Caldwell, C. A. & Millen, A. E. (2010). Conservatism in laboratory microsocieties: unpredictable payoffs accentuate group-specific traditions. *Evolution and Human Behavior*, *31*, 123-130.

Conservatism in Laboratory Microsocieties: Unpredictable Payoffs Accentuate Group-Specific Traditions

Christine A. Caldwell
Ailsa E. Millen
Department of Psychology, University of Stirling, UK

Theoretical work predicts that individuals should strategically increase their reliance on social learning when individual learning would be costly or risky, or when the payoffs for individually-learned behaviors are uncertain. Using a method known to elicit cumulative cultural evolution in the laboratory, we investigated the degree of within-group similarity, and between-group variation, in design choices made by participants under conditions of varying uncertainty about the likely effectiveness of those designs. Participants were required to build a tower from spaghetti and modeling clay, their goal being to build the tower as high as possible. In one condition, towers were measured immediately on completion, and therefore participants were able to judge the success of their design during building. In the other condition, participants' towers were measured five minutes after completion, following a deliberate attempt to test the tower's stability, making it harder for participants to judge whether an innovative solution was liable to result in a good score on the final measurement. Cultural peculiarity (i.e. the extent to which a design could be identified as belonging to a particular chain) was stronger in the delayed measure condition, indicating that participants were placing greater reliance on social learning. Furthermore in this condition there was only very weak evidence of successive improvement in performance over learner generations, whereas in the immediate measure condition there was a clear effect of steadily increasing scores on the goal measurement. Increasing the risk associated with learning for oneself may favor the development of arbitrary traditions.

1. Introduction

1.1. Background

The term "cumulative cultural evolution" refers to situations in which social transmission allows for successive improvements to performance over generations of learners, generated by the accumulation of modifications to the transmitted behaviors (Boyd & Richerson, 1996; Caldwell & Millen 2008a; Tomasello et al., 1993). Cumulative culture presents a mechanism by which humans can inherit aspects of phenotypic flexibility, producing a "system for the inheritance of acquired variation" (Boyd & Richerson, 1988, p30). Such accumulated traditions therefore offer a powerful means by which populations can adapt to their local surroundings. Boyd and Richerson (1996) have noted that the human capacity for cumulative culture has allowed our species to occupy a much wider range of habitats than any other animal.

The phenomenon of cumulative culture depends on a combination of social learning and innovation (Tomasello, 1999). Individuals must occasionally invent novel strategies which improve on existing methods, and these modifications must be faithfully transmitted in order to be maintained within the population. Theoretical work predicts that individuals should make use of social information in a strategic way (Laland, 2004), relying more heavily on social learning when individual learning

is likely to be risky or costly, or when there is good reason to believe that the socially acquired information is highly reliable (for example, because it comes from a more experienced, or prestigious, individual, or because it represents the choice of the majority), (see also Boyd & Richerson, 1985). Previous experiments demonstrate that human participants do strategically alter their reliance on social information in line with the predictions of these models (e.g. Kameda and Nakanishi, 2002; McElreath et al., 2005; Mesoudi, 2008).

Particularly relevant from the point of view of the current study are experiments which show that participants rely more heavily on social information when individual judgments are likely to be inaccurate. Deutsch and Gerard (1955) found that the more uncertain an individual was about the correctness of their judgment, the more susceptible they were to social influences on their decisions. Likewise, Baron et al. (1996) found that as task difficulty increased, participants were more likely to conform to inaccurate group norms generated by confederates. In an explicit test of evolutionary predictions concerning the strategic use of social information, McElreath et al. (2005) presented participants with a two-armed bandit decision task, finding that they more often opted to view the choices of others as the difficulty of predicting the best option increased. Nonetheless, it remains unclear what the outcome of a greater reliance on social learning might be, in terms of emerging traditions.

