each row is a single chain, with each plot representing a participant. Training data are shown in green, while test responses are shown in blue.

Table 1: Results of ANOVA comparing average RMSE of chains across conditions

comparison	df_1	df_2	F	p
all conditions	2	23	15.12	6.4e-5
pure v. mixed data	1	13	23.97	0.0003
pure v. posterior passing	1	14	20.54	0.0005
mixed data v. posterior passing	1	19	8.67	0.0083

Table 2: Results of ANOVA with the RMSE values of individual participants across conditions in each generation

Generation	F	p	Generation	F	p
1	1.11	0.3480	5	7.19	0.0038
2	1.03	0.3747	6	6.24	0.0068
3	4.44	0.0235	7	3.82	0.0368
4	9.06	0.0013	8	4.13	0.0293

Note: In all cases $df_1 = 2, df_2 = 23$.

chains proceed quickly to an RMSE of around 0.5. Given that responses were limited to $[0,1]$, this shows the total decoupling of the chain from the true function, as expected. We note that the mixed data and posterior passing chains had a highly varied mix of values, where the error changes dramatically over the generations. However, it is also clear that the only chains which get at all close to the true function (for instance, RMSE < 0.2) are posterior passing chains. Table 2 gives the results of comparing the RMSE across conditions at each generation via a one-way ANOVA, showing that this difference is statistically significant after just three generations.

The separation between the conditions becomes clearer if we track the cumulative average of the RMSE for each chain. Figure 3(c) shows these cumulative averages for each chain as a function of generation, and Figure 3(d) averages these values for each condition. *Posterior passing* chains as a group tended to have lower error than the *mixed data* chains. The difference in performance of chains across conditions was significant. Table 1 (top) gives the results of comparing the average RMSE for each chain across conditions, showing highly significant differences between conditions.

Discussion

We have used the iterated learning framework to define several simple models of cultural evolution, casting the beliefs of each generation as a random variable in a Markov chain. This makes it possible to analyze the convergence properties of these Markov chains, making predictions about the conditions under which cumulative cultural evolution will emerge. Our experimental results are consistent with these predictions: Cumulative cultural evolution does not occur when learners simply observe the behavior of a previous learner, but does occur when learners communicate their beliefs about the state of the world.

The ability to keep the inferential contributions of previous generations is, we claim, what allows the cultural ratchet to work. Whether in making tools or inferring properties of the world around us, retaining actual ideas and hypotheses over time prevents cultural evolution from backsliding. In the *posterior passing* case, we observed that a sequence of participants is able to make successively better inferences over several generations. This is in stark contrast to the *mixed data* case, where the last participant looks just as confused as the first. The inferential steps taken by each generation actually degrade any progress made by the previous generation.

This work introduces several questions to be answered by future investigation. A skeptical reader might ask, for instance, what the progress of *mixed data* chains looks like when the amount of data from the world and data from the previous participant changes. Could it merely be that receiving the posterior beliefs of the preceding generation functions

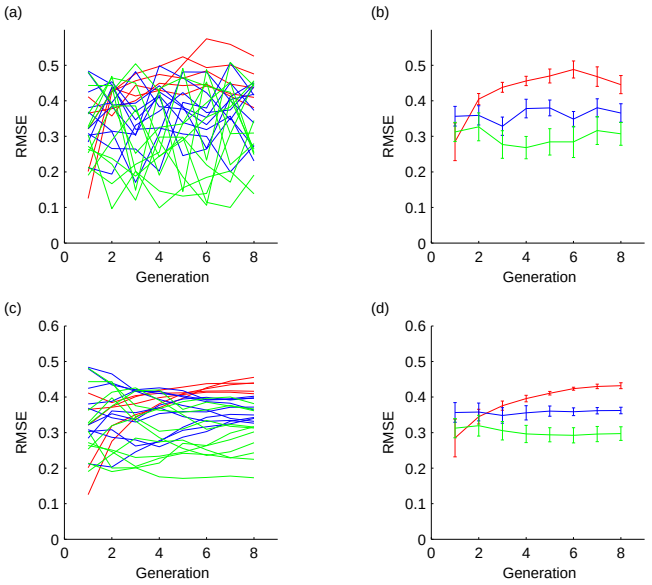

Figure 3: Root mean squared error (RMSE) to the true function at each generation of iterated learning. In all plots, the *pure iterated learning* condition is given in red, the mixed case is given in blue, and the posterior-passing case is given in green. (a) RMSE for each chain. (b) Average RMSE for each condition. (c) Cumulative average RMSE for each chain. (d) Average cumulative RMSE for each condition.

just like receiving more data? Our model indicates that no amount of observational data is sufficient to produce a ratchet effect.

Two other related extensions involve agents that reason about the inferences of other agents. First, what if in the *mixed data* case, datapoints were labeled with their source? If each learner knows which datapoints are predictions from a previous agent, it is conceivable that learners would use these to recover an estimate of the teacher's posterior beliefs. Secondly, what behavior would we see if in the *mixed data* case both teachers and learners were both aware that predictions would be passed between generations (as in Shafto & Goodman, 2008)? We might expect teachers to tailor their predictions to help the learners, implementing the kind of communication that might convey information about plausible hypotheses from one generation to the next.

Other approaches to simulating cultural evolution in the lab may also be useful, particularly for exploring the teacher/learner interaction. For instance, Caldwell and Millen (2008) introduced a method for exploring cumulative cultural evolution in the laboratory using a problem solving task (building better paper airplanes, and taller spaghetti/clay towers). Cultural evolution was simulated by replacing participants one by one, allowing them to directly interact with the previous generation. This establishes a richer setting to model, and the opportunity to observe interactions to determine the kind of information being transmitted.

The key result of our mathematical and empirical analyses is that communication of theories about the world, and not just observational learning, is important in order for the inferences of each generation to be maintained. In the broadest possible scope, our findings tell us that cumulative cultural evolution requires communication, not merely observation. Considering the cognitive capacities that would support this kind of communication may provide a means of understanding what makes humans unique in being able to exploit the cultural ratchet.

Acknowledgments. This work was supported by grant number BCS-0704034 from the National Science Foundation.

References

Boyd, R., & Richerson, P. J. (1993). Rationality, imitation, and tradition. In R. H. Day & P. Chen (Eds.), *Nonlinear dynamics and evolutionary economics.* Oxford: Oxford University Press.

Busemeyer, J. R., Byun, E., DeLosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K. Lamberts & D. Shanks (Eds.), *Concepts and categories* (p. 405-437). Cambridge: MIT Press.

Caldwell, C. A., & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, *29*, 165-171.

Cover, T., & Thomas, J. (1991). *Elements of information theory*. New York: Wiley.

Doucet, A., Freitas, N. de, & Gordon, N. (2001). *Sequential Monte Carlo methods in practice.* New York: Springer.

Gergely, G., & Csibra, G. (2006). The social construction of the cultural mind : Imitative learning as a a mechanism of human pedagogy. *Interaction Studies*, *6*, 43-48.

Griffiths, T. L., & Kalish, M. L. (2007). A Bayesian view of language evolution by iterated learning. *Cognitive Science*, *31*, 441-480.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review*, *14*, 288-294.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, *5*, 102-110.

Robert, C. P. (1994). *The Bayesian choice: A decision-theoretic motivation*. New York: Springer.

Shafto, P., & Goodman, N. D. (2008). Teaching games : statistical sampling assumptions for learning in pedagogical situations. *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*.

Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495-552.