Accepted for publication in *Proceedings of the Royal Society B: Biological Sciences*: https://doi.org/10.1098/rspb.2020.1885

The value of teaching increases with tool complexity in cumulative cultural evolution

Amanda J. Lucas[1], Michael Kings[1], Devi Whittle[1], Emma Davey[1], Francesca Happé[2],
Christine A. Caldwell[3] & Alex Thornton[1*]

[1] Centre for Ecology and Conservation, University of Exeter, Penryn Campus, TR10 9FE, UK

[2] Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, SE5 8AF, UK

[3] Department of Psychology, University of Stirling. FK9 4LA, UK

*Author for correspondence: alex.thornton@exeter.ac.uk

Abstract

Human cumulative cultural evolution (CCE) is recognised as a powerful ecological and evolutionary force, but its origins are poorly understood. The longstanding view that CCE requires specialised social learning processes such as teaching has recently come under question, and cannot explain why such processes evolved in the first place. An alternative, but largely untested, hypothesis is that these processes gradually co-evolved with an increasing reliance on complex tools. To address this, we used large-scale transmission chain experiments (624 participants), to examine the role of different learning processes in generating cumulative improvements in two tool types of differing complexity. Both tool types increased in efficacy across experimental generations, but teaching only provided an advantage for the more complex tools. Moreover, while the simple tools tended to converge on a common design, the more complex tools maintained a diversity of designs. These findings indicate that the emergence of cumulative culture is not strictly dependent on, but may generate selection for, teaching. As reliance on increasingly complex tools grew, so too would selection for teaching, facilitating the increasingly open-ended evolution of cultural artefacts.

Keywords: coevolution; cumulative cultural evolution; social learning; teaching; tool-making

1. Introduction

Progressive improvements in tools, technologies and institutions enabled human populations to spread around the world and ushered in the Anthropocene, shaping not only our own evolution but also that of other species [1,2]. These far-reaching consequences have inspired a large body of research into the behavioural, cognitive and neural mechanisms through which humans transmit and build on cultural information (reviewed in [3–5]). Nevertheless, despite much theorising, the mechanisms that enabled the initial emergence of cumulative culture in the human lineage remain poorly understood.

For many authors, cumulative culture represents a Rubicon between humans and all other animals [2,6–8]. While it is clear that animals across a range of taxa exhibit socially learned cultural traditions [9–11], the cultural achievements of our species have no obvious parallel in nature. Various explanations for this apparent human uniqueness have pinpointed cognitive processes such as episodic memory [12], metacognition [13] and technical reasoning [5] as potential prerequisites for CCE, but the most influential focus on the importance of high-fidelity social learning. In particular, processes such as imitation and active teaching, thought to be restricted or absent in other species, are often argued to be necessary in order to transmit information faithfully and so preserve and build upon innovations [6,14,15]. However, current evidence is limited and contradictory, with bodies of theoretical and empirical research seemingly supporting [7,16–18] or contradicting [19–23] the theory. More fundamentally, theories stipulating specialised human learning processes as prerequisites for CCE fail to explain why such processes evolved in the first place, and do little to advance our understanding of the initial emergence of the phenomenon.

An alternative, gradualist approach considers the “fully fledged” CCE seen in modern human populations as the outcome of a long history of co-evolutionary feedback loops (c.f. [24–26]). At its core, CCE involves sequential improvements in the performance of an innovation over successive rounds of cultural transmission (“core criteria” for CCE, as defined by Mesoudi & Thornton [27]). Adding to previous circumstantial evidence (reviewed in [11]), a number of experimental studies now provide compelling evidence that some non-human animals fulfil these core criteria [23,28,29]. For instance, iterated bouts of social learning can allow homing pigeons to find the optimal, shortest route between two points [29]. Thus, it is possible that relatively simple, phylogenetically conserved learning processes akin to those found in other animals may have allowed ancestral hominins to produce modest, sequential improvements to simple tools. These tools, like those manufactured by our great ape relatives, are likely to have been made of perishable materials that leave no trace in the archaeological record. As reliance on these increasingly complex tools grew, so too would the selection pressure for social learning processes that facilitated the transmission of high-performing innovations that would be difficult for individuals to invent from scratch. Over time, such co-evolutionary feedback could eventually enable the production of tools whose mode of production and causal structure is opaque, or difficult to ascertain through emulation of existing artefacts alone [30]. Thus, rather than simply solving problems with a single, optimal solution (as in the pigeon example [29]), CCE could begin to open up design space and facilitate the open-ended diversity that characterises modern human culture. This open-endedness reflects Mesoudi & Thornton's "extended criteria" for CCE (see [27] for details), which to date have only been observed in humans.

Theoretical models demonstrate the plausibility of the argument that increasing tool complexity generates selection for high-fidelity social learning processes [25,31], but relevant empirical data is lacking. In particular, we lack clear evidence of the central assumption that the value of specialised learning processes in generating cumulative improvements increases as artefacts become more complex. For instance, one recent experiment showed that while participants could copy simple knot designs through emulation alone, they required teaching from an expert when the design was complex [32]. However, this study did not examine accumulation of improvements. To examine the potential for cumulative culture, researchers commonly use transmission chain experiments, in which participants solve a task or produce an artefact and are gradually replaced by new participants, who have opportunities to learn from their predecessors. Here, each round represents a "generation" and improvements in performance across successive generations are indicative of CCE [4]. It is notable that, across different transmission chain experiments, high fidelity social learning processes such as teaching or imitation were necessary to preserve or improve performance in tasks that seem relatively complex (e.g. flint knapping [33] or making virtual fishing nets [16]), but not in apparently simpler tasks like building paper aeroplanes [19], spaghetti towers [21] or home-made baskets [20]. This seems superficially consistent with the argument that increasing tool complexity generates selection for high-fidelity processes. However, we must be cautious in comparing these different tasks because (1) we have no objective measures of task complexity and (2) the studies employed very different procedures. To address this important gap in the literature, here we present the first study to examine the role of different learning processes in generating cumulative improvements in two types of tool that differ in their degree of complexity (as defined by the relative causal opacity of their mode of production).

