Title

The evolution of category systems within and between learners

Permalink

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Ferdinand, Vanessa
Perfors, Amy

Publication Date

2020

The evolution of category systems within and between learners

Vanessa Ferdinand (vanferdi@gmail.com)
University of Melbourne School of Psychological Sciences

Amy Perfors (perfors@gmail.com)
University of Melbourne School of Psychological Sciences

Abstract

How do cumulative cultural evolution and individual learning differ? In an abstract computational sense, both are optimisation processes that search a space of possible explanations and previous work has identified deep parallels in the mathematical models used to describe them (Suchow, Bourgin, & Griffiths, 2017). However, there are obvious differences as well: for example, individual learning involves a single agent characterised by one set of prior beliefs, representational capabilities, and so forth, while cultural evolution involves multiple agents who may vary along these factors. We argue that this difference implies that the process of cumulative cultural evolution should involve searching a more restricted set of hypotheses and converge on simpler ones. In two iterated category learning experiments, we test this prediction and find that transmission chains composed of single individuals, who learn based on their previous performance, consider both a wider variety and more complex categorisation schemas than do chains involving multiple people. **Keywords:** cumulative cultural evolution; learning; complexity, categorisation

Introduction

What is the root of human evolutionary success? Many have argued that it can be found primarily in our individual cognitive skills, ranging from domain-general representational abilities to language (e.g., Lieberman, 1991; Premack, 2007; Penn, Holyoak, & Povinelli, 2008). Others suggest that individual cognition has a more indirect effect and that the true "secret of our succcess" is cumulative cultural evolution – the ability to pass on cultural traits and skills (technology, knowledge, etc) across generations to create a cumulative ratchet effect (Henrich, 2015; Mesoudi & Thornton, 2018). The process of cumulative cultural evolution relies on individual-level skills such as the capacity to imitate and reason about other minds (Tomasello, Kruger, & Ratner, 2018) but its success has been argued to derive at least in part from characteristics of the system rather than individual learners. For instance, cultural systems may allow for greater breadth of innovation or higher overall information processing capacity than any single learner can attain (e.g., Caldwell & Millen, 2008; Muthukrishna, Shulman, Vasilescu, & Henrich, 2014).

This seems quite sensible, but it has also been argued that culture-level evolution and individual-level learning can be characterised by the same abstract mathematical and computational descriptions (Harper, 2009; Suchow et al., 2017). On the individual level, changes in beliefs are in many cases well-described as a process of Bayesian inference by which learners integrate prior knowledge with data from the world (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). On the cultural level, standard mathematical models of evolution involve replicator dynamics in which agents produce offspring in proportion to their fitness (Nowak, 2006). There are deep mathematical parallels in the form of these two equations (Harper, 2009; Suchow et al., 2017), reflecting the fact that on an abstract level both individual learning and cultural evolution are optimisation processes involving searching a set of alternatives and identifying the best of those according to some metric.

This parallel can be seen in the iterated learning framework (Smith, Kirby, & Brighton, 2003), which investigates how information changes when passed between people. The behaviour of the resulting transmission chains can be experimentally explored in the lab (Griffiths, Christian, & Kalish, 2008; Kirby, Cornish, & Smith, 2008; Claidière, Smith, Kirby, & Fagot, 2014) as well as modelled mathematically (Griffiths & Kalish, 2007; Suchow et al., 2017).

These chains are often modelled as a sequence of Bayesian agents, where each agent observes some data d from the previous agent and forms a hypothesis h about what knowledge the previous agent used to generate the data. As Bayesian reasoners, each agent performs this computation according to Bayes Rule, calculating the posterior probability of each hypothesis given the data $P(h|d)$ so that it is a function of their prior belief in that hypothesis $P(h)$ as well as the likelihood of the data given that hypothesis $P(d|h)$. This conceptualisation reduces the process of cultural transmission to a Markov chain, which means that it is possible to derive the convergence properties of the outcome of the process. In this simple case, in fact, the stationary distribution of the chain is just the prior distribution (Griffiths & Kalish, 2007).

