Cultural Evolution with Sparse Testimony: when does the Cultural Ratchet Slip?

Andrew Whalen (aczw@st-andrews.ac.uk)
Department of Biology, University of St Andrews, St Andrews, Fife KY16 9TH UK

Luke Maurits (luke.maurits@berkeley.edu)
Michael Pacer (mpacer@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California, Berkeley, Berkeley, CA 94720 USA

Abstract

Humans have accumulated a wealth of knowledge over the course of many generations, implementing a kind of "cultural ratchet". Past work has used models and experiments in the iterated learning paradigm to understand how knowledge is acquired and changed over generations. However, this work has assumed that learners receive extremely rich testimony from their teacher: the teacher's entire posterior distribution over possible states of the world. We relax this assumption and show that much sparser testimony may still be sufficient for learners to improve over time, although with limits on the concepts that can be learned. We experimentally demonstrate this result by running an iterated learning experiment based on a classic category learning task.

Introduction

The sciences are impressive; humanity can be proud. However the work of science can hardly be conceived of, let alone realized, in a single generation. Data needs to accrue over time to shine light on different theories; with new information, the landscape of theories changes, making some plausible while rendering others unthinkable. In addition to science, humanity has accumulated a vast body of practical knowledge and technology which has permitted our adaptation to nearly every environment on Earth (Boyd & Richerson, 1988). Our species is distinguished by this accumulation of knowledge, but our understanding of the process underlying this "cultural ratchet" (Tomasello, 1994, 1999) is in its early stages.

One aspect of the cultural ratchet's operation that may have significant consequences for cultural evolution is the amount and kind of information that is passed between generations. Beppu and Griffiths (2009) investigated this aspect of the cultural ratchet in the context of an iterated function learning task. When participants in this task provided testimony to future participants which consisted only of demonstrated data (assumed to be undifferentiated from data observed "in the wild"), groups performed no better than single learners who received only data from the world — no ratcheting effect occurred. However, when participants provided their entire theory about how the world works — in Bayesian learning terms, their complete posterior beliefs — then the groups eventually learned the correct function.

We know that when copious and rich information is passed from generation to generation, the cultural ratchet works flawlessly; given a steady stream of data from the world, science marches forward. We also know that under conditions of much poorer information passing, the cultural ratchet "slips"; a constant stream of data does not guarantee progress. Between these two extremes are a range of possibilities, which may lie closer to actual human information passing than either extreme. What forms of testimony passing are needed for the cultural ratchet to catch more often than slip? We take a first step toward addressing this question in this paper.

We construct a simple model of iterated learning with "sparse" testimony and consider three forms of evaluating social testimony. We apply this model to a category learning task and find that limited social testimony may lead to iterative improvements across generations; however we also find that these improvements may not allow learners to find the correct hypothesis. We find that in hard category learning tasks, with limited personal data, learners may not perfectly learn the category, although they perform better than the initial learners in the chain. These predictions are confirmed by an iterated learning experiment using a similar category learning task. In the experiment, we find that participant accuracy improved across generations, however in most of the conditions the amount of improvement is limited depending on the difficulty of the task and the amount of private data received. These results suggest that while passing limited testimony can still be sufficient to improve the accuracy of groups compared to receiving no testimony, it may not be enough to learn particularly challenging tasks with limited data.

Iterated Learning and Cultural Evolution

Iterated learning is a widely used computational and experimental paradigm for understanding how inductive biases might shape linguistic preferences over the course of multiple generations and influence how languages develop and change (Kirby, 2000, 2001; Perfors & Navarro, 2011). It has since been generalized beyond this setting. Griffiths and Kalish (2007) showed that if learners receive only social testimony then the long term distribution of beliefs of the population will be the same as the prior beliefs of each learner.