In our experiment, participants were tasked with building a tool to carry as many marbles as possible: either (a) a floating container made from a single sheet of waterproof paper or (b) a carrying container made from pipe-cleaners. We chose these two tool types for their differences in causal opacity; while the paper tools are relatively simple and easy to copy by inspecting previously made exemplars (i.e. via emulation), pipe-cleaners can be attached together in a wide variety of different ways and their "furriness" makes it difficult to see how the individual elements join and overlap. Pilot studies confirmed the differences in opacity: naïve participants could readily reproduce paper tools simply by inspecting them, but needed the original maker to teach them to accurately reproduce pipe-cleaner tools (see "Pilot study" in supplementary material).

Within each tool category, we divided participants into transmission chain groups where experienced individuals were gradually replaced with new, naïve group members over a series of ten "generations". There were three social learning conditions: *Emulation*, *Imitation* and *Teaching*. In the *Emulation* condition, participants could inspect the tools made by previous chain members and were informed of each tool's performance score. In the *Imitation* condition, participants could observe previous chain members making their tools (and were also made aware of the tools' performance scores), while in the *Teaching* condition individuals that had finished building used verbal communication to help subsequent chain members. In addition we ran an *Asocial* learning condition, where participants built 10 successive tools with no opportunities to learn from others.

To address the co-evolutionary hypothesis for the emergence of CCE, we made five key predictions. First (1), given the hypothesis that CCE can emerge in the absence of high-fidelity social learning processes, we predicted that cumulative improvements in tools would arise across all social learning conditions, as well as in the asocial condition where individuals could learn from their own prior experiences (c.f. [3]). Second (2), if selection for high-fidelity processes arises as tools become more causally opaque, we predicted that imitation or teaching would only provide any advantage in generating cumulative improvements in the pipe-cleaner tool task, generating steeper slopes of improvement across generations compared to the emulation treatment. Specifically, we predicted that these processes would facilitate the transmission of high-performing innovations in pipe-cleaner tool design, generating successors that (3) also performed well and (4) were similar in design. Finally (5), we predicted that paper tools would tend to converge on similar designs, reflecting cases of CCE where there is a single peak in the adaptive landscape, whereas pipe-cleaner tool designs would show evidence of diversification, reflecting open-ended exploration of design space.

2. Methods

(a) Participants

624 participants took part in the main experiment. Of these, 600 participated in “transmission chain” groups of 10 individuals. Groups were pseudo-randomly allocated to tasks (building a tool out of either paper or pipe-cleaners) within one of three social learning conditions

(*Emulation*; *Imitation* or *Teaching*), giving 10 replicate groups of each task and social learning condition. The remaining 24 participants were allocated to the *Asocial* learning condition, in which they made 10 consecutive paper tools (N=12 participants) or pipe-cleaner tools (N=12) with no opportunity to learn from others. While most previous transmission chain experiments have enrolled only university students, we increased the diversity of participants by recruiting from local community groups (N = 38 groups of 10 individuals and 15 individuals in the *Asocial* condition; age 16-89 years) as well as the student body at the University of Exeter and Truro College (N = 22 groups and 9 *Asocial*; age 16-56). In all cases, group members knew each other, as would be expected in ancestral hominin groups (see supplementary materials for a full list of participating groups and further discussion of the potential impacts of group composition). We incentivised participation with a £1000 reward for the groups that produced the highest-performing tool of each type.

(b) Procedure

We ran experiments in classrooms, laboratories and community group rooms, with screens to separate areas for building and testing tools. Before starting the experiment, each participant read an information sheet and completed a consent form. We randomly allocated participants from social learning conditions to a position from one to ten within their transmission chain.

Each participant in turn was called into the experimental room. Here, they sat at a desk and were given written and verbal instructions to build, within five minutes, a tool from the materials provided (one sheet of waterproof paper or 30 identical, 30cm long pipe-cleaners) to carry as many marbles as possible. Participants were allowed to inspect the marbles, which were of two different sizes, before they began building, but did not have access to the marbles during building. The instructions specified that (a) paper tools must float on water before receiving marbles and (b) pipe-cleaner tools must be held by one or more handles incorporated in the design. A stopwatch clearly displayed the time elapsed and we updated builders periodically on their remaining time.

After the allocated building time elapsed, participants moved into a screened-off testing area, which contained a bowl filled with marbles of the two different sizes (totalling 3kg) and a scoop. Builders of paper tools were asked to float the tool in a tray filled with water and load as many marbles as possible into it without it sinking. In the pipe-cleaner task, builders were asked to load as many marbles as possible into the tool before carrying it to a set of weighing scales 5m away (see supplementary materials for further details of the testing procedure). The time available for testing was unrestricted, so the staggering of transmission chains had an element of fluidity (mean testing time = 3 mins; range 2-5 mins; see supplementary materials; Fig S2). During testing, we recorded the number of marbles of each size and whether or not the paper tools took on water. After testing, participants were either guided to a waiting area or, for participants in *Teaching* treatments, asked to stay behind to help other group members. At the end of the procedure participants filled in a debrief form that included a Likert scale question regarding their experience with handiwork or craft-making on a scale of 0 to 4.