Interestingly, at least under some assumptions, the mathematical analysis is identical whether each generation of the chain is a different agent (as in cultural evolution) or represents the same agent learning from their own previous data (Griffiths et al., 2008). One such assumption is that all agents in the chain share the same prior; this is trivially the case when the chain consists of a single learner, but is not necessarily the case when each agent is a distinct individual. If these individuals do not share a prior, then the information transmission will be systematically distorted by those with more extreme priors (Navarro, Perfors, Kary, Brown, & Donkin, 2018). A prior in this sense simply means the *a priori* distri-

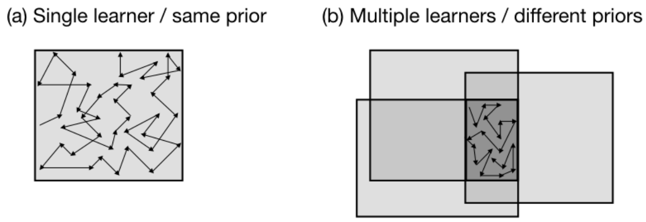

Figure 1: Schematic depiction of how individual learning and cultural evolution might search the hypothesis space differently. Each box represents the hypothesis spaces considered *a priori* by learners, and the arrows indicate the search path of the chain through the hypotheses. (a) When there is a single learner or multiple learners with the same prior, the chain searches the entire space and converges on the prior distribution. (b) When there are multiple learners, the bottleneck imposed by needing to transmit information amongst all of them means that the space effectively searched by the chain is much smaller and consists of the shared overlap between them all.

bution of weight amongst all hypotheses in the space, which can be shaped by many factors that humans vary substantially on (memory, attention, executive function, motivation, intelligence, or previous experience). It thus seems likely that information transmission through a chain of distinct agents (as in cumulative cultural evolution) might have different convergence properties and dynamics compared to the process of information transmission within a single learner over a chain of distinct learning phases.

What differences might we expect to see, and where might we expect to see them? One idea is to consider not just the convergence properties of the chain, but also the space of hypotheses investigated and the dynamics of that investigation over time. Consider the case illustrated schematically in Figure 1. When a single agent creates a transmission chain by learning from their own input — or when multiple agents with exactly the same priors do the same – the process of information transmission does not distort the search at all: any information that is capable of being represented is passed on in proportion to its prior probability and ease of representation. This is why such transmission chains converge to the prior. However, when multiple agents with different priors are in a transmission chain, the process of transmission itself affects the search. Only the set of hypotheses shared between all of the agents gets searched effectively, because only those are capable of being transmitted veridically between everyone. This is why agents with more "extreme" priors (i.e., less overlap with other agents) have a distorting effect on the chain, as found by Navarro et al. (2018).

This analysis suggests that, as long as individuals differ in their priors, transmission chains composed of distinct people (cultural evolution) should generally search a smaller part of the hypothesis space. Moreover, since simpler hypotheses are more likely to be shared by more people, cultural evolution should favour simpler hypotheses in a way that individual learning does not. We test both of these predictions here. Over the course of two experiments, we give people a category learning task by putting them into transmission chains where they either learn from another person's data (the CULTURAL condition) or where they learn from their own (the INDIVIDUAL condition). We find that, as predicted, the space of hypotheses explored by people in the cultural condition is smaller and the hypotheses considered are less complex.

Experimental design

Experiment 1

Participants 297 participants were recruited on Mechanical Turk and took our experiment online via Google App Engine. Location was restricted to the USA. Participants were paid a base rate of 1.50 USD and bonused to ensure they made 10 USD per hour. There were 90 participants in the INDIVIDUAL condition and 207 in the CULTURAL condition.

