

Caldwell, C. A. & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, 29, 165-171 (doi:10.1016/j.evolhumbehav.2007.12.001).

## Experimental Models for Testing Hypotheses about Cumulative Cultural Evolution

Christine A. Caldwell

Ailsa E. Millen

Department of Psychology, University of Stirling, UK

The rapid appearance (over evolutionary time) of the cognitive skills and complex inventions of modern humans has been attributed to “cumulative cultural evolution” (henceforth CCE), the accumulation of knowledge and skills over generations. To date, researchers have only been able to speculate about the reasons for the apparent absence of this phenomenon in nonhumans, and it has not been possible to test hypotheses regarding the mechanisms underlying it. Here we show that it is possible to demonstrate CCE under laboratory conditions, by simulating generational succession through the repeated removal and replacement of human participants within experimental groups. We created “microsocieties” in which participants were instructed to complete simple tasks using everyday materials. In one of our procedures, participants were instructed to build a paper aeroplane which flew as far as possible, and in the other, they were instructed to construct a tower of spaghetti which was as tall as possible. We show that, in both cases, information accumulates within the groups such that later generations produce designs which are more successful than earlier ones. These methods offer researchers a window to understanding CCE, allowing for experimental manipulation and hypothesis testing.

### Introduction

The rapid appearance (over evolutionary time) of the cognitive skills and complex inventions of modern humans has been attributed to “cumulative cultural evolution” (henceforth CCE), (Boyd & Richerson, 1996; Richerson & Boyd, 2005; Tomasello, 1999). The term CCE is used to describe the way that knowledge accumulates in human populations over time, such that each generation makes use of behaviours and artefacts invented by previous generations, which they would be unlikely to have been able to invent by themselves (Boyd & Richerson, 1996; Richerson & Boyd, 2005; Tomasello, 1999). It has been argued that, although social learning is relatively common in the animal kingdom, CCE is extremely rare, possibly restricted to humans (Boyd & Richerson, 1996; Galef, 1992; Tomasello, 1999). It has also been suggested that CCE may even be dependent on learning mechanisms which are unique to humans, and is consequently not possible in nonhumans (Tomasello, 1999), although this remains contentious (e.g. see Whiten et al, 2003). Understanding CCE may therefore represent an important element in understanding human nature, particularly as it has allowed humans to develop powerful technologies, assemble complex societies, use symbolic forms of communication, and exploit an unusually wide range of habitats (Boyd & Richerson, 1996). However, to date, research on CCE has been restricted to historical approaches, such as those which classify and sequence human artefacts (Basalla, 1989; O’Brien et al, 2001), and comparative approaches, which draw comparisons between human behaviour and that of other animals, such as chimpanzees (Boesch, 2003; Tomasello et al, 1993; Whiten et al, 2003). Therefore, researchers have only been able to speculate about the reasons for its apparent absence

(or at least its relative rarity, e.g. see Boesch, 2003) in other species, and the abilities upon which it depends in humans.

This has led to considerable debate (Whiten, 2005), and little consensus. For example, Boyd and Richerson (1996) and Tomasello (1999; Tomasello et al, 1993) have suggested that CCE may depend on specific social learning mechanisms, in particular imitation and/or teaching. Since imitation and teaching have traditionally proven notoriously difficult to identify in animals (Caldwell & Whiten, 2002; Caro & Hauser, 1992), their arguments have provided a conveniently neat explanation for the apparent absence of CCE. However, this has not gone undisputed (Heyes, 1993; Laland & Hoppitt, 2003). Laland (2004) has suggested that CCE may instead depend on an ability to appraise the relative effectiveness of behavioural alternatives, and that this could be beyond the capabilities of nonhumans. In contrast, Whiten (2005; Whiten et al., 2003) has proposed that the crucial factor may be the unusual complexity of human behaviours, and that this accounts for CCE, rather than particular social learning mechanisms.