(c) Experimental conditions

We gave each participant written and spoken instructions relevant to their experimental condition. For participants in the social learning conditions (*Emulation*, *Imitation* and *Teaching*) our transmission chains operated very similarly to an earlier study [19], whereby participants had five minutes (as described above) to build their implements before being replaced by the next participant in the chain, who then had five minutes to build their own implement. To address an important confound of most previous studies (c.f. [34]; see supplementary material for further information), we ensured that participants had access to social information for a standardised amount of time (seven minutes) across conditions. A visual depiction of the staggering of the chains for the three social learning conditions can be seen in the supplementary materials (Figure S2).

In the *Emulation* condition, participants could not observe or communicate with other team members, but could examine the tools that they made (as well as being informed of the scores). Each new participant (from the third participant onwards) could inspect the two most recently constructed tools for two minutes before starting building, as well as having access to them during the five minutes building time, giving a total of seven minutes of access to social information (Figure S2; see supplementary material for further details).

In the *Imitation* condition, participants were able to observe earlier chain members building their tools, but could not communicate or touch the materials. Each new participant (from the third member of the chain onwards) observed the participant two steps ahead and the participant one step ahead for six minutes. Building commenced once the participant two steps ahead finished testing their tool (and the focal participant was informed of their score) (See Figure S2). While building, participants were also free to continue to observe the participant one step ahead in the chain as they completed their final one minute of building (and were informed of that participant's score as it was recorded), providing a total of seven minutes social learning time.

In the *Teaching* condition participants returned to the building area after testing their tool in order to help the next members of their group. During this "*Teaching* role" they could communicate with group members, but could not physically assist in building or touch the materials. Each participant (from the third participant in the chain onwards) received two minutes of teaching before commencing building. Teachers continued to guide and instruct throughout the five minute build, totalling seven minutes of teaching time. Each participant had one teacher (the person two steps before them in the chain) present for the full seven minutes, with an additional teacher (the chain member three steps ahead) joining once they had finished assisting the participant one step ahead in the chain (see supplementary materials for further details; Figure S2).

Finally, in the *Asocial* condition, participants were asked to build and test ten tools in succession, each time attempting to improve upon their previous score, with no opportunity to observe or communicate with others. The participant's previous two tools were left on display after each round of building

(d) Similarity measures

We used online surveys, built and administered using Qualtrics (www.qualtrics.com), to determine the similarity between different tools within transmission chains. Raters (blind to hypotheses and experimental conditions) were given detailed instructions and multiple tests of comprehension of the instructions, which they had to pass in order to proceed with the survey. Each survey question displayed two tools, and raters had to rate their similarity in terms of (a) shape and features and (b) underlying construction, using a slider on a continuous scale from 0.00 to 4.00 (see supplementary material for details). As similarity scores are bounded, they were analysed as continuous proportions, with logit transformation [35].

We conducted two separate surveys for each tool type. Survey 1 quantified the similarity of every tool to its successor(s) within the same transmission chain. For each tool type, a total of 151 raters each rated the similarity of 20 different pairs of tools, such that each pair was rated by at least three different raters. We then used the mean rating as the measure of similarity for analyses. Survey 2 followed the same format, but compared randomly selected pairs of tools from the same generation (either generation 1, 5 or 10) across different transmission chains to provide measures of divergence or convergence in tool designs. Every pair of tools was scored by ten different raters, and we used the mean value as the measure of similarity.

(e) Statistical analyses

We analysed data in R 3.6.3 [36], using the package lme4 for linear (mixed) models (LMMs). We assessed model fit using standard residual plot techniques. Response variables were transformed when necessary to meet model assumptions (transformations are specified in the statistical tables in the supplementary material), and we checked for potentially highly influential datapoints by calculating Cook’s distances. We adopted an information theoretic approach to model selection, ranking models by Akaike Information Criterion corrected for small sample sizes (AICc). The top model set contained models within AICc ≤6 of the lowest AICc value, and we applied the “nesting rule” [37], in which simpler versions of a nested model are favoured over more complex versions. In preliminary analyses of the factors influencing tool performance, using the entire dataset for both tool types, the best model included interactions between generation and both tool type and condition (Table S1; Table S2). For ease of interpretation, all subsequent analyses were therefore conducted on each tool type separately (see supplementary materials for full details of variables and data distributions in each model).

3. Results

(i) Tool performance:

(a) Paper tools

Paper tools showed clear improvements across generations, carrying more marbles irrespective of the experimental condition. The best supported model (Table S3) contained only effects of generation (LMM: β (s.e.) = 0.389 (0.055), t = 7.039, p <0.001, CI (0.280, 0.499), Figure 1a) and craft, with people with more craft experience building better tools (β (s.e.) = 0.452 (0.135), t = 3.360, p = 0.001, CI (0.188, 0.719), Table S4). Model comparisons provide little support for effects of condition, or for an interaction between generation and condition (Table S3).

(b) Pipe-cleaner tools

The improvement in pipe-cleaner tools across generations depended on the experimental condition. The best supported model (Table S5) included an interaction between generation and condition: compared to asocial learning, the slope of improvement was lower in *Emulation* and *Imitation* chains, but did not differ between *Asocial* learning and *Teaching* chains (Figure 1b; Table S6). Additional post-hoc comparisons indicate that *Teaching* chains showed a steeper slope of improvement compared to *Imitation* chains (β (s.e.) = 0.234 (0.102), t = 2.307, *p* = 0.025; CI (0.031; 0.437)), but not compared to *Emulation* ((β (s.e.) = 0.155 (0.120), t = 1.288, *p* = 0.208; CI (-0.091; 0.400); Table S7). The top model also included a positive effect of craft experience (Table S5; Table S6).