Stimuli & Method Participants were trained and tested on a mapping between 2 labels and 10 stimuli for several rounds. All stimuli had identical shape (a seashell), but each had a different fill color, varying from dark gray, RGB(25,25,25), to light grey, RBG(250,250,250), in increments of 25. One pair of labels from the following list was randomly assigned to each participant for the duration of the experiment: (buv,kal) (dap,mig) (pon, fud) (vit,lem) (seb,nuk) (gos,tef).

Each round consisted of 30 training trials and 10 testing trials. On each training trial, a shell and two possible labels for it were displayed. The participant clicked on one of the labels and received feedback with the correct label from an alien biologist who was training the participant to identify the shells on planet Zorg.[1] During the testing phase, participants received 10 trials (one per stimulus) and chose labels without feedback. This set of 10 labels will be called the *category system* that participants produced on round *x*.

Conditions The experiment consisted of two transmission conditions: CULTURAL and INDIVIDUAL. In the CULTURAL condition, each person participated in only one round of trials, and they were organized into 45 transmission chains. The first participant in each chain was trained on a random category system in which five tokens of each label were randomly assigned to the ten stimuli. Each subsequent participant in the chain was trained on the category system that the previous participant had produced on their test trials. Chains ran until convergence (i.e. when two participants produced identical category systems in a row).

In the INDIVIDUAL condition, each person participated in two to eight rounds of trials. In the first round, they were trained on a random category system as defined above. In subsequent rounds, they were trained on the most recent category system they produced (although they did not know this), constituting an iterated learning chain within one participant. Chains ran until convergence (i.e. when the participant produced identical category systems on two consecutive rounds) or until they completed eight rounds.

[1]This was the cover story told to participants, not the actual experimental setup...

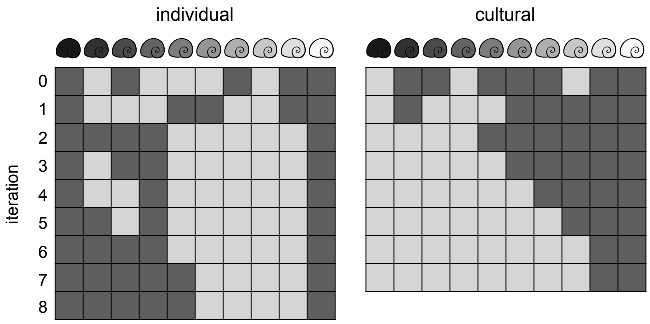

Figure 2: Example evolutionary trajectories in the INDIVIDUAL condition (left) and the CULTURAL condition (right) from Experiment 1. The ten stimuli are shown above and their labels are shown by the cells below. Dark cells mean label A was used for the stimulus and light cells mean label B was used. Iteration 0 are the random initial systems. Iteration x is the category system produced by the participant on round x in the INDIVIDUAL condition or on generation x in the CULTURAL condition. The trajectory from the INDIVIDUAL condition begins with random mapping ABABBBABAA and explores 7 systems before converging on the two-boundary system AAAAABBBBA, while the trajectory from the CULTURAL condition begins with BAABAAABAA and converges on the one-boundary system BBBBBBBBAA.

Experiment 2

In order to ensure that the results from Experiment 1 were robust, we performed a pre-registered[2] replication of it here.

Participants 313 participants were recruited under the same conditions described in Experiment 1. There were 90 participants in the individual condition and 223 in the cultural condition.

Stimuli & Method The method is identical to Experiment 1 with one exception: stimuli varied in color on a blue-purple continuum, as in Levari et al. (2018). Color values for the 10 stimuli ranged from RGB(100,0,155) to RGB(1,0,254) in R and B increments of 11.

Results

In this section, we analyze how the evolutionary trajectories produced in the CULTURAL and INDIVIDUAL conditions differed. We ask two main questions. First, what was the *breadth* of the space of category systems explored in each? Did the chains composed of multiple people tend to consider fewer hypotheses than the chains composed of a single individual? Second, what was the *complexity* of the category systems explored? Did the chains with multiple people tend to consider systems with fewer boundaries than the chains with single individuals?