Our aim was to demonstrate that CCE could be studied under laboratory conditions. Such a demonstration would allow this debate to move from theoretical speculation into the realms of empirical testability. We therefore wanted to show that improvement in performance on a task could be passed on within groups, over miniaturised “generations” of learners. We used a micro-society design, in which generational succession is simulated through the repeated removal and replacement of participants within groups (e.g. Baum et al, 2004; Jacobs & Campbell, 1961). This method also has similarities with the transmission chain method, originally pioneered by Bartlett (1932), and more recently applied by Mesoudi (e.g. Mesoudi & Whiten, 2004; Mesoudi, Whiten & Dunbar, 2006; Mesoudi, 2007). We presented groups of participants with challenges involving the construction of simple artefacts. In one of our tasks, groups of participants were asked to build a paper aeroplane from a sheet of paper that would fly as far as possible. In the other task, ten groups of participants were asked to construct a tower from spaghetti and modelling clay, which was as high as possible. It was predicted that the performance of these chains of individuals would improve over successive generations. It was also predicted that the artefacts produced by participants would themselves show physical evidence of social learning, in that structures would be more similar within chains than across them.

## Methods

*Participants.* Participants were recruited on campus at the University of Stirling, and from two local secondary schools. For the paper planes study, ten chains of ten participants took part. Their mean age was 20 years ( $SD=6.02$ , youngest=13, eldest=48), and the ratio of males to females was approximately 50:50 (53 males, 47 females). Ten chains of ten participants also took part in the spaghetti towers study. Their mean age was 21 years ( $SD=7.14$ , youngest=11, eldest=47), and the ratio of males to females was approximately 40:60 (39 males, 61 females). As the participants for both studies were drawn from the same pool of participants (predominantly undergraduates from the University of Stirling) it is possible that some individuals took part in both parts. We did not consider it necessary to exclude individuals from participating in one study if they had already taken part in the other.

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.

*Materials.* Paper plane builders were provided with a single sheet of A4 paper, and spaghetti tower builders were provided with a standard 500g packet of spaghetti and 78g of modelling clay (Early Learning Centre “Modelling Material”).

*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, the participants’ start times were staggered, such that every two and a half minutes a new person entered the group (see Table 1 for information on 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 artefact, followed by five minutes of building time, during which they had to construct their own artefact. 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 group, two of whom were observing, and two of whom were actually engaged in the task (see Table 1). So, for example, a chain would begin with participant 1 building their artefact, 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 group as an observer. The aim was to simulate a miniaturised society, in which one generation would have the opportunity to interact with and observe individuals from the previous two generations, but not those further back. However, we did retain all artefacts for inspection by later participants, to reflect the more permanent record generated by material culture.

Table 1  
Group composition over time in the micro-society design

Time (min)	Group composition (by participant number)	
	Observing	Constructing
0.00–2.30	2, 3	1
2.30–5.00	3, 4	1, 2
5.00–7.30	4, 5	2, 3
7.30–10.00	5, 6	3, 4
10.00–12.30	6, 7	4, 5
12.30–15.00	7, 8	5, 6
15.00–17.50	8, 9	6, 7
17.50–20.00	9, 10	7, 8
20.00–22.30	10	8, 9
22.30–25.00		9, 10
25.00–27.30		10

Note. Generational succession is simulated through the repeated removal of experienced participants and introduction of naïve participants. Each row of the table shows the group composition at any given time, made up of observing participants and participants actually engaged in the task (constructing).

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 (i.e. to build a paper aeroplane that flew as far as they could make it go, or to build a spaghetti tower that was as high as they could build it) and of their time restrictions (five minutes of observation time followed by five minutes in which to build their own artefact). They were also informed that they were permitted to communicate with other members of the group regarding the task, and that they were allowed to observe and learn from 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 artefact was evaluated. For spaghetti towers, this involved the experimenter measuring the height of the tower after it had been standing for 30s. For paper aeroplanes, this involved the participant throwing their plane three times, with the experimenter recording the distances flown, and then taking the best of the three measurements (to allow for mis-throws). This would generally take less than 30s, so feedback to the group was fairly rapid, and it was possible for later participants to revise their designs on the basis of this feedback. The artefacts were then retained for display, for later members of the group to inspect, and the experimenter wrote down the measurements next to each, so that this information was also available. Participants left the testing area once their artefact had been evaluated.

Examples of the artefacts produced by participants are provided as Supplementary Information (Figs. S1 and S2).