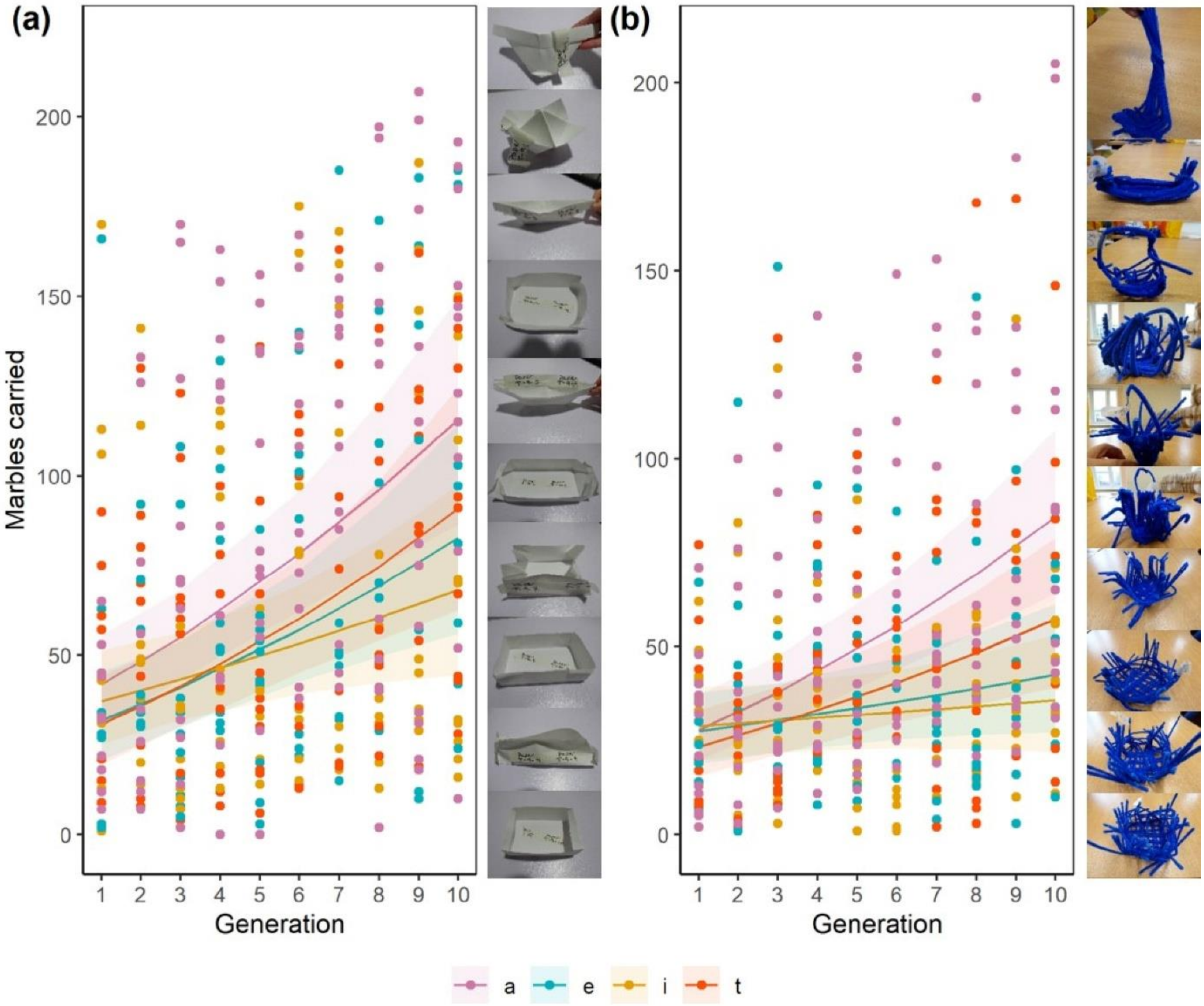

Figure 1. Slopes of improvement in (a) paper and (b) pipe-cleaner tools across experimental conditions: a = asocial learning; e = emulation; i = imitation; t = teaching. Images show illustrative examples of transmission chains from generation 1 (top) to 10 (bottom).

(ii) Improvements across the chain: performance of tools and their successors

(a) Paper tools

There was a negative relationship between the performance of a paper tool and the relative performance of its successor (defined as the tool two steps later in the chain, given that social learning from this tool was available across all three social learning conditions; Fig. S2). The best supported model included a negative effect of total marbles carried (Table S8): if a tool performed badly, its successor was likely to do better (positive difference score); if a tool performed very well, its successor is likely to do worse (negative difference score: LMM: β (s.e.) = -0.615 (0.066), t = -9.251, *p* <0.001, CI (-0.752,-0.466), Figure 2a; Table S8). In addition, participants with greater craft experience obtained better relative scores (Table S8; Table S9). There was no clear evidence of an effect of condition: the top model set included an interaction between total marbles carried and condition, but this was not robust (total*condition=*Imitation*: β (s.e.) = 0.017 (0.163), t = 0.103, *p* = 0.918; CI (-0.30; 0.32); total*condition=*Emulation*: β (s.e.) = 0.195 (0.172), t = 1.132, *p* = 0.259; CI (-0.15; 0.52)).