In the current study, we aim to investigate the emergence of spontaneous cultural traditions under controlled laboratory conditions. Mesoudi (2007; Mesoudi & Whiten, 2008) has advocated experimental approaches to the study of culture, emphasizing the potential value of such studies over those using observational methods, or mathematical models and computer simulations. In experiments, variables of interest can be manipulated, and the record of cultural change is more complete than is generally possible with real world traditions. At the same time, studies with real human participants are liable to have greater validity than mathematical models or computer simulations. Using a method known to elicit cumulative cultural evolution in the laboratory (demonstrated in Caldwell & Millen, 2008b), we intend to examine cumulative culture under two contrasting conditions. In one of our conditions participants will be readily able to judge the success of their own solution, and therefore the likely payoffs. In the other condition this will be more difficult for participants to judge. We expect that participants in the latter condition will show greater reliance on social learning, in line with previous theoretical and empirical research.

Caldwell and Millen (2008b) documented the emergence of apparently arbitrary design traditions within chains of participants ("laboratory microsocieties", e.g. Baum et al., 2004). Designs produced by participants from the same chain were more similar to one another than those which had been built by participants from different chains. In the current experiment, stronger evidence of within-group similarity and between-group variation is expected in the condition under which success on the task is harder to predict. Using Caldwell and Millen's (2008b) tower building task, we have generated two conditions with differing levels of uncertainty about likely payoffs, by instructing participants to work towards slightly different goals. In one condition, participants are to be instructed that towers will be measured immediately on completion, and in the other, they are to be informed that towers will be measured following a delay and intentional disturbance. Participants in the latter condition are therefore being implicitly encouraged to build towers that are relatively stable and resilient, probably at the expense of some initial height. Although the

ostensive goals of the experimental groups are different, designs can nonetheless be directly compared with one another, as towers in both conditions will in fact be treated identically and the experimenter will always take both measurements.

1.2. Hypotheses

1.2.1. Hypothesis 1 – Cumulative culture. Consistent with Caldwell and Millen (2008b), it is predicted the later generations in each chain will typically produce better solutions (in terms of the goal measure of height), compared with earlier generations.

1.2.2. Hypothesis 2 – Adaptive specialization. Adaptive specialization, as a result of cumulative culture, will be assessed in several ways. Firstly, the heights of the towers for both the goal and non-goal measure will be used. It is predicted that, for each measure, towers will be higher in the condition for which that measure was made salient. So, for the immediate measurement, towers in the immediate measure salient (IMS) condition are expected to be higher than those in the delayed measure salient (DMS) condition. For the delayed measurement, towers in the DMS condition are expected to be higher than towers in the IMS condition. Since this specialization is expected to happen as a result of cumulative learning, it is also predicted that such effects will be stronger in later generations, compared with earlier ones. Also, adaptive specialization is likely to influence design choices. Designs from the same experimental condition are predicted to be more similar to one another than designs from different conditions, i.e. those from the IMS condition will tend to be more similar to one another than they are to those from the DMS condition, and vice versa.

1.2.3. Hypothesis 3 – Design traditions. Following Caldwell and Millen (2008b) it is predicted that designs will be more similar to those from the same chain, compared with those from different chains within the same experimental condition.

1.2.4. Hypothesis 4 – Stronger design traditions in DMS condition. It is predicted that there will be a greater difference between the within- and between-chain similarity ratings for the DMS condition, compared with the IMS condition, indicating greater social influence in this condition.

2. Methods

2.1. Participants

Participants were recruited on campus at the University of Stirling. For each of the two experimental conditions, ten chains of ten participants took part. Their mean age was 21.70 years (*SD*=6.47, youngest=17, eldest=56). Sixty-one males took part, and 139 females. The majority of participants (150) were Psychology undergraduate students, taking part in return for a research participation credit, but the remainder (50) took part in exchange for a £3 participation fee.

Two chains (20 participants) of the DMS condition consisted entirely of participants recruited outside of the Psychology participant pool, and who therefore received the fee rather than the credit. Apart from these two chains, the distribution of paid participants across the two conditions was very similar. In the eight remaining DMS chains there were a total of 16 paid participants. Three chains consisted entirely of students taking part for a participation credit, and the remaining five contained between one and six paid participants. In the IMS condition there were a total of 14 paid participants. Four chains consisted entirely of students taking part for a participation credit, and the remaining six contained between one and six paid participants.

All participants (whether given the credit or the fee) were also given a monetary performance incentive to encourage them to score as well on the task as possible (see Procedure).

Ethical approval for this research was provided by the University of Stirling Department of Psychology Ethics Committee. The procedure was explained to all participants in advance, and they each gave written consent to participation.