Much of human learning does not take place purely on the basis of socially transmitted information. Human learners also receive data directly from the world: be they scientists measuring the behavior of particles in a laboratory or hunter-gatherers testing new tools in an unfamiliar environment. When learners receive outside data as well as testimony, the convergence to the prior shown by Griffiths and Kalish does not hold, and learners' long-term behavior de-

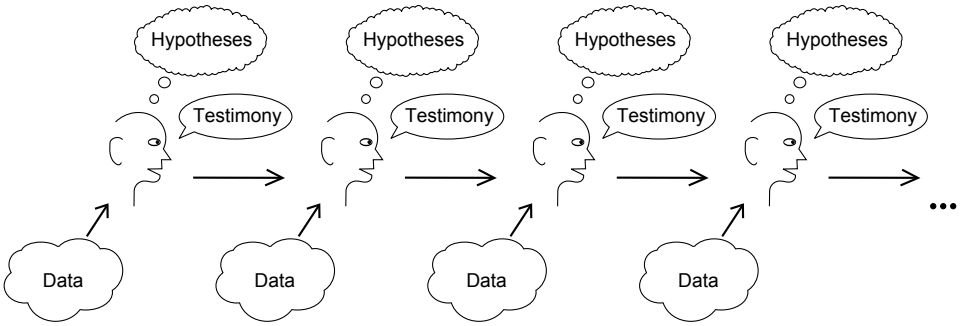

Figure 1: A graphical diagram of iterated learning with external data. Learners receive testimony from the previous learner and private data from the world which they use to evaluate hypotheses and produce testimony.

pends crucially on the type of testimony being passed. Beppu and Griffiths (2009) were the first to explore what happens when an additional source of knowledge was added alongside testimony. They found that when learners were presented social testimony in the form of teacher-generated data that was undifferentiated from environment-generated data, learners' actions at the end of the chain were no different than a learner who just received environmental data. However, they also found that when given the previous learner's entire set of beliefs, learners far down the iterated learning chain would accurately learn the true state of the world, no matter how little private data was given at each step of the way.

Actual testimony likely falls between these two extremes. Human testimony is richer than just providing extra, undifferentiated examples. Even in the case when learners provide examples, the examples are often crafted to teach the learner about a more general phenomenon, and are interpreted by the learner in a different way than non-socially produced examples (Csibra & Gergely, 2009). Humans also have the ability to directly transmit abstract concepts through writing and speech. These abstract concepts are likely richer than examples and may provide direct support for certain theories about how the world works. Yet a teacher almost certainly does not tell the learner everything they know about the world, or even a single subject.

When given social testimony to evaluate, learners face the task of integrating this testimony with concrete data from their own experiences. This is a probabilistic inference problem that has not previously been addressed in the context of iterated learning. Given that passing a full posterior may be either costly or inefficient, it is important to understand how knowledge can accumulate under limited information transmission. To do this, we analyze a model of iterated learning from testimony.

Modeling the Effects of Testimony

Past work on iterated learning has focused on understanding the stationary distribution of the beliefs of learners after a large number of generations. In order to understanding how knowledge can accumulate under the condition of limited social testimony, we analyze learners who use Bayes' rule to evaluate hypotheses, h, about the world given private data, d, and testimony t,

$$p(h|d,t) \propto p(d,t|h)p(h). \tag{1}$$

Formal models of iterated learning calculate the probability that a learner at generation i adopts a belief, h_i after hearing testimony from a learner at time $i-1$ with belief h_{i-1}. This can be expressed as the conditional probability $p(h_i|h_{i-1})$, and can often be found by marginalizing over the testimony and the data the learner receives,

$$p(h_i|h_{i-1}) = \sum_{d,t} p(d|h_{i-1})p(t|h_{i-1})\frac{p(d,t|h_i)p(h_i)}{p(d,t)}. \tag{2}$$

Beppu and Griffiths (2009) analyzed this model for two different types of testimony. They first looked at a data passing condition, where teachers selected a hypothesis from their posterior belief and then generated data consistent with that hypothesis. The learners receive a mix of data generated by their teachers and data generated from the world. Beppu and Griffiths found that at the end of the chain, the performance of individual learners was no different from the performance of learner who just received data from the world; social learning led to no long-term benefit.

Conversely Beppu and Griffiths analyzed a posterior passing condition, where teachers passed their entire posterior belief about the system. In this model, learners use their teachers' posterior belief to form their own prior belief, which is then refined using the learner's private data. This analysis showed that posterior-passing was equivalent to each teacher passing the accumulation of all of the private data witnessed previously in the chain to each new learner. A basic result in Bayesian learning gives that with increasing amounts of data, the posterior likelihood of the correct hypothesis will tend to one; social learning will lead to perfect accuracy for learners far down the chain.