(b) Pipe-cleaner tools

As with the paper tools, we found that as the success of a pipe-cleaner tool increased its successor was likely to do worse. However, teaching attenuated this negative relationship. The best performing model included an interaction between total marbles carried and condition (Table S10): the successors of high-performing tools showed reduced loss of performance in *Teaching* conditions compared to *Emulation* (β (s.e.) = 0.003 (0.001), t = 3.667, *p* < 0.001; CI (0.002; 0.005) and *Imitation* conditions ((β (s.e.) = 0.004 (0.001), t = 4.468, *p* < 0.001; CI (0.001; 0.004); Table S11; Table S12; Fig 2b).

There was some evidence that the relationship between the performance of pipe-cleaner tools and their successors differed between student and community groups, as the top model set included an interaction between total and group type (Table S11). In community groups the successors of high-performing tools showed a steeper loss of performance compared to student groups (total*grouptype=students: β (s.e.) = 0.002 (0.001), t = 2.214, $p = 0.028$; CI (0.001; 0.004).

(iii) Similarity between tools and their successors

In Survey 1, similarity measures in terms of (a) shape and features, and (b) underlying construction were very strongly correlated in all cases ($R^2 > 0.8$). Analyses of (a) and (b) gave qualitatively the same results, so only the former are reported here.

(a) Paper tools

Analysis of the similarity between each tool and its successor indicated that designs that performed well were more likely to be replicated. The best supported model included a positive effect of the total number of marbles carried: if a paper tool was particularly effective, its successor was more likely to be similar (Table S13; LMM, β (s.e.) = 0.011 (0.001), t = 6.225, $p < 0.001$, CI (0.008; 0.015); Table S14; Figure 2c). There was no evidence of any differences between experimental conditions (Table S13).

(b) Pipe-cleaner tools

Again, analyses suggested that high-performing designs were more likely to be replicated, though this relationship was only clearly apparent in *Teaching* and *Emulation* conditions. The best supported model included an interaction between the total number of marbles carried and condition (Table S15; Table S16; Figure 2d). Post-hoc comparisons confirmed that, compared to *Imitation* chains, *Teaching* and *Emulation* chains showed a stronger positive relationship between the performance of a pipe-cleaner tool and the similarity of its successor (β (s.e.) = 0.022 (0.006), t = 3.54, p < 0.001, CI (0.010; 0.035); Table S17). The relationship tended to be steeper in *Teaching* than *Emulation* chains, but the evidence was weak (β (s.e.) = 0.009 (0.005), t = 1.69, p = 0.093, CI (-0.001; 0.018); Table S17).

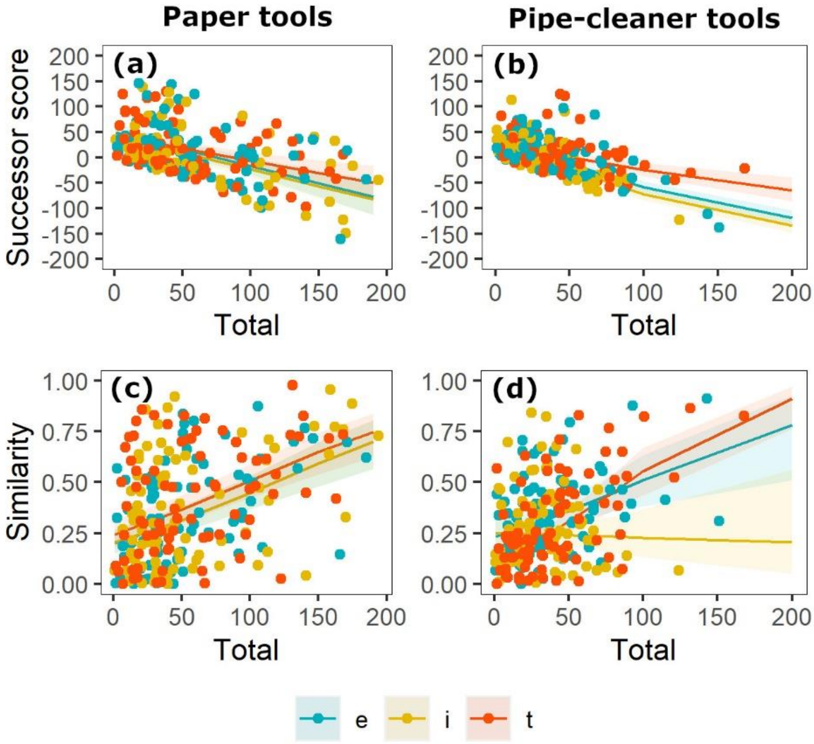

Figure 2. Relationship between the performance (Total marbles carried) of (a) paper and (b) pipe-cleaner tools and their successors across social learning conditions (e = emulation; i = imitation; t = teaching). (c) Paper tools that carried larger numbers of marbles produced more similar successors. For (d) pipe-cleaner this the relationship was particularly steep in the teaching condition.

(iv) Convergence and diversification of designs: between-chain comparisons

(a) Paper tools

Across different chains, paper tools from generation 10 were significantly more similar to each other than were paper tools from generation 1 (Fig S3a; Table S18; similarity in terms of shapes and features: $\beta = 0.789$, s.e. = 0.376, $t = 2.10$, $p = 0.042$, CI (0.053, 1.526); underlying construction $\beta = 0.692$, s.e. = 0.311, $t = 2.26$, $p = 0.032$, CI (0.083, 1.301). In the final generation, most paper tools had converged on similar, flat-bottomed designs (Fig S3c).

(b) Pipe-cleaner tools

Unlike the paper tools, the top model did not include an effect of generation on the similarity of pipe-cleaner tools across different chains (Table S19; Fig S3b), and there was little evidence that pipe-cleaner tools converged on similar designs (Fig S3d).