In order to address these questions we must first define the space of all possible category systems. This experimental design implicitly defines $2^{10} = 1024$ possible category systems since there are ten stimuli that can each take one of two labels. These 1024 systems vary in complexity, ranging from simple (AAAAABBBBB) to interesting (ABBBAABBBA) to random-looking (ABABBAAAAB). In the following analyses, we operationalize complexity as the number of category boundaries a system has. This value ranges from zero (ex: AAAAAAAAAA) to nine (ex: ABABABABAB). There is an interesting debate about whether randomness should be considered high or low complexity (e.g., Still & Crutchfield, 2007) and this provisional operationalisation falls on the side of calling both interesting and random-looking categories "complex."

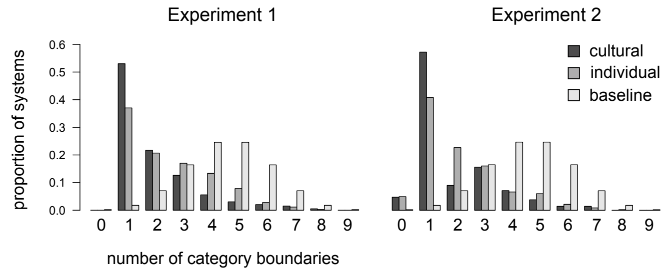

Figure 3: Proportion of category systems produced in Experiment 1 (left) and Experiment 2 (right), grouped by their number of category boundaries (x-axis), for generations 1-8 in the CULTURAL condition (dark grey), rounds 1-8 in the INDIVIDUAL condition (light grey), and the theoretically-defined baseline one would observe if the two category labels were randomly assigned to stimuli (white).

In addition to defining the space of systems and operationalising complexity, it is also useful to define a baseline to compare behaviour in both conditions against. A natural baseline to consider is what we would expect to see if participants were randomly choosing labels during the test trials. Figure 3 (baseline) shows the probability that random choice would produce a category system with k boundaries. The total number of category systems with k boundaries is $2\binom{n}{k}$ where n is the maximum number of boundaries possible.

All of the analyses that follow are restricted to the category systems obtained in the first 8 iterations of their evolutionary history (i.e. rounds 1-8 in the INDIVIDUAL condition and generations 1-8 in the CULTURAL condition). This ensures that the features we're comparing between conditions (like the number of category boundaries) had equal time to evolve.

Question 1: Exploratory breadth

Does the set of category systems explored in the INDIVIDUAL and CULTURAL transmission conditions differ from the baseline and differ from one another? Figure 3 shows the distributions of category systems from each condition and each experiment, in terms of the number of category boundaries. The set of attested systems under each transmission condi-

[2] `http://aspredicted.org/blind.php?x=ej9yg4`

Table 1: Percent of the 1024 possible category systems explored within each condition.

	INDIVIDUAL	CULTURAL
Experiment 1	19.2%	8.8%
Experiment 2	18.6%	8.3%

tion seem to exhibit less diversity than the random baseline, suggesting that there were constraints on the evolution of category systems in each of these conditions. The two conditions appear to differ from one another as well: evolutionary search in the CULTURAL condition is more focused around one-boundary systems than the INDIVIDUAL condition is.

One way to measure exploratory breadth is simply to count up the number of possible 1024 category systems that were explored in each condition. This is somewhat crude because it does not naturally compare performance to a sensible baseline (which we do below). However, it is useful to give an intuition of what is going on. As Table 1 shows, in both experiments, people in the CULTURAL condition considered fewer category systems overall than the INDIVIDUAL condition.