*Similarity Ratings.* We also wished to know whether there was evidence for cultural variation between groups, as well as successive improvement over generations, as this would be another emergent feature of CCE. Photographs were therefore taken of all of the artefacts that had been produced by participants, and these were rated by naïve coders. Each rater was given one of the photographs from the set and asked to rate it in comparison to all of the others. There were therefore a total of 100 raters for the paper aeroplane photographs, and 97 raters for the spaghetti tower photographs (three photographs were missing from this particular set and therefore could not be rated). The comparison photographs were randomly ordered for each rater. The raters were provided with written instructions, informing them that we were only interested in the structural similarity of the designs, and that backgrounds (and, in the case of the spaghetti towers, also colours) were to be ignored. They were provided with a seven point scale, which they were to refer to in making their ratings. A rating of 7 indicated the most similar photographs, and a rating of 1 the least similar ones. Each individual rater therefore provided a total of 99 ratings, rating a particular target photograph against all of the others. Inter-rater reliability for the similarity ratings of the photographs could therefore be readily assessed from our dataset, as for every comparison we in fact had two ratings. In each case, one of these ratings came from a rater who had been provided with photograph A as a target, and had rated it in comparison to photograph B along with all of the others, and the other came from a rater who had been provided with photograph B as a target, and had rated it in comparison to photograph A along with all of the others. Therefore there were a total of 9,900 ratings, giving 4950 pairs of comparisons, for a full set of photographs (in fact this was slightly lower in both cases due to three photographs being missing from the spaghetti tower set, and one rating missing from the paper aeroplane data). These pairs of comparisons were correlated in order to establish the reliability of the ratings

of similarity. For the paper planes  $r = 0.387$ ,  $n = 4949$ ,  $p < 0.001$ , and for the spaghetti towers  $r = 0.449$ ,  $n = 4656$ ,  $p < 0.001$ . The ratings were therefore significantly correlated in both cases, although it should be noted that these correlations are relatively weak, as it was clearly relatively difficult to objectively judge the similarity of these photographs.

## Results

Consistent with the idea that information accumulated within the groups, later artefacts were more successful in terms of the goal measures (of distance flown and height) than earlier ones. The planes produced by the first participants in each group flew a mean best distance of 226.5cm ( $SD = 104.1$ ), and those produced by the tenth participants flew a mean best distance of 714.6cm ( $SD = 334.9$ ). A related  $t$  test confirmed a significant difference between the first and tenth planes ( $t_9 = 4.706$ ,  $p = 0.001$ , two-tailed). The towers produced by the first participants in each group had a mean height of 33.5cm ( $SD = 20.3$ ), and those produced by the tenth participants had a mean height of 53.3cm ( $SD = 25.6$ ). The towers also showed a significant improvement when comparing first and tenth artefacts, although this was only significant as a one-tailed test (related  $t$  test:  $t_9 = 1.864$ ,  $p = 0.048$ , one-tailed).

Page's  $L$  Trend Test (Page, 1963) was used to analyse the complete data over the ten generations to explicitly test for successive improvement. There was a highly significant effect for paper planes ( $L = 3351$ ,  $k = 10$ ,  $n = 10$ ,  $p < 0.001$ ) and for spaghetti towers ( $L = 3285.5$ ,  $k = 10$ ,  $n = 10$ ,  $p = 0.001$ ). Figure 1 displays the data for both planes (panel A) and towers (panel B). The complete data, showing each individual chain's performance, is provided as Supplementary Information (Fig. S3).

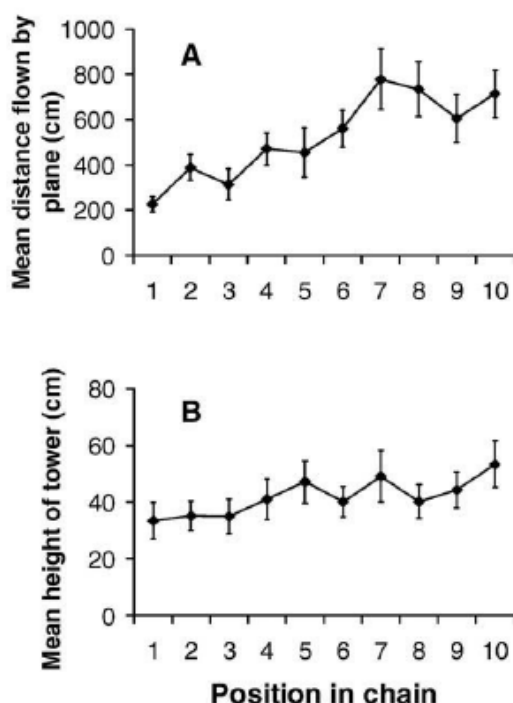


Fig. 1. Measures of success over generations (error bars indicate  $\pm$ S.E.M.). Panel A displays the data for paper aeroplanes, and Panel B displays the data for spaghetti towers.