2.2. Materials

Each participant was provided a standard 500g packet of uncooked spaghetti and 78g of modeling clay (Early Learning Centre "Modelling Material"). They were also provided with a rectangular melamine tray (42cm x 35cm) onto which they were to build their tower.

2.3. Apparatus

A Sony DCR-HC94 camcorder mounted on a tripod was used to record the testing area. A tape measure was fixed to the wall so that the height of towers could be measured with ease (both in real time, and from the video record later on). A twelve inch desk fan was placed on a table in the testing area. Figure 1 depicts the testing area.

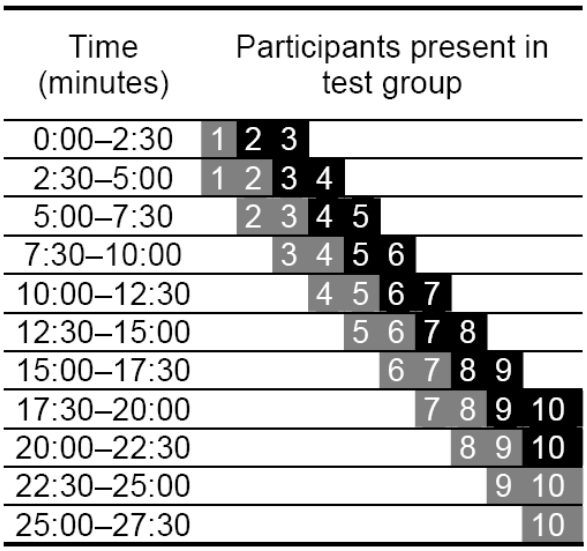
Fig. 1. The testing area. Participants were seated on the floor in front of a table, and built their towers on trays. A measuring tape ran from the floor to the tabletop, and then from the tabletop up the wall. Completed towers were moved from the floor, on their trays, to the tabletop. A desk fan was located on the tabletop, directed towards the towers.

2.4. Procedure

Participants were randomly assigned to the positions 1 to 10 in each chain. The participants were informed that they were about to take part in a team challenge, and that they would be called in turn to engage in the task. In order to simulate generational succession, we used the replacement method pioneered by Jacobs & Campbell (1961). Participants start times were staggered, such that every two and a half minutes a new person entered the test group (see Fig. 2 for information on test

group composition at any given time). While they were in the test group, each participant had five minutes of observation time, during which they could watch the previous participants building their tower, followed by five minutes of building time, during which they had to construct their own tower. Once their time was up, they left the test group. The staggered start and finish times had the effect that, at any given time (except at the very start and very end of any given chain) there were four individuals together in the test group, two of whom were observing, and two of whom were actually engaged in the task (see Fig. 2). So, for example, a chain would begin with participant 1 building their tower, with participants 2 and 3 observing. Then, two and a half minutes in, participant 2 would also start building, and participant 4 would join the test group as an observer.

Fig. 2. Schematic of test group membership at different time points over the course of a single complete trial. Shading indicates the role of the participant: black indicates observing, and grey indicates building.



While they waited their turn to join the group in the test area, participants sat together in an adjoining area from which the test area could not be seen. When participants joined the group in the testing area they were provided with written instructions about the nature of the task. They were informed of the aim of the task. For participants in the IMS condition this was simply to build a tower that was as high as possible, which would be measured immediately on completion. For participants in the DMS condition this was to build a tower that would be as high as possible following a five minute delay after completion. Participants in the DMS condition were also made aware that during this five minute delay the tower would be moved from its position on the floor to a new location on a table, where it would be directly in the stream of a desk fan.

All participants were informed that they would receive an incentive fee in proportion with their performance on the task (a penny for each centimeter of height achieved on their goal measure, which resulted in participants receiving an average of 53 pence in addition to their participation fee or credit). For IMS participants it was the immediate measurement which determined their payment, and for DMS participants it was the delayed measurement. Participants were also informed of their time restrictions (five minutes of observation time followed by five minutes in which

to build their own tower), and were told that they were permitted to communicate with other members of the test group regarding the task, as well as learning from observing others.