These two conditions provide an upper and lower bound on the richness of testimony that can be passed between learners. However when learners pass testimony that is neither data generated from a teacher's posterior belief (undifferentiated from data from the world), nor is a full accounting of the

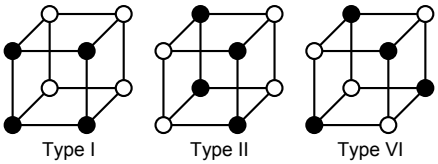

Figure 2: A visualization of three of the six category types from Shepard et al. (1961). Black dots indicate category members, the axes represents values of three binary features.

teacher's posterior belief, the outcome of iterated learning is unknown. In many cases it may be intractable to formal analysis. To understand this problem, we analyze the specific case where learners are given a category learning task and pass a "sparse" form of testimony.

Bayesian Category Learning

In a traditional category learning task, participants learn how to use the features of a set objects to place the objects into categories. In our task learners must separate eight objects which vary along three binary features into two categories of four objects each. Not all categories are equally easy to learn. Shepard, Hovland, and Jenkins (1961) classified the possible categorizations into six types and found that the the difficulty in learning these categories were in the following order: Type I $<$ Type II $<$ (Type III, Type IV, Type V) $<$ Type VI. Figure 2 gives a visualization of Types I, II, and VI. The remaining categories, Types III-V, can be described as a Type 1 category with a single exception. The difficulty of learning each category type has been replicated a number of times (e.g. Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994).

Category learning is an inference problem, where learners are given private data, d, in the form of the category membership of n objects, and are asked to infer the correct categorization of the remaining $8-n$ objects. In the iterated learning case, learners may receive social testimony t to help guide their decision. We use Bayes' rule to compute the probability of a given category c,

$$p(c|d,t) \propto p(d|c)p(t|c)p(c). \quad (3)$$

We compute this probability in three parts: the likelihood of the private data, the likelihood of the social testimony and the prior probability of each category. We consider these three parts in turn.

Private Data

We assume learners calculate the likelihood of the private data, d, that they received by assuming that each example was drawn randomly from examples of the category, c. If the categorization of any object is inconsistent with the category c, we set $p(d|c) = 0$. Otherwise, we set $p(d|c) = 1/\binom{8}{n}$ which is the chance of drawing a specific set of n objects from a set of 8 objects.

Evaluating Testimony

We analyze three models of evaluating testimony: a posterior passing model, a testimony generalization model and a no generalization model.

Posterior Passing In the posterior passing model, we assume that teachers are able to pass their entire posterior beliefs about the system on to the next learner in the chain. Learners then use their teacher's beliefs to evaluate their own private data. In this model, we assume that each learner uses their teacher's posterior belief, $p(c|d',t')$, as their prior belief, where d' and t' are the teacher's observed data and testimony. We set $p(t|c)p(c) \propto p(c|d',t')$.

Testimony Generalization In the testimony generalization model, learners receive testimony from their teacher in the form of a single, complete categorization of the objects. We assume that learners use this category to infer the teacher's beliefs about similar categorizations. Let c_t be the category given as testimony, and $d(c,c_t)$ be the number of objects that differ between c_t and c. We let $p(t|c) \propto q^{d(c,c_t)}$ where q is a free parameter between 0 and 1 that governs how much weight is placed on other hypotheses. High values of q make similar hypothesis more likely.

No Testimony Generalization In the no testimony generalization learners do not generalize support for a single category to similar categories. We set $p(t|c) = 1-\varepsilon$ if t supports the categorization c, and $p(t|c) = \delta$ otherwise.

Prior probability of categories

Past work on category learning has discovered that individuals have a non-trivial prior belief on categories, preferring "simpler" categories over more complex ones (Kemp, 2012). We assume that a learner's prior on the six category types are free parameters with the following order: $p(\text{Type I}) > p(\text{Type II}) > p(\text{Type III}), p(\text{Type VI}), p(\text{Type V}), p(\text{Type VI})$. We fit all model parameters to minimize the mean-squared error with participant responses from the two example condition of the experiment presented later in this paper.[1] The parameters of the posterior passing, testimony generalization, and no testimony generalization models were fit separately.