4. Discussion

Our findings are consistent with the argument that teaching coevolved with the manufacture of increasingly complex and causally opaque tools. In our experiments, both paper and pipe-cleaner tools showed clear cumulative improvements, increasing in efficacy across experimental generations. However, while there were no differences between the learning conditions in the relatively simple paper tool task, we found evidence that teaching provided important advantages in the production of the more causally opaque pipe-cleaner tools. Moreover, whereas paper tools tended to converge on a common, flat, tray-like design, pipe-cleaner tools maintained a diversity of designs; a key feature of modern human cumulative culture which seems to be absent in other species [27].

Our results add to the weight of evidence that high fidelity social learning processes are not fundamental pre-requisites for cumulative cultural evolution (CCE). In our experiments, simply having the opportunity to inspect tools produced by others was sufficient to generate cumulative improvements in performance of both tool types. This clearly fulfils the "core criteria" for CCE [27] (though note that some authors argue that CCE must result in behaviours or products that no individual could invent within their lifetime [6,38]; a criterion that has been criticised on both practical and conceptual grounds [20,27]). Thus, our results, alongside other similar findings [19,20] and recent research on non-human animals [23,28,29], indicate that CCE can occur in the absence of specialised forms of human social learning, and raise the possibility that CCE may be more common in nature than previously assumed. Our findings also speak to important debates in the literature on human culture. For instance, researchers have long debated whether human ecological dominance derives from our intrinsic individual intelligence [39] or as a collective outcome of CCE [2]. Our results blur this distinction, suggesting that cultural change cannot be understood without considering aspects of individual cognition such as instrumental learning (note that craft experience improved performance in our experiments), causal reasoning to reverse-engineer and improve artefacts [5,40,41], and strategies for deciding when to rely on social learning [42]. Similarly, there are longstanding debates as to whether cultural evolution rests on mechanisms for preserving or transforming learned information (reviewed in [43]). Our results suggest both are important: learners tended to make similar copies of tools that performed well, but were more likely to modify tool designs if their predecessors performed badly.

Although not strictly necessary for CCE to occur, we find that teaching provides important advantages, but only when the task is relatively causally opaque. While we found no effects of experimental condition in the paper task, in the pipe-cleaner task *Teaching* was the only social learning condition to show equivalent slopes of improvement to the *Asocial* condition. Importantly, asocial learners had direct access (via memory) to accumulated experience across *all* previous attempts (whereas social learners could only acquire information directly from their immediate predecessors) and were not subject to the constraints inherent in transmitting learned expertise *between* individuals. In the pipe-cleaner task, teaching was the only form of social learning to overcome these constraints, resulting in slopes that resembled those of the asocial condition. Given that our experimental design simulates change across generations, one might argue that this implies that teaching chains showed cumulative improvements equivalent to ten “lifetimes” of individual learning. Thus, teaching could generate important savings in terms of time and effort (critical if teaching is to be favoured by selection [25,44]). Participants in generation 10 of our teaching chains were, following a single bout of teaching, producing pipe-cleaner tools as effective as those of asocial learners who had been refining their tools over 10 rounds. Nevertheless, as is clear from the similar slopes of improvement in *Teaching* and *Asocial* conditions, the importance of individual learning must not be downplayed (see also [3,22]). In naturalistic settings, the interplay between asocial and social learning is likely to be critical, as experience will often allow individuals to refine and hone their (socially acquired skills) before they are transmitted to others.

Within the scope of the experiment, the advantages of teaching in generating steeper slopes of improvement were relatively modest, with post-hoc tests revealing a clear-cut difference in comparison to *Imitation*, but not *Emulation* chains. One possible explanation for this is that participants in the *Imitation* chains may have been relatively disadvantaged. A consequence of balancing the amount of social learning time available across conditions was that *Imitation* participants were not able to able to observe the full construction process of the predecessor two steps ahead of them in the chain (See Figure S2). This is different from both the *Emulation* and *Teaching* chains in which the full design of the implement two steps ahead could be either inspected or described. Nevertheless, the analyses comparing the performance of tools with their successors indicate that teaching may be vital in retaining and improving upon high-performing innovations. As one might expect, participants found it more difficult to improve upon tools that performed particularly well, resulting in a negative relationship between the performance of a given tool and the relative performance of its successor. However, in the pipe-cleaner task, teaching attenuated this decline in performance, and analyses of tool similarity provide some evidence that it facilitated the retention of high-performing designs. These finding parallels results from a recent experimental study on the transmission of flint-knapping [33], which found that teaching reduced the loss of cultural information compared to other forms of social learning and suggested that human teaching and language coevolved with the emergence of Oldowan stone tool-making around 2.5 million years ago. Our findings suggest that selection for verbal teaching may in fact pre-date and perhaps scaffolded the evolution of stone tools. Compared to other learning processes, teaching provides the distinct advantage that teachers can convey information and advice about how designs may be improved and what *not* to do, and focus their pupils' attention on elements of task design and the manufacturing process that are difficult to infer through observation alone (c.f. [20,30,32]). Mechanisms of teaching, including components of language such as syntax and recursion, may thus have come under selection long before the emergence of stone tools (for related arguments, see [45,46]). This could have allowed our ancestors to produce increasingly effective and opaque tools by combining elements made from perishable materials, similar to what we see in our pipe-cleaner task (see also [47]).