Within the test group, participants were kept aware of their current role (observing, constructing), and the time elapsed, by a computer display and reminders from the experimenter. Once an individual's five minute construction period was up, their tower could be evaluated. For participants in the IMS condition, the experimenter recorded the height of the tower immediately upon completion and reported this score to the participant. For participants in the DMS condition, this immediate measurement was taken from the video record after the experiment had been completed. The experimenter took a photograph of the tower following its completion. For participants in both conditions, the tower was then moved (complete with the tray on which it had been built) from its current position on the floor, to the top of the table. Whilst participants in the DMS condition were made aware that this was a crucial part of the evaluation process, those in the IMS condition were simply informed that the tower was being moved to create room for the later participants in the test group. Each tower remained on the table for five minutes, during which time it stood in the stream of a desk fan. Again this detail of the procedure was made salient for the participants in the DMS condition, but was not mentioned to those in the IMS condition. At the end of the five minute delay, towers from the DMS condition were measured by the experimenter, who reported the height to the participant. For the IMS towers, this measurement was taken from the video. After the five minute delay, towers were removed and dismantled to create room for further towers. This meant that participants in the test group could always view the two most recently completed towers, as well as those currently under construction. Participants left the testing area once they had been informed of their tower's height. Examples of the towers produced by participants are provided as Supplementary Information (Fig. S1).

2.5. Similarity ratings

Photographs were taken of all of the towers that had been produced by participants, and these were rated by a naïve coder. The rater was given each of the photographs from the set one-by-one, and asked to rate it in comparison to all of the others. The comparison photographs were randomly ordered each time, for each target photograph. The rater was informed to attend only to the structural similarity of the designs, and that backgrounds and colors were to be ignored. Ratings were given on a seven point scale, with a rating of 7 indicating the most similar photographs, and a rating of 1 the least similar ones. The reliability of the similarity ratings of the photographs could be readily assessed from our resulting dataset, as we had two ratings for every comparison. For each such pair of ratings, one resulted from comparing photograph A as a target, in comparison with photograph B (along with all of the others), and the other resulted from comparing photograph B as a target, in comparison with photograph A (along with all of the others). Therefore there were a total of 39,800 ratings, giving 19,900 pairs of comparisons. These pairs of comparisons were correlated in order to establish the reliability of the ratings of similarity ($r = 0.817$, $N = 19900$, $p < 0.0005$).

3. Results

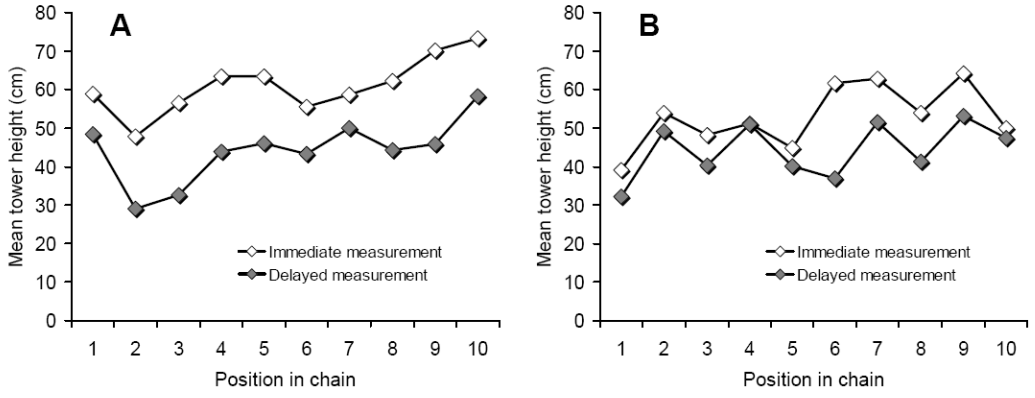
As noted in the Methods, two chains of participants from the DMS condition consisted entirely of participants who had been recruited from outside the Psychology participant pool. The data reported here include these chains. However, all statistics reported were also carried out having excluded these two chains, with no change to the pattern of results.