As a more principled measure, we also computed the ratio of explored to unexplored evolutionary space by estimating the Shannon entropy of the experimental and baseline distributions.[3] The estimates are 2.63 bits for the baseline, 2.47 bits for the INDIVIDUAL condition, and 2.17 bits for the CULTURAL condition. This means that for the INDIVIDUAL condition, 94% of the space was explored and 6% was not, while for the CULTURAL condition, 83% was explored and 17% was not. Figure 4 shows the bootstrap 95% confidence intervals around these estimates. In Experiment 1, these intervals do not overlap, providing strong evidence that all three distributions come from different generative processes. In Experiment 2, confidence intervals do not overlap with the baseline, showing that both search processes were significantly more constrained than random generation of category systems, but the two transmission conditions overlap very slightly.

Overall, it appears that the cultural transmission regime imposes more constraints on evolutionary search than the individual transmission regime does for this particular categorization task. As we address in the discussion, the next interesting step will be identifying what these constraints are.

Question 2: Category system complexity

The previous analysis suggests that individuals and cultures explore different proportions of the space of all possible hypotheses in this categorisation task. Do they also differ with respect to the complexity of the category systems they found?

[3] We used the R entropy library's minimax estimator for the experimental data and the plug-in estimate for the known baseline. Because finite data sets exhibit lower entropy than the generative process they are sampled from, entropy estimation procedures correct for this by raising the estimate more for smaller N. Our experiment had $N = 90$ INDIVIDUAL chains and $N = 45$ CULTURAL chains. It is possible that the values in Table 1 are higher for the INDIVIDUAL condition because it was sampled more. However, even after the entropy corrections were made, the CULTURAL condition still exhibits less diversity than the INDIVIDUAL condition.

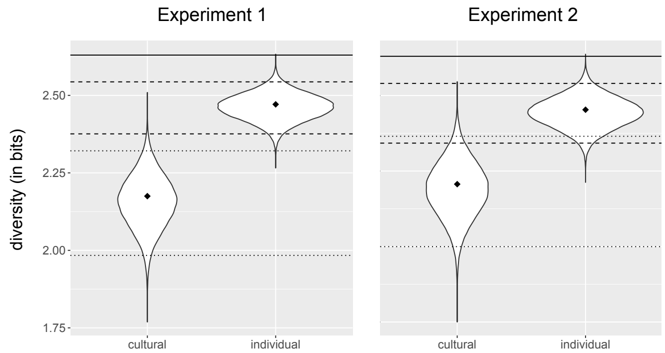

Figure 4: Differences in search space coverage between conditions in Experiment 1 (left) and Experiment 2 (right). The solid line is the maximum entropy of the search space, according to the random baseline. Diamonds are the estimated Shannon entropy of each distribution in Figure 3. Dashed lines denote the 95% confidence intervals around the INDIVIDUAL estimate and dotted lines denote those around the CULTURAL estimate. The violin plot shows the density of 100,000 bootstrap resamples of the experimental data used to compute the 95% confidence intervals.

To address this question we compared the number of category boundaries within the INDIVIDUAL and CULTURAL conditions in a linear mixed effects regression analysis. Number of category boundaries was the dependent variable while condition and iteration were the independent variables. Trajectory is included as a random effect because category systems within a chain are not independent from one another, since they share a common evolutionary history. The best-fit model was determined by systematically comparing a full model with all variables and interactions to reduced models omitting each effect, as outlined in Winter (2013).

In Experiment 1, the best-fit model incorporated condition ($\beta = 0.527, SE = 0.223, t(5) = 2.37$) and iteration ($\beta = -0.222, SE = 0.023, t(5) = -9.65$) but no interaction. In Experiment 2, the best-fit model contained condition ($\beta = -0.358, SE = 0.247, t(6) = -1.45$) and iteration ($\beta = -0.377, SE = 0.047, t(6) = -8.10$) as well as their interaction ($\beta = 0.186, SE = 0.055, t(6) = 3.41$). Thus, in both experiments, category complexity (as measured by the number of category boundaries) was significantly affected by both transmission condition and time (i.e., iteration). Complexity decreased over time in both conditions and was significantly lower in the CULTURAL condition.