Model Predictions

We estimated learners' performance on a series of iterated learning chains with this category learning task. Each chain represents the average number of errors at each generation, marginalized over all possible chains. We varied the amount of private data people received (either two, four, or six examples), and the category learning type (Type I, Type II, and Type VI). To make the total number of examples shown in each chain equal across conditions we ran the two-example chain for 30 generations, the four-example chain for 15 generations and the six example chain for 10 generations. We found that the amount of private information received heavily impacted the accuracy of learner's in each chain. Accuracy

[1] Final parameter values for the posterior passing model were: $p(\text{Type I}) = .56$, $p(\text{Type II}) = .15$, $p(\text{Type III, VI, V, VI}) = .07$; testimony generalization: $p(\text{Type I}) = .56$, $p(\text{Type II}) = .15$, $p(\text{Type III, VI, V, VI}) = .07$, $q = .52$; no testimony generalization: $p(\text{Type I}) = .48$, $p(\text{Type II}) = .17$, $p(\text{Type III, VI, V, VI}) = .09$, $\varepsilon = .2$, $\delta = .01$. For all models, the priors were fit without constraints on relative ordering.

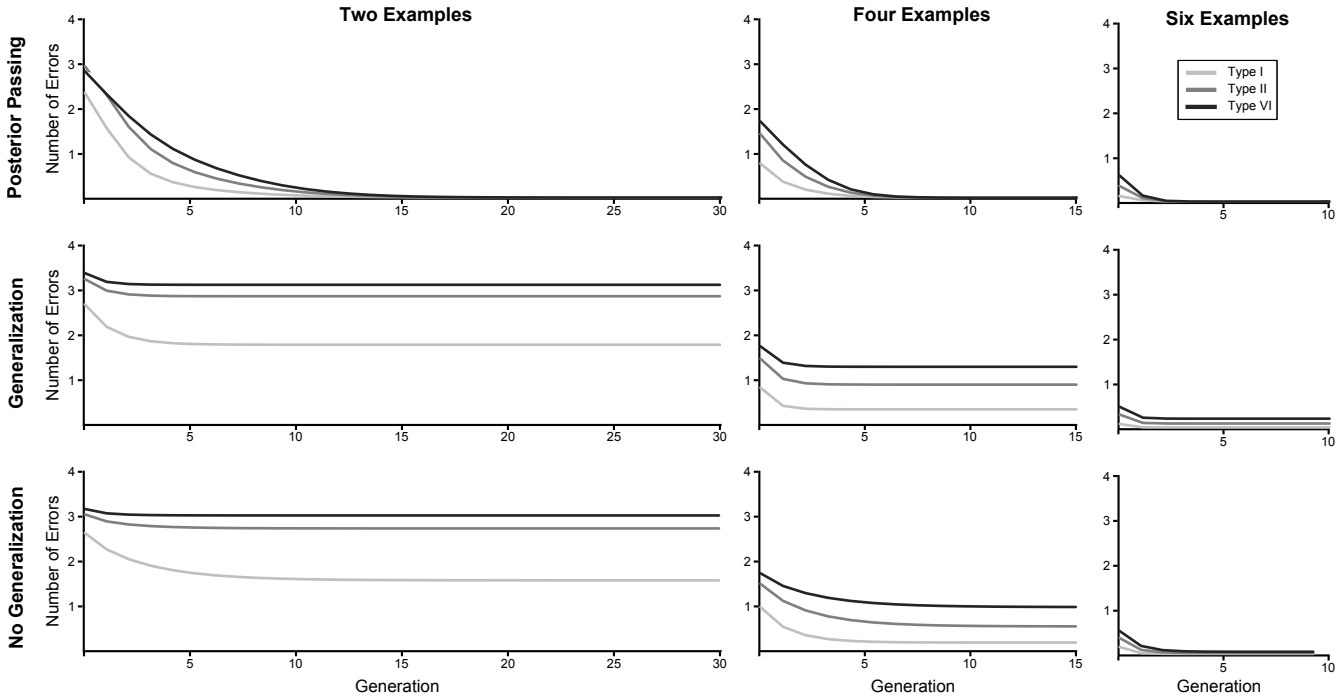

Figure 3: Model predictions for the category learning task for three forms of testimony. Model parameters were fit to the experimental data independently for each form of testimony.

also depended on the type of category being learned. Model predictions are shown in Figure 3.