3.1. Hypothesis 1 – Cumulative culture

It was predicted that there would be evidence of cumulative learning in the two experimental conditions in relation to their respective goal measures (i.e. the immediate measure for the IMS participants and the delayed measure for the DMS participants). Page's *L* Trend Test (Page, 1963) was used to analyze the complete data over the ten generations to explicitly test for successive improvement (Caldwell & Millen, 2008b). Consistent with the hypothesis, this was significant for the IMS participants using the immediate measure (L = 3249, k = 10, N = 10, p = 0.005). Although not predicted, cumulative improvement was also found for IMS participants using the delayed measure (L = 3212.5, k = 10, N = 10, p = 0.016) although this was not the goal measurement that they had been instructed to aim for. The immediate and delayed measures were strongly correlated however, for both conditions (IMS condition: r = 0.408, N = 100, $p < 0.0005$; DMS condition: r = 0.691, N = 100, $p <$ 0.0005), so success on the immediate measure was in fact a good predictor of success on the delayed measure.

The effect of improvement over generations did not reach significance for the DMS participants, when using their goal (delayed) measure (L = 3108.5, k = 10, N = 10, p = 0.169). Despite this (and although this was not explicitly predicted) there was a significant effect of improvement over generations when using the immediate measure (L = 3170, k = 10, N = 10, p = 0.048). Figure 3 displays the data for the IMS (panel A) and DMS (panel B) conditions, using both the immediate and delayed measurements.

Fig. 3. Mean height of towers produced by participants in the IMS (immediate measure salient) condition (panel A), and in the DMS (delayed measure salient) condition (panel B).



3.2. Hypothesis 2 - Adaptive specialization

3.2.1. Evidence of adaptive specialization from height measurements. In order to compare differential effects for early and late generations in the chain of participants, and therefore the role of cumulative learning in any contrasts found, data were collapsed across the first three (positions 1, 2 and 3) and last three (8, 9 and 10) participants in each chain. Table 1 displays the means and standard deviations. Tests indicated that the distribution of these values was not significantly different from the normal distribution. These collapsed values therefore allowed us to perform a 2x2x2 ANOVA, with generation (early and late) and measurement type (immediate and delayed) as repeated measures variables, and experimental condition (IMS and DMS) as a between-subjects variable.

The ANOVA identified a main effect of measurement type ($F_{1,18} = 47.350$, $p < 0.0005$) with the towers being significantly shorter for the delayed measure. There was also a main effect of generation ($F_{1,18} = 9.346$, $p = 0.007$), with the towers being taller for the later generations, compared with the earlier ones. There was no main effect of experimental condition ($F_{1,18} = 0.828$, $p = 0.375$), so neither condition showed an overall advantage in terms of the height of their towers. There was no interaction between measurement type and generation ($F_{1,18} = 0.110$, $p = 0.744$), and no interaction between generation and experimental condition ($F_{1,18} = 0.640$, $p = 0.434$). There was a significant interaction between measurement type and experimental condition ($F_{1,18} = 8.049$, $p = 0.011$), with the delayed measure having a stronger detrimental effect on the towers built by participants in the IMS condition. However there was no three-way interaction between measurement type, generation and experimental condition ($F_{1,18} = 0.006$, $p = 0.941$).

3.2.2. Evidence for adaptive specialization from design similarity. Using the similarity ratings obtained (see Methods) each tower was given a score for a score for within-condition similarity and between-condition similarity. In order to generate the within-condition similarity scores, a mean rating was calculated based on the comparisons of each individual tower with the 90 others from different chains in the same experimental condition. The between-condition similarity scores were calculated based on the comparisons of each tower with the 100 others in the other condition. Tests indicated that the distribution of these measures was significantly different from the normal distribution so non-parametric statistics were applied. The median similarity rating when comparing towers with those from the same experimental condition was 3.57 (lower quartile 2.15, upper quartile 4.05). The median when comparing those from different experimental conditions was 3.58 (lower quartile 2.13, upper quartile 4.04). These were not significantly different (Wilcoxon test: $Z = 0.322$, $N = 200$, $p = 0.748$), therefore there was no evidence that designs from particular conditions overall conformed to specific types.

3.3. Hypothesis 3 – Design traditions

The similarity ratings were also used to address the hypothesis concerning design traditions within chains (and variation between them). We calculated the mean similarity rating for each tower in relation to the nine towers from the same chain, and in relation to the 90 towers from other chains in the same experimental condition. Tests indicated that the distribution of these measures was significantly different from the normal distribution so non-parametric statistics were applied.