All of the analyses so far have suggested that the search process differed between conditions, but do the end products differ as well? For instance, it might be the case that all trajectories in both conditions could have ended with one-boundary systems, but gotten there via very different search dynamics. To address this question, we looked at the subset of category systems that converged within the first eight iterations. Table 2 gives the percentage of systems that converged within the first 8 iterations and Table 3 breaks them down by the number

Table 2: Number of category systems that converged within the first eight iterations. A similar proportion of systems converged in both conditions.

Experiment 1		
INDIVIDUAL	79/90	88%
CULTURAL	40/45	89%
total	119/135	88%
Experiment 2		
INDIVIDUAL	74/90	82%
CULTURAL	42/45	93%
total	116/135	85%

Table 3: Number of category boundaries in the converged category systems (from Table 2). INDIVIDUAL chains tended to converge to more complex systems with more category boundaries than CULTURAL chains did.

boundaries	0	1	2	3	4	5	6	7	8	9
Experiment 1										
INDIVIDUAL	0	44	22	11	1	1	0	0	0	0
CULTURAL	0	34	5	1	0	0	0	0	0	0
Experiment 2										
INDIVIDUAL	8	49	16	1	0	0	0	0	0	0
CULTURAL	5	36	1	0	0	0	0	0	0	0

of category boundaries in the final system in the chain.

The distribution of converged category systems differ by condition in Experiment 1 ($\chi^2(12) = 30$, p = .003) and Experiment 2 ($\chi^2(12) = 30$, p = .003). This suggests that individual transmission supports evolutionary endpoints with more diversity in number of category boundaries, whereas cultural transmission largely converges to one-boundary systems.

Discussion

This work compared the dynamics of evolutionary search within and between individuals, in the domain of category learning. In two iterated learning experiments, we presented participants with a novel one-dimensional category and investigated how that category changed over time. Participants in the INDIVIDUAL transmission condition, who learned based on their own previous data, ended up searching a larger part of the overall hypothesis space in the aggregate and tended to consider more complex hypotheses with more category boundaries. In contrast, participants in CULTURAL transmission chains, who learned based on data produced by the previous participant, searched a smaller portion of the hypothesis space in the aggregate and considered simpler hypotheses with fewer category boundaries.

These findings might appear surprising in light of previous work that argues that transmission-chain models of cumulative cultural evolution and this kind of individual learning are identical (Griffiths et al., 2008; Suchow et al., 2017). However, although they are identical in the case where all agents in the population have the same prior, transmission plays a different role when they do not: in effect, it creates a filter which means that the system effectively searches only the space of hypotheses and priors that are shared by everyone.

How can we square this with experimental work that appears to show similar individual and cultural learning? For instance, Griffiths et al. (2008) presented both individuals and chains of learners with a category-learning task based on the six category types of Shepard, Hovland, and Jenkins (1961). They found that both individuals and chains showed similar behaviour for all six category types. However, unlike in our experiment (and most experiments with iterated learning chains), instead of passing on the labelled data from the previous agent, they had each agent select which of the six category types they believed was correct and generated the data from that. This subtle difference may have had two effects. First, it meant that the space of hypotheses was much smaller than ours (effectively only six) and all participants were shown all six hypotheses explicitly. To the extent that individual differences in priors or hypothesis spaces might stem from individual differences in the ease of conceptualising or thinking of them, making them explicit would have removed this source of differences. Second, having participants select hypotheses rather than pass on their data may have meant that there was little pressure toward data simplification of the sort that occurs when people's errors during learning and labelling occur asymmetrically in one direction (towards simplicity). Taken together, it is possible that there was little scope within their experiment to notice system-level differences in the complexity of the hypotheses favoured or the extent of the space searched.