The posterior passing results are consistent with previous work on Bayesian iterated learning (Beppu & Griffiths, 2009), indicating that learners will always be able to learn the correct category when passing their posterior beliefs about the world. Interestingly, this is not the case in either the generalization or no-generalization conditions. In these conditions social learning does improve the performance of the group. However, the benefit that social learning provides plateaus after a small number of generations. Our sparse form of testimony passing results in qualitatively novel long-term learning behavior. When fit to to experimental data, the generalization and no-generalization models provide similar predictions.

Testing the Model Predictions

To test our model predictions, we performed an iterated learning experiment with a category learning task. Participants were shown eight blocks that had to be split into two categories, "blickets" and "non-blickets". Participants were provided with external data by showing them between two to six of the eight blocks being placed on a "blicket detector", which provided (noisy) evidence of whether or not a block was a blicket. They also received a complete categorization of all eight blocks from the previous participant, indicating whether or not the previous participant thought each block was a blicket. This is equivalent to the type of testimony passed in both the "testimony generalization" and "no testimony generalization" models, above.

We manipulated the category being learned (Type I, II or VI), and the number of examples participants are given (either two, four, or six). To match the total number of examples shown in each chain, we ran the two example chains for 30 generations, the four example chains for 15 generations, and the six example chains for 10 generations.

Based on our model, we predict that the average number of errors will decrease over the first few generations, but will then plateau, making learners farther down the chain no more accurate than previous learners.

Methods

Participants A total of 927 participants were recruited through Amazon Mechanical Turk (`http://www.mturk.com`). Participants were compensated $0.50 for their time. They were randomly assigned to one of nine conditions: learning from Type I, II, and VI categories, and seeing either 2, 4, or 6 blocks on the machine. Five chains were run for each condition.

Stimuli and Procedure The experiment was a web-administered survey. Participants were placed in one of 45 chains (five per condition). Participants received testimony from the previous person in the chain.

During the experiment, participants were shown eight blocks and were told that some of them were blickets. The blocks varied along three binary features: cubes or spheres, blue or red, and the presence of a black or white diagonal stripe. Blickets were determined by the condition (Type I, II or IV; see Figure 2 for a visualization). To control for feature salience, we randomized how categories mapped on to a