Using the ratings provided by the naïve coders (see Methods) we calculated the mean similarity rating for each artefact in relation to artefacts from the same chain ( $n = 9$  for each individual artefact), and in relation to artefacts from other chains ( $n = 90$  for each individual artefact). The mean similarity rating for paper planes (of which there were a total of 100, so  $n = 100$ ) within chains was 3.34, ( $SD = 0.85$ ), compared with 3.09 ( $SD = 0.56$ ) between chains. The planes were rated as being significantly more similar within chains than between them (related  $t$  test:  $t_{99} = 3.413$ ,  $p = 0.001$ , two-tailed). The mean similarity rating for spaghetti towers within chains was 3.37 ( $SD = 1.14$ ), compared with 2.72 ( $SD = 0.74$ ) between chains. The towers were significantly more similar within chains than between them (related  $t$  test:  $t_{96} = 6.622$ ,  $p < 0.001$ , two-tailed). As previously, the  $n$  for the towers is slightly lower on account of the missing photographs.

If a process of descent with modification was being observed, we would also expect that, within chains, artefacts would be more similar to those closer to them in the chains, compared with those that were more distant. For the purposes of comparing the ratings of artefacts which were close and distant within the chains, only artefacts from positions 1-3 and 8-10 were used for this analysis (artefacts 4-7 were excluded as they were not sufficiently distant from others in the chain for a useful comparison to be made). The mean similarity rating was calculated for each photograph under six conditions: close (0-2 positions apart in chain), intermediate (3-5 positions apart), and far (6-9 positions apart), both within and between chains. A 3x2 two-way repeated measures ANOVA was used to analyse the data. For paper planes, it was found that there was a main effect of distance, such that those close together were more similar than those far apart,  $F_{2,118} = 16.570$ ,  $p < 0.001$ , as well as a main effect of chain, such that those from the same chain were more similar than those from different chains,  $F_{1,59} = 5.595$ ,  $p = 0.021$ . There was also an interaction between chain and distance, confirming that the distance effect was stronger within chains, compared with between  $F_{2,118} = 9.583$ ,  $p < 0.001$ . For spaghetti towers, it was found that there was a main effect of distance, such that those close together were more similar than those far apart,  $F_{2,114} = 7.108$ ,  $p = 0.001$ , as well as a main effect of chain, such that those from the same chain were more similar than those from different chains,  $F_{1,57} = 21.220$ ,  $p < 0.001$ . The trend towards an interaction was in the same direction as that for the paper aeroplanes, although this was not significant for the spaghetti towers,  $F_{2,114} = 1.413$ ,  $p = 0.247$ . The data are displayed in Fig. 2. The pairwise comparisons shown in Fig. 2 were carried out using related  $t$  tests. Comparisons marked with three ( $p < 0.005$ ) or four ( $p < 0.001$ ) asterisks remain significant following correction of the alpha level to take account of multiple comparisons (Bonferroni correction for nine comparisons: alpha = 0.0056). All comparisons were two-tailed.