To our knowledge, there is little additional experimental work beyond the present experiment that directly compares between-subject transmission chains involving multiple people (cumulative cultural evolution) to within-subject transmission chains involving a single learner. Every transmission event is an opportunity for a culturally-evolving artifact to be restructured by selection pressures (such as a learner's prior bias) or to acquire information about the environment in which it is evolving (across a network of agents vs within a single mind). Most iterated learning research focuses on understanding the behaviour of between-subject, multi-generational chains on their own (Kalish, Griffiths, & Lewandowsky, 2007; Kirby et al., 2008) or compared to a baseline involving individuals learning in what are effectively single-generation chains (e.g., Reindl & Tennie, 2018; Silvey, Kirby, & Smith, 2019). However, if we want to know exactly how cultural evolution creates cumulative changes that "no single individual could invent on their own" (Boyd & Richerson, 1996), it seems fair to compare artifacts of similar evolutionary age. The only other work we are aware of that compares the two directly is Sasaki and Biro (2017), which finds cumulatively more efficient flight paths in homing pigeons compared to individual controls matched for number of flight iterations, and Claidière et al. (2014), which found that between-subject chains of baboons created more structured patterns (tetrominos) in a pattern copying task over time than did within-subject chains (which produced more non-tetrominos). Although they did not analyse complexity explicitly, if tetrominos are of lower complexity than non-tetrominos, their results are in line with ours.

Recent work by Carr, Smith, Culbertson, and Kirby (2018) and Silvey et al. (2019) provides thorough experimental investigation of the cultural evolution of category systems. Both studies find that category system complexity[4] goes down over time when transmitted in between-subject iterated learning chains. Our CULTURAL condition replicates these findings. Silvey et al. (2019) also investigated category systems created from scratch by single individuals and found that their complexity was low and did not significantly differ from the final generation of the between-subject transmission chain. However, their individual condition wasn't designed to be directly compared to their cultural condition, as their research question was, instead, aimed at the role of communication in cultural transmission. Our experiment was designed to make this absolute comparison possible by 1) using the same initial starting conditions for all chains, 2) implementing transmission identically in the two conditions, and 3) controlling for artifact age, such that we're comparing complexity between sets of artifacts that all had the opportunity to evolve for 8 iterations. By doing this, we have shown that the process of transmission leads to a decrease in complexity over time in both conditions, but that cultural transmission chains arrive at simpler systems, and discover fewer systems overall, than individual transmission chains do.

Although our work suggests that the dynamics and nature of the information search between people (cultural evolution) may differ in important ways from the dynamics and information search within people (individual learning), this conclusion is still preliminary. Even if we view evolution and learning entirely through the simplifying lens of transmission chains, we have not considered more complex kinds of information transfer like verbal instruction or posterior passing (Beppu & Griffiths, 2009). These and other techniques may permit cultures to avoid the kind of simplifying bottleneck that drove our results. It is also possible that these dynamics would be affected by horizontal transfer within generations (e.g., Fay et al., 2018), bidirectional communication within generations (e.g., Silvey et al., 2019), and imposing external pressures for expressivity, and thus the maintenance of complexity in communication systems (e.g., Carr, Smith, Cornish, & Kirby, 2017).

In future work, we plan to focus in on the specific, cognitive differences that exist between individual and cultural transmission chains. In addition to differing in diversity of priors, as mentioned in the introduction, individual chains have memory of previous iterations, get more practice with the task at hand, and may have decreased attention or increased fatigue toward the end of the chain. Does eliminating one of these factors unleash the cumulative power of cultural transmission? For example, when individuals do not have memory of an artifact's evolutionary past, does this free them up to innovate more? Or to re-analyse a linguistic construction? For example, people with direct memory of *Watergate* in 1972, may be less likely to invent *x*-gate as a suffix meaning "scandal of type *x*", such as in *Ubergate*.[5] In addition to exploring the cognitive differences across conditions, we would also like to know how properties of the search process differ between conditions. Currently, we are developing formal models of our intuition that simple hypotheses are over-represented in the intersection of a group of learners' hypotheses, and that the process of cultural transmission over-explores this intersection.