Mesoudi and Thornton [27] recently made a distinction between the core criteria for CCE, which may be met in other species, and extended criteria including the diversification of cultural lineages, which current evidence suggests are restricted to humans. Our findings provide some indication of how the latter may arise from the former through gradual co-evolutionary processes. As in recent experimental studies on non-human animals [29], our paper task was played out in a simple adaptive landscape with a single optimal solution. Accordingly, paper tools from different transmission chains became more similar to each other as the generations progressed, tending to converge on wide, flat-bottomed designs. In contrast, the pipe-cleaner tools from the final generation retained a diversity of different designs, and were no more similar to each other than those from the first generation. This suggests that the production of distinct lineages of cultural artefacts may emerge as a product of the gradual cultural evolution of increasingly causally opaque implements. Our experimental design precluded the transmission of information between groups, but in natural settings transmission of information between social sub-units could also facilitate the recombination of designs across cultural lineages, generating ever-more complex adaptive landscapes (see [48]).

As teaching involves a costly investment in helping others to learn, it is only expected to evolve if it provides advantages over other forms of learning [44]. While we cannot rule out the possibility that human teaching evolved for some other function, our results are consistent with theoretical modelling which suggests that the initial emergence of cumulative culture generated selection pressure for teaching that is absent in other great apes [25]. These arguments assume that the differences between human and non-human culture began to emerge as a result of coevolutionary processes linked to our ancestors' increasing reliance on tools following the split from other great ape lineages. A greater emphasis on the ultimate adaptive benefits of tool-making [49,50] alongside proximate factors like cognition [4] and demography [48] is therefore vital to understand both the ancient origins of human technology and its subsequent elaboration into the powerful, world-changing force we see today.

Data accessibility: Data and R code are available on Figshare: https://doi.org/10.6084/m9.figshare.12759626.v1

Author contributions: AT, CAC and FH conceived the idea. AL led the design and execution of the experiments with guidance from AT and input from CAC and FH. AL, DW and ED recruited the participants and collected the data. AT and MK analysed the data and AT drafted the manuscript. All authors gave final approval for publication.

Acknowledgements: Charlie Savill provided invaluable assistance in data collection. We are deeply indebted to participants from diverse community groups across Cornwall who gave up their time to take part in the experiment and to the University of Exeter and Truro College for their help with recruiting student participants.

Funding: This work was supported by an ESRC grant (ES/M006042/1) to AT, CAC and FH. CC also received support from a European Research Council (ERC) consolidator grant (No. 648841 RATCHETCOG ERC-2014-CoG).

Ethical statement: This study followed the guidelines of the British Psychological Society's Code of Human Research and the 1964 Declaration of Helsinki. The methods were approved by the University of Exeter Biosciences Research Ethics Committee (2014/538) and all participants provided written consent before taking part.

References

1. Laland KN. 2018 *Darwin's unfinished symphony: how culture made the human mind*. Princeton, NJ: Princeton University Press.

2. Henrich J. 2017 *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.

3. Stout D, Hecht EE. 2017 Evolutionary neuroscience of cumulative culture. *Proc Natl Acad. Sci.* **114**, 7861–7868. (doi:10.1073/pnas.1620738114)

4. Caldwell CA, Atkinson M, Blakey KH, Dunstone J, Kean D, Mackintosh G, Renner E, Wilks CEH. 2020 Experimental assessment of capacities for cumulative culture: Review and evaluation of methods. *Wiley Interdiscip. Rev. Cogn. Sci.* **11**, e1516. (doi:10.1002/wcs.1516)

5. Osiurak F, Reynaud E. 2020 The elephant in the room: what matters cognitively in cumulative technological culture. *Behav. Brain Sci.* **43**, e156: 1-66.

6. Tennie C, Call J, Tomasello M. 2009 Ratcheting up the ratchet: on the evolution of cumulative culture. *Phil. Trans. R. Soc. B* **364**, 2405–2415. (doi:10.1098/rstb.2009.0052)

7. Dean LG, Kendal RL, Schapiro SJ, Thierry B, Laland KN. 2012 Identification of the social and cognitive processes underlying human cumulative culture. *Science* **335**, 1114–1118. (doi:10.1126/science.1213969)

8. Hill K, Barton M, Hurtado AM. 2009 The emergence of human uniqueness: characters underlying behavioral modernity. *Evol. Anthropol.* **18**, 187–200. (doi:10.1002/evan.20224)

9. Thornton A, Clutton-Brock T. 2011 Social learning and the development of individual and group behaviour in mammal societies. *Phil. Trans. R. Soc. B* **366**, 978–987. (doi:10.1098/rstb.2010.0312)

10. Aplin LM. 2018 Culture and cultural evolution in birds: a review of the evidence. *Anim. Behav.* **147**, 179–187. (doi:10.1016/j.anbehav.2018.05.001)

11. Whiten A. 2019 Cultural evolution in animals. *Annu. Rev. Ecol. Evol. Syst.* **50**, 27–48. (doi:10.1146/annurev-ecolsys-110218-025040)

12. Vale GL, Flynn EG, Kendal RL. 2012 Cumulative culture and future thinking: Is mental time travel a prerequisite to cumulative cultural evolution? *Learn. Motiv.* **43**, 220–230. (doi:10.1016/j.lmot.2012.05.010)

13. Heyes C. 2016 Who Knows? Metacognitive Social Learning Strategies. *Trends Cogn. Sci.* **20**, 204–213. (doi:10.1016/j.tics.2015.12.007)

14. Galef BG. 1992 The question of animal culture. *Hum. Nat.* **3**, 157–178. (doi:10.1007/BF02692251)

15. Boyd R, Richerson PJ. 1996 Why culture is common, but cultural evolution is rare. In *Evolution of Social Behaviour Patterns in Primates and Man* (eds WG Runciman, J Maynard Smith, RIM Dunbar), pp. 77–93. Oxford: Oxford University Press.