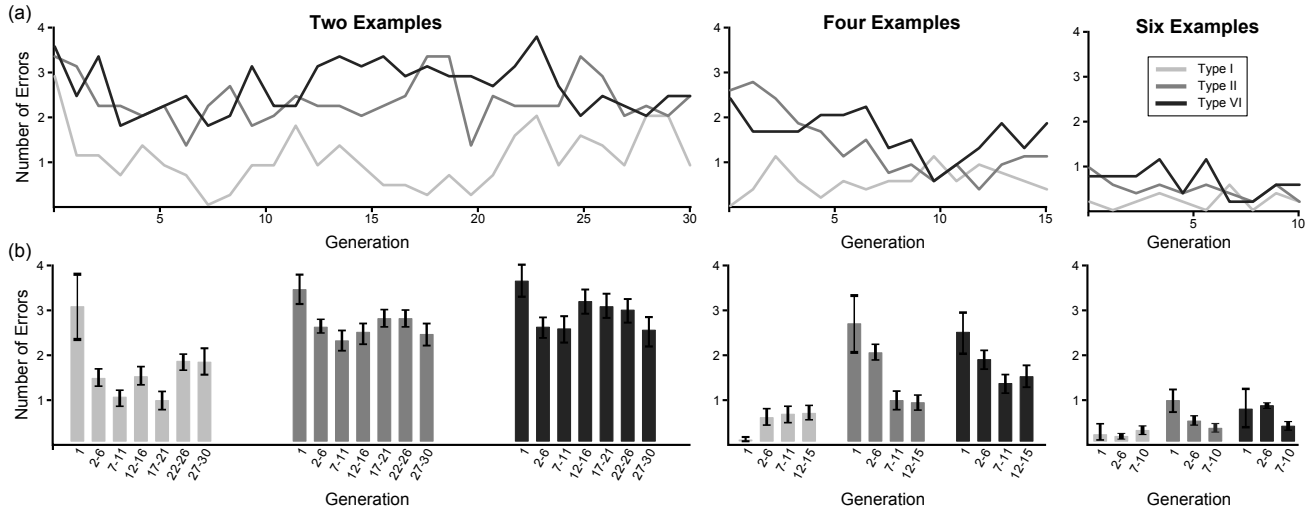

Figure 4: Results for Experiment 1. (a) Average number of errors in each generation across chains. (b) Five generation wide averages of the number of errors in each generation. Error bars represent one standard error.

specific set of features (see Love & Markman, 2003).

Participants were given testimony from the previous participant, and were told the previous participant saw two, four, or six of the blocks tested on the machine. The testimony was a categorization of the eight blocks into blickets or non-blickets. The first person in each chain received no testimony. Participants were then shown either two, four, or six of the blocks placed on a machine. Each block was placed on the machine five times. The machine lit up 90% of the time if the block was a blicket and 10% of the time otherwise.

At the end of the experiment, participants were told to choose which blocks they thought were blickets. This information was passed along to the next participant in the chain. At the end of the survey, we asked participants to give a written description of what made the blocks blickets. These answers were not passed on to the next participant and were not analyzed by the researchers.

Participants were excluded from the survey (and their results not passed on to future participants) if they either failed an attention check (how many green blocks did you see? Correct answer: zero) or inaccurately categorized three or more example blocks. A total of 93 participants failed an attention check and four were dropped for inaccuracy.

Results

We analyze changes in participant accuracy across generations, average participant accuracy compared to chance levels, and participant accuracy across learning tasks.

Iterated Improvement The model predicted that learners in all conditions should improve after the first generation, but may not substantially improve after that. To analyze this, we split up each chain into five-generation long blocks and analyzed the difference in averages between blocks.

We found that when learning from two or four examples the first social learning block performed better than the initial generation of learners. The difference between the first generation of learners and the next five generations of learners was significant in all conditions when learners received two data points (Type 1: $t(28) = 2.45$, $p < 0.01$; Type II: $t(28) = 1.73$, $p < 0.51$; Type VI: $t(28) = 1.90$, $p < .05$).[2] When learners received four data points, it seems as if there was some small iterated improvement in the first five generations. Learners in the first five generations did not significantly perform better than learners in the first generation (Type II: $t(28) = 1.23$, $p = 0.11$; Type VI: $t(28) = 1.04$, $p = 0.15$), but learners in the next five generations performed significantly better than learners in both the first generation(Type II: $t(28) = 2.5$, $p < 0.01$; Type VI: $t(28) = 1.78$, $p < 0.05$), and only significantly better than the first five generation in the Type II condition ($t(48) = 3.58$, $p < 0.001$) and only marginally better than the Type VI condition ($t(48) = 1.47$, $p = 0.07$). We also saw that when learning from six examples, social testimony did not seem to greatly increase participant accuracy, instead they had enough data to perform well on the task without using social information.

Average Participant Accuracy The benefit of social information can also be examined by examining how well participants performed on unobserved blocks. We found that participants correctly categorized the unobserved blocks above chance levels on Type I categories (two examples: $t(149) = -12.61$, $p < 0.001$; four examples: $t(149) = -12.2$, $p < 0.001$; six examples: $t(149) = -11.9$, $p < 0.001$), and Type II categories (two examples: $t(149) = -2.5$, $p < 0.01$; four examples: $t(149) = -3.6$, $p < 0.001$; six examples: $t(149) = -4.8$, $p < 0.001$) and generally above chance on Type VI categories, except for learning from two examples (two examples: $t(149) = -0.12$, $p = 0.45$; four examples: $t(149) = -1.69$, $p < 0.05$; six examples: $t(149) = -2.9$,

[2]Unless otherwise noted, statistical tests were one tailed t-tests based on the model predictions.

$p < 0.01$). This result suggests that even though continual improvement across chains may not occur, participants perform better than a single learner would on their own.