Acknowledgments

Thanks to Charles Kemp, Kenny Smith, the members of the CoCoSci Melbourne lab, and three reviewers, two anonymous and Jon Carr, for providing useful feedback on previous versions of this work. Special thanks to Jane Ferdinand for drawing the stimuli used in the experiment. Funding was provided from Australian Research Grant DP180103600.

References

Beppu, A., & Griffiths, T. (2009). Iterated learning and the cultural ratchet. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Boyd, R., & Richerson, P. J. (1996). Why culture is common, but cultural evolution is rare. In *Proceedings-british academy* (Vol. 88, pp. 77–94).

Caldwell, C., & Millen, A. (2008). Cultural learning. *Evolution and Human Behavior*, *29*(3), 165–171.

Carr, J., Smith, K., Cornish, H., & Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science*, *41*, 892–923.

Carr, J., Smith, K., Culbertson, J., & Kirby, S. (2018). Simplicity and informativeness in semantic category systems.

Claidière, N., Smith, K., Kirby, S., & Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society B: Biological Sciences*, *281*.

Fay, N., Ellison, T., Tylen, K., Fusaroli, R., Walker, B., & Garrod, S. (2018). Applying the cultural ratchet to a social artefact: The cumulative cultural evolution of a language game. *Evolution and Human Behavior*, *39*(3), 300–309.

Griffiths, T., Christian, B., & Kalish, M. (2008). Using category structures to test iterated learning as a method for identifying inductive biases. *Cognitive Science*, *32*, 68–107.

[4]Complexity is operationalised similarly in their studies and ours, in terms of contiguity, convexity, and number of category system boundaries. Also, all studies initialized chains with category systems near the maximum complexity in their particular search spaces.

[5]This example is from Rob Boyd.

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

Harper, M. (2009). The replicator dynamic as an inference dynamic. *https://arxiv. org/pdf/0911.1763.pdf*.

Henrich, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

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

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686.

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, *360*(6396), 1465–1467.

Lieberman, P. (1991). *Uniquely human: The evolution of speech, thought, and selfless behavior*. Harvard University Press.

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1880).

Muthukrishna, M., Shulman, B., Vasilescu, V., & Henrich, J. (2014). Sociality influences cultural complexity. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1774).

Navarro, D., Perfors, A., Kary, A., Brown, S., & Donkin, C. (2018). When extremists win: Cultural transmission via iterated learning when populations are heterogeneous. *Cognitive Science*, *42*, 2108–2149.

Nowak, M. (2006). *Evolutionary dynamics: Exploring the equations of life*. Belknap Press.

Penn, D., Holyoak, K., & Povinelli, D. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2), 109–130.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences*, *104*(35), 13861–13867.

Reindl, E., & Tennie, C. (2018). Young children fail to generate an additive ratchet effect in an open-ended construction task. *PLoS ONE*, *13*(6).

Sasaki, T., & Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nature communications*, *8*(1), 1–6.

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

Silvey, C., Kirby, S., & Smith, K. (2019). Communication increases category structure and alignment only when combined with cultural transmission. *Journal of Memory and Language*, *109*, 104051.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, *9*, 371–386.

Still, S., & Crutchfield, J. P. (2007). Structure or noise? *arXiv preprint arXiv:0708.0654*.

Suchow, J., Bourgin, D., & Griffiths, T. (2017). Evolution in mind: Evolutionary dynamics, cognitive processes, and bayesian inference. *Trends in Cognitive Sciences*, *21*(7).

Tenenbaum, J., Kemp, C., Griffiths, T., & Goodman, N. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Tomasello, M., Kruger, A., & Ratner, H. (2018). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495–511.

Winter, B. (2013). Linear models and linear mixed effects models in r with linguistic applications. *arXiv preprint arXiv:1308.5499*.