For the towers built by participants in the IMS condition, the median similarity rating within chains was 3.72 (lower quartile 2.78, upper quartile 4.67). The median when comparing those from different experimental conditions was 3.67 (lower

quartile 2.13, upper quartile 4.08). These were significantly different (Wilcoxon test: $Z = 4.160$, $N = 100$, $p < 0.0005$), so designs from the same chain were rated as more similar to one another compared with designs from different chains.

For the towers built by participants in the DMS condition, the median similarity rating within chains was 4.33 (lower quartile 2.67, upper quartile 5.44). The median when comparing those from different experimental conditions was 3.51 (lower quartile 2.10, upper quartile 4.03). These were significantly different (Wilcoxon test: $Z = 6.504$, $N = 100$, $p < 0.0005$), so designs from the same chain were rated as more similar to one another compared with designs from different chains.

3.4. Hypothesis 4 – Stronger design traditions in DMS condition

It was also predicted that there would be evidence of stronger design traditions in the DMS condition, compared with the IMS condition, as indicated by higher within-chain similarity in relation to between-chain similarity. In order to test this hypothesis, each individual tower was given a *cultural peculiarity* score. Similarity ratings were given on a scale of 1-7 (as detailed in Methods) so these were first transformed by subtracting 1 from each to give a scale of 0-6. This scale was then used to calculate each tower's within-chain similarity in relation to its between-chain similarity. A proportion [within-chain similarity / (within-chain similarity + between-chain similarity)] was calculated, such that the range of possible values would run from 0 to 1, with values greater than 0.5 indicating a higher within-chain similarity rating, and values less than 0.5 indicating a greater between-chain similarity rating. This proportion was calculated for each individual tower, the resulting score providing an indication of the likelihood of identifying that particular tower as coming from that particular chain (hence its cultural peculiarity). Tests indicated that the distribution of these cultural peculiarity scores was significantly different from the normal distribution so non-parametric statistics were applied.

The median cultural peculiarity score for the IMS condition was 0.548 (lower quartile 0.477, upper quartile 0.606). The median cultural peculiarity score for the DMS condition was 0.591 (lower quartile 0.532, upper quartile 0.626). As predicted, the cultural peculiarity scores were significantly higher in the DMS condition (Mann-Whitney U test: $U = 3891$, $n_1 = 100$, $n_2 = 100$, $p = 0.007$). Pictures of the towers built by two of the DMS chains are provided as Supplementary Information (Fig. S1).

A more stringent test of this hypothesis would involve treating the *chains* as the unit of analysis, rather than each individual tower. Given that designs were more similar within chains than between them, the result reported above could potentially be driven by a very small minority of chains in the DMS condition using designs that were particularly readily copied. In order to perform an analysis which treated the chains as the unit of analysis, the mean cultural peculiarity score was calculated for each chain, and these were used to compare across the two experimental conditions (giving $n = 10$ for each). Tests indicated that the distribution of these values was significantly different from the normal distribution so non-parametric statistics were applied. The median cultural peculiarity score for the IMS condition was 0.521 (lower quartile 0.496, upper quartile 0.546). The median cultural peculiarity score for the DMS condition was 0.561 (lower quartile 0.532, upper quartile 0.599). The cultural peculiarity scores remained significantly higher in the DMS condition using this more conservative analysis (Mann-Whitney U test: $U = 24$, $n_1 = 10$, $n_2 = 10$, $p = 0.049$).

4. Discussion

We expected to find evidence of cumulative culture in both conditions, using their respective goal measures. In fact this was only apparent for the IMS condition. In the DMS condition there was evidence for improvement over generations on the non-goal (immediate) measurement, but not on the goal (delayed) measure.

We also expected to find evidence of adaptive specialization as a result of cumulative culture, i.e. that designs would become better tailored towards the particular measure over generations. Although a two-way interaction between measurement type and experimental condition indicated a fit of design to measurement type, the three-way interaction involving generational effects was not significant. So any fit of design to measurement type in this experiment appeared to be attributable to individual decision making rather than cumulative learning.

We did find strong evidence of design traditions, as towers were rated as more similar to those from the same chain, compared with those from different chains. Furthermore, this effect appeared to be stronger in the DMS condition, as predicted. The greater difference between the within-chain and between-chain similarity revealed stronger design traditions in this condition. Interestingly, similarity effects also appeared to be restricted to within-chain influences, rather than demands of the experimental condition, as there was no overall difference between the designs produced by participants in the two conditions.