Category Learning Difficulty We also reconstruct the traditional ordering on category learning difficulty in each condition. We found that participants generally performed better on Type I categories than Type II categories (two examples: $t(298) = 7.44$, $p < .001$; four examples: $t(298) = 4.73$, $p < .001$; six examples: $t(298) = 2.28$, $p < .05$), and Type II categories than Type VI categories (two examples: $t(298) = 1.72$, $p < .05$; four examples: $t(298) = 1.29$, $p = .09$; six examples: $t(298) = 1.2$, $p < .11$). However the difference between Type II and Type VI categories was smaller than previously thought. This may be due to the way we measure participant accuracy. Unlike previous work which tends to track whether or not the participant learned the entire category correctly, we track the number of errors in their category predictions. This measurement provides substantially better scores for learning Type III, IV, and V categories which all can be closely approximated by a Type I category. It provides slightly better scores for Type II and VI categories.

Discussion

We examined how passing limited testimony can change how much knowledge accumulates in a group of learners. We presented three computational models: one looks at a maximum amount of information that can be passed, the entire posterior belief, and the other two examine learning from testimony that supports a single hypothesis. In all three models learners' accuracy will increase over time. However, unlike the posterior passing model, when learners learn from limited testimony the improvement in accuracy will plateau; after a few generations no significant improvement in accuracy will occur. The models predict that the difficulty of the task and the number of examples individuals are shown will change where this plateau is, and how how long it takes to reach it.

These predictions were confirmed by a category learning experiment. In the experiment learners who received social information performed better than learners at the beginning of each chain, who did not. We also found slight evidence for iterative improvement. When learners learned a Type II or Type VI category and received four examples, learners in the first five generations were more accurate than the initial set of learners, and that learners in the next five generations were even more accurate, but future learners were no more accurate than these learners. This pattern was mirrored (although not significantly) when learners were given six examples. The experiment confirms the model predictions: individuals who received sparse testimony were more accurate than those who receive no testimony, however in all cases the accuracy of the group plateaued; the cultural ratchet reached a point beyond which it could not "catch". Category difficulty and the concentration of data influence where and when the ratchet fails.

Our work speaks to the conditions needed for a culture to accumulate knowledge over time. If each individual in a culture sees only a limited piece of the world, difficult-to-learn ideas may never fully be learned. Our model demonstrates that a culture is then not just a shared repository of data. The information that individuals pass alters how well knowledge is accrued. These results shine light on the process by which information is gained across generations. People often times only pass along small amounts of abstract information to others. We demonstrate that even with this limited testimony, cultural knowledge can still accumulate. However we only examine a single task with a single form of testimony. This work can be extended by examining a broader range of learning tasks and richer forms of testimony passing. This will allow us to shine light on a central question: how exactly has humanity gathered such a large amount of cultural knowledge, and how can we gain even more?

Acknowledgments. This work was supported by grant number IIS-1018733 from the National Science Foundation, grant number FA9550-13-1-0170 from the Air Force Office of Scientific Research, a National Defense Science and Engineering Graduate Fellowship, and John Templeton Foundation grant #40128, Exploring the Evolutionary Foundations of Cultural Complexity, Creativity and Trust.

References

Beppu, A., & Griffiths, T. L. (2009). Iterated learning and the cultural ratchet. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2089–2094).

Boyd, R., & Richerson, P. (1988). *Culture and the evolutionary process*. University of Chicago Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148–153.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*(3), 441–480.

Kemp, C. (2012). Exploring the conceptual universe. *Psychological review*, *119*(4), 685.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. *The evolutionary emergence of language: Social function and the origins of linguistic form*, *302*, 323.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *Evolutionary Computation, IEEE Transactions on*, *5*(2), 102–110.

Love, B. C., & Markman, A. B. (2003). The nonindependence of stimulus properties in human category learning. *Memory & Cognition*, *31*(5), 790–799.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & cognition*, *22*(3), 352–369.

Perfors, A., & Navarro, D. (2011). Language evolution is shaped by the structure of the world: An iterated learning analysis. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 477–482).

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.

Tomasello, M. (1994). The question of chimpanzee culture. In R. Wrangham, W. McGrew, F. de Waal, & P. Heltne (Eds.), *Chimpanzee cultures*. Harvard University Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.