16. Derex M, Godelle B, Raymond M. 2012 Social learners require process information to outperform individual learners. *Evolution* **67**, 688–697. (doi:10.5061/dryad.5ck3n)

17. Wasielewski H. 2014 Imitation is necessary for cumulative cultural evolution in an unfamiliar, opaque task. *Hum. Nat.* **25**, 161–179. (doi:10.1007/s12110-014-9192-5)

18. Van Der Post DJ, Franz M, Laland KN. 2017 The evolution of social learning mechanisms and cultural phenomena in group foragers. *BMC Evol. Biol.* **17**, 1–15.

(doi:10.1186/s12862-017-0889-z)

19. Caldwell CA, Millen AE. 2009 Social learning mechanisms and cumulative cultural evolution: is imitation necessary? *Psychol. Sci.* **20**, 1478–1483.

20. Zwirner E, Thornton A. 2015 Cognitive requirements of cumulative culture: teaching is useful but not essential. *Sci. Rep.* **5**, 16781. (doi:10.1038/srep16781)

21. Reindl E, Apperly IA, Beck SR, Tennie C. 2017 Young children copy cumulative technological design in the absence of action information. *Sci. Rep.* **7**, 1788. (doi:10.1038/s41598-017-01715-2)

22. Truskanov N, Prat Y. 2018 Cultural transmission in an ever-changing world: trial-and-error copying may be more robust than precise imitation. *Phil. Trans. R. Soc. B Biol. Sci.* **373**, 20170050. (doi:10.1098/rstb.2017.0050)

23. Saldana C, Fagot J, Kirby S, Smith K, Claidière N. 2019 High-fidelity copying is not necessarily the key to cumulative cultural evolution: A study in monkeys and children. *Proc. R. Soc. B* **286**, 20190729. (doi:10.1098/rspb.2019.0729)

24. Castro L, Toro MA. 2014 Cumulative cultural evolution: the role of teaching. *J. Theor. Biol.* **347**, 74–83. (doi:10.1016/j.jtbi.2014.01.006)

25. Fogarty L, Strimling P, Laland KN. 2011 The evolution of teaching. *Evolution* **65**, 2760–2770. (doi:10.1111/j.1558-5646.2011.01370.x)

26. Thornton A, Happé F, Caldwell CA. 2020 Supporting the weight of the elephant in the room: technical intelligence propped up by social cognition and language. *Behav. Brain Sci.* **43**, e156: 43-44.

27. Mesoudi A, Thornton A. 2018 What is cumulative cultural evolution? *Proc. R. Soc. B* **285**, 20180712. (doi:10.1098/rspb.2018.0712)

28. Fehér O, Wang H, Saar S, Mitra PP, Tchernichovski O. 2009 De novo establishment of wild-type song culture in the zebra finch. *Nature* **459**, 564–568. (doi:10.1038/nature07994)

29. Sasaki T, Biro D. 2017 Cumulative culture can emerge from collective intelligence in animal groups. *Nat. Commun.* **8**, 15049. (doi:10.1038/ncomms15049)

30. Csibra G, Gergely G. 2011 Natural pedagogy as evolutionary adaptation. *Phil. Trans. R. Soc. B* **366**, 1149–1157.

31. Lotem A, Halpern JY, Edelman S, Kolodny O. 2017 The evolution of cognitive mechanisms in response to cultural innovations. *Proc. Natl. Acad. Sci.* **114**, 7915–7922. (doi:10.1073/pnas.1620742114)

32. Caldwell CA, Renner E, Atkinson M. 2018 Human Teaching and Cumulative Cultural Evolution. *Rev. Philos. Psychol.* **9**, 751–770. (doi:10.1007/s13164-017-0346-3)

33. Morgan TJH *et al.* 2015 Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat. Commun.* **6**, 6029. (doi:10.1038/ncomms7029)

34. Miton H, Charbonneau M. 2018 Cumulative culture in the laboratory: methodological and theoretical challenges. *Proc. R. Soc. B* **285**, 20180677. (doi:10.1098/rspb.2018.0677)

35. Warton DI, Hui FKC. 2011 The arcsine is asinine: the analysis of proportions in ecology. *Ecology* **92**, 3–10.

36. R Core team. 2020 R: a language and environment for statistical computing. https://www.r-project.org/

37. Richards SA, Whittingham MJ, Stephens PA. 2011 Model selection and model averaging in behavioural ecology: The utility of the IT-AIC framework. *Behav. Ecol.*

Sociobiol. **65**, 77–89. (doi:10.1007/s00265-010-1035-8)

38. Reindl E, Gwilliams AL, Dean LG, Kendal RL, Tennie C. 2020 Skills and motivations underlying children's cumulative cultural learning: case not closed. *Palgrave Commun.* **6**, 106. (doi:10.1057/s41599-020-0483-7)

39. Pinker S. 2010 The cognitive niche: coevolution of intelligence, sociality, and language. *Proc. Natl. Acad. Sci.* **107**, 8993–8999. (doi:10.1073/pnas.0914630107)

40. Vaesen K. 2012 The cognitive bases of human tool use. *Behav. Brain Sci.* **35**, 203–218. (doi:10.1017/S0140525X11001452)

41. Penn DC, Holyoak KJ, Povinelli DJ. 2008 Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* **31**, 109–130. (doi:10.1017/S0140525X08003543)

42. Kendal RL, Boogert NJ