The apparent absence of cumulative effects in the DMS condition is intriguing, especially given that significant cumulation was nonetheless found for the non-goal (immediate) measure. These results provoke consideration of the conditions necessary for cumulation to occur. The imperfect information available to those in the DMS condition made it difficult to judge whether an innovation would be effective (i.e. that it would result in an improvement on the goal delayed measure). The inhibition of innovation is likely to be connected to both the greater reliance on social information in this condition, and the lack of cumulation. As noted in the introduction, social learning and innovation are both necessary for cumulative cultural evolution. Mesoudi (2008) has also found evidence to suggest that an increase in social learning within a population can inhibit exploration of an adaptive landscape of possible design choices, reducing the chances of finding globally optimal designs. All the same, it should be noted that we did not find a significant interaction between experimental condition and generation effects so these interpretations must be viewed cautiously. It is not possible to conclude with certainty that our two experimental conditions were different in this regard.

Our findings are consistent with the notion that learners should rely more heavily on social information when "decision-making forces" (e.g. Richerson & Boyd, 2005, p116) are weak. In situations where payoffs are transparent, little cultural heritability is expected. In contrast, when payoffs are difficult to judge, or feedback about payoffs is delayed until after an individual's own decisions have been made, greater heritability is likely. Choices which are made as relatively once-in-a-lifetime decisions, such as the choice of a particular University, or career, are more likely to show cultural heritability compared with decisions which can much more easily be revised and adapted on the basis of immediate and transparent payoff information, such as which filling station to buy fuel from. Our findings may help to generate novel predictions about the circumstances under which arbitrary group variation should be expected in real human populations.

Acknowledgements

This project was funded by a research grant from the Economic and Social Research Council (RES-061-23-0072). We are grateful to the staff and students of the University of Stirling for their cooperation and participation, and to Beth Richardson who carried out the ratings of the photographs.

References

Baron, R. S., Vandello, J. A. & Brunsman, B. (1996). The forgotten variable in conformity research: impact of task importance on social influence. *Journal of Personality and Social Psychology, 71,* 915-927.

Baum, W. M., Richerson, P. J., Efferson, C. M., & Paciotti, B. M. (2004). Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. *Evolution and Human Behavior, 25*, 305-326.

Boyd, R. & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: Chicago University Press.

Boyd, R. & Richerson, P. J. (1988). An evolutionary model of social learning: the effects of spatial and temporal variation. In T. Zentall & B. G. Galef (Eds.), *Social learning: A psychological and biological approach* (pp. 29-48). Hillsdale, NJ: Erlbaum.

Boyd, R. & Richerson, P. J. (1996). Why culture is common but cultural evolution is rare. *Proceedings of the British Academy, 88*, 77-93.

Caldwell, C. A. & Millen, A. E. (2008a). Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society B, 363,* 3529–3539.

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

Deutsch, M. & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology, 51,* 629-636.

Jacobs, R. C., & Campbell, D. T. (1961). The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. *Journal of Abnormal and Social Psychology, 62,* 649-658.

Kameda, T. & Nakanishi, D. (2002). Cost-benefit analysis of social/cultural learning in a nonstationary uncertain environment: an evolutionary simulation and an experiment with human subjects. *Evolution and Human Behavior, 23,* 373-393.

Laland, K. N. (2004). Social learning strategies. *Learning and Behavior, 32,* 4-14.

McElreath, R., Lubell, M., Richerson, P. J., Waring, T. M., Baum, W., Edsten, E., Efferson, C., & Paciotti, B. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior, 26*, 483-508.

Mesoudi, A. (2007). Using the methods of experimental social psychology to study cultural evolution. *Journal of Social, Evolutionary, and Cultural Psychology, 1,* 35-58.

Mesoudi, A. (2008). An experimental simulation of the "copy-successful-individuals" cultural learning strategy: adaptive landscapes, producer-scrounger dynamics, and informational access costs. *Evolution and Human Behavior, 29*, 350-363.

Mesoudi, A. & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London Series B, 363,* 3489–3501.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: a significance test for linear ranks. *Journal of the American Statistical Association, 58*, 216-230.

Richerson, P. J. & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago University Press.

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

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