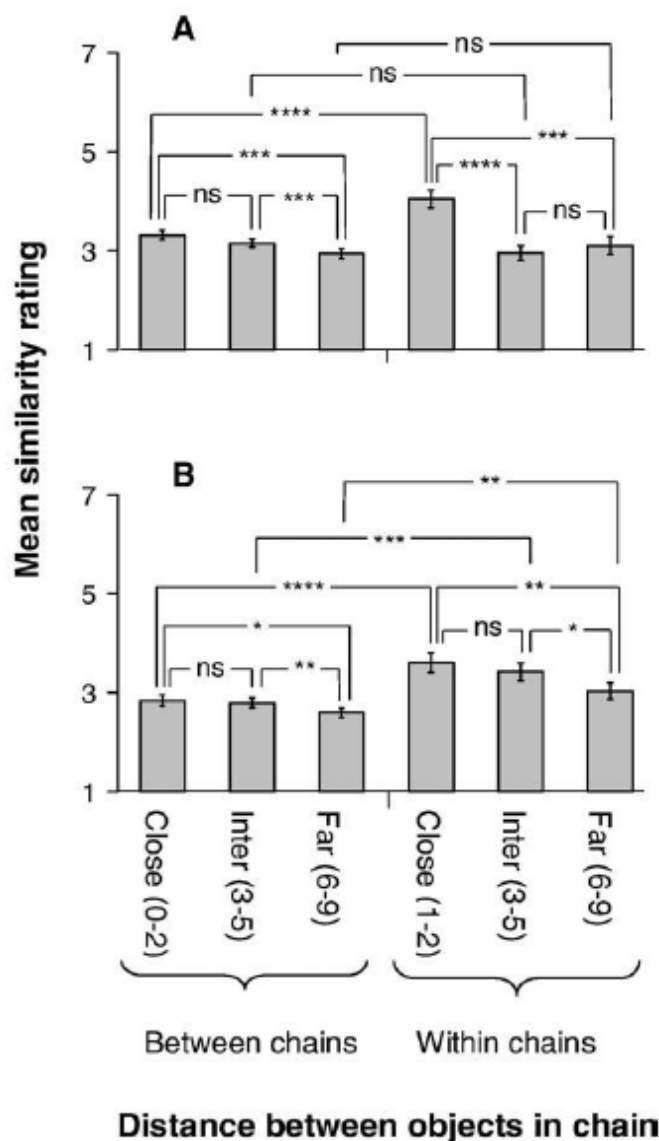


Fig. 2. Ratings of similarity for pairs of artefacts. Pairs were rated from the same (within) and different (between) chains, for those which were close, intermediate, and far apart, in terms of their positions in the chains. Panel A displays the data for paper aeroplanes, and Panel B displays the data for spaghetti towers. Pairwise comparisons were carried out between conditions, which are indicated with brackets and asterisks ( $*p < .05$ ;  $**p < .01$ ,  $***p < .005$ ,  $****p < .001$ ). Error bars indicate  $\pm$ S.E.M.

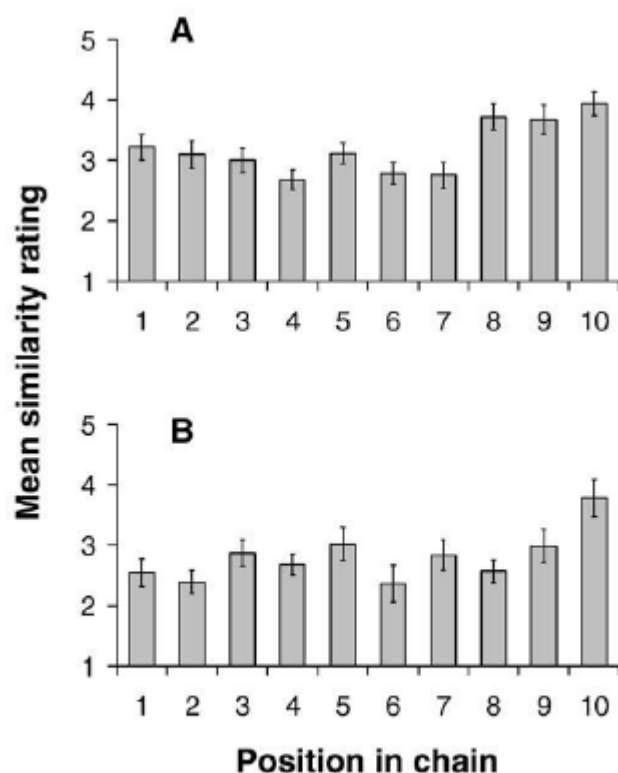


Fig. 3. Ratings of similarity for pairs of artefacts in the same position across chains. Panel A displays the data for paper aeroplanes, and Panel B displays the data for spaghetti towers. Error bars indicate  $\pm$ S.E.M.

The similarity ratings also allowed us to test for convergent evolution. If certain designs were inherently more successful, then these designs should be favoured in all chains, and later designs should be more similar to one another across chains, as compared with earlier designs. In order to analyse this, we took the similarity ratings for all pairs of photographs which were in the same position across chains. There were therefore 450 pairs of photographs considered in this analysis, 45 pairs for each position in the chain (although again this number was slightly lower for the towers due to the missing photographs). For each position in the chain, the mean similarity was calculated, and these are displayed in Fig. 3. One-way ANOVA was used to determine whether there were differences in similarity across positions. For the planes,  $F_{9,449} = 4.688$ ,  $p < 0.001$ . Fisher's *LSD* was used to determine which conditions were significantly different from one another. Planes in position 1 were significantly less similar than planes in position 10. Planes in positions 2, 3, 4, 5, 6 and 7, were all significantly less similar than planes in positions 8, 9 and 10. No other comparisons were significant. A positive correlation between position and similarity confirmed the upward trend towards greater similarity later in the chains of paper planes,  $r = 0.173$ ,  $n = 450$ ,  $p < 0.001$ . The one-way ANOVA was also significant for the towers:  $F_{9,422} = 2.682$ ,  $p = 0.005$ . Fisher's *LSD* revealed that the towers in position 10 were significantly more similar than the planes in positions 2, 3, 4, 5, 6, 7, 8 and 9. No other comparisons were significant. A positive correlation between position and similarity again confirmed the upward trend towards greater similarity later in the chains of spaghetti towers,  $r = 0.137$ ,  $n = 423$ ,  $p = 0.005$ .



Due to the design of the experiments, certain individuals in the chains were actually able to interact with a greater number of fellow participants than others. For instance, participant 1 would only meet participants 2, 3 and 4, whereas participant 4 would meet participants 1, 2, 3, 5, 6 and 7, and participant 10 would only meet participants 7, 8 and 9, whereas participant 7 would meet participants 4, 5, 6, 8, 9 and 10 (see Table 1 for further information). It is possible that greater input favoured success on the task, so this was also analysed. Twenty planes from our set (positions 1 and 10) were produced by participants which had access to a total of three potential advisors, twenty were produced by participants with four potential advisors (positions 2 and 9), a further twenty were produced by participants with five potential advisors (positions 3 and 8), and forty were produced by participants with six potential advisors (positions 4, 5, 6 and 7). The pattern was the same for the towers. One-way ANOVA was used to analyse the influence of number of advisors. However, there was no significant effect of number of advisors for either planes ( $F_{3,99} = 0.546$ ,  $p = 0.652$ ) or towers ( $F_{3,99} = 0.423$ ,  $p = 0.737$ ).

## Discussion

We conclude that our results show that it is possible to demonstrate CCE under controlled laboratory conditions. We found improvement in performance over generations of learners within the group, consistent with the accumulation of knowledge and skills (or a “ratchet effect”, Tomasello, 1999). We also found evidence of descent with modification within the chains, as we found clear evidence that designs were more similar within chains than across them, and also that designs that were closer together were more similar than those that were far apart. We also found evidence of convergent evolution, with later designs (in different chains) being more similar than earlier ones.

Interestingly, the finding that the artefacts show convergent evolution may help to explain why artefacts that are closer together are rated as more similar, even across chains (a significant main effect was found for distance between artefacts in the analysis of similarity ratings, see Results, and also Fig. 2). Clearly, artefacts towards the end of the chain have design features in common, even across chains (see Fig. 3), and even the *absence* of some of these successful features may result in artefacts earlier in the chain being more similar to one another than they are to later designs.

We have been able to demonstrate these phenomena using two very distinct tasks, which helps to confirm that the process that we are observing is a very general one. We selected the two particular tasks for a variety of reasons, and therefore they necessarily share certain key attributes. Each task can be evaluated using a highly objective measure of success, and each is also easy and simple enough to complete within a short time frame, and yet difficult enough for participants to gain real benefits from opportunities for social learning. However, there are also important differences between the two tasks. Firstly, whilst many people have some prior experience of having built a paper aeroplane, the spaghetti tower task is far more novel and participants have few preconceived ideas about how to approach the task. Secondly, whilst feedback on performance on the spaghetti tower task is continual during construction (participants can of course see exactly how high their tower currently is), feedback on the performance on the paper aeroplane task is delayed until construction is complete. We believe that these differences between the tasks are quite fundamental, and we are therefore pleased that the process can be demonstrated using both of these procedures. This is consistent with the idea that CCE is a pervasive

phenomenon throughout human society (Boyd & Richerson, 1996; Richerson & Boyd, 2005; Tomasello, 1999).

Interestingly, the differences between the tasks may lend themselves to testing hypotheses based on theories of social learning. Less prior experience ought to provoke greater reliance on social learning (e.g. Boyd & Richerson, 1985). The spaghetti towers task would therefore be predicted to show greater fidelity within the chains, compared with the paper aeroplanes. Likewise, the paper aeroplanes might be expected to show a greater rate of accumulation (consistent with novel input from individuals). It is somewhat difficult to make quantitative statistical comparisons across the different tasks in order to test this directly. Firstly, it is likely that raters are using very different notional criteria when rating the 'similarity' of a pair of paper aeroplanes, than they are when rating the 'similarity' of a pair of spaghetti towers. Also, with regard to the rates of accumulation, likewise it is difficult to determine what degree of 'improvement' in a paper aeroplane would correspond to particular degree of 'improvement' in a spaghetti tower, as they may not improve in a linear fashion.

All the same, informal inspection of the data on similarity and accumulation suggests that these predictions may be upheld. Looking at the similarity ratings, both the paper planes and spaghetti towers were given similar similarity ratings within chains. However, the difference in the ratings given for the between chains comparisons (mean = 3.09,  $SD = 0.56$ , for the paper planes, and mean = 2.72,  $SD = 0.74$ , for the spaghetti towers) suggests that, on average, unrelated planes were being rated as more similar than were unrelated towers. Therefore, it seems that the spaghetti towers were showing a greater effect of within-chain fidelity, when compared with typical ratings across chains. Also, cautious inspection of the rates of accumulation (Fig. 1) suggests that the paper aeroplanes may have shown a more rapid ratcheting effect. Further experiments actively manipulating levels of prior experience with these tasks would prove extremely illuminating, and offer another possibility for application of these methods.

In conclusion, we expect that the methods that we have described will prove to be extremely useful in addressing the debates surrounding CCE. The procedures described will allow researchers to run controlled experiments under laboratory conditions, within which manipulations can be carried out to test for the influence of different learning conditions, and the availability of different sources, and amounts, of information.

### **Acknowledgements**

This work was supported by a grant from the ESRC to C.A.C. (RES-061-23-0072). We thank C. Hill and J. Davies for help with data collection and pilot work. We are very grateful for the cooperation of staff at Lornshill High School and Grangemouth High School.

### **Supplementary Information**

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.evolhumbehav.2007.12.001.

## References

- Basalla, G. (1989). *The evolution of technology*. Cambridge: Cambridge University Press.
- Bartlett, F. C. (1932). *Remembering*. Oxford: Macmillan.
- 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.
- Boesch, C. (2003). Is culture a golden barrier between human and chimpanzee? *Evolutionary Anthropology*, 12, 82-91.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: Chicago University Press.
- 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., & Whiten, A. (2002). Evolutionary perspectives on imitation: Is a comparative psychology of social learning possible? *Animal Cognition*, 5, 193-208.
- Caro, T. M., & Hauser, M. D. (1992). Is there teaching in nonhuman animals? *Quarterly Review of Biology*, 67, 151-174.
- Galef, B. G. (1992). The question of animal culture. *Human Nature*, 3, 157-178.
- Heyes, C. M. (1993). Imitation, culture and cognition. *Animal Behaviour*, 46, 999-1010.
- 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.
- Laland, K. N., & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology*, 12, 150-159.
- Laland, K. N. (2004). Social learning strategies. *Learning and Behavior*, 32, 4-14.
- 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., & Whiten, A. (2004). The hierarchical transformation of event knowledge in human cultural transmission. *Journal of Cognition and Culture*, 4, 1-24.
- Mesoudi, A., Whiten, A. & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, 97, 405-423.
- O'Brien, M. J., Darwent, J., & Lyman, L. (2001). Cladistics is useful for reconstructing archaeological phylogenies: Paleoindian points from the Southeastern United States. *Journal of Archaeological Science*, 28, 1115-1136.
- 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: 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.
- Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993). Imitative learning on objects by children, chimpanzees, and enculturated chimpanzees. *Child Development*, 64, 1688-1705.

Caldwell, C. A. & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, 29, 165-171 (doi:10.1016/j.evolhumbehav.2007.12.001).

Whiten, A. (2005). The second inheritance system of chimpanzees and humans. *Nature*, 437, 52-55.

Whiten, A., Horner, V., & Marshall-Pescini, S. (2003). Cultural panthropology. *Evolutionary Anthropology*, 12, 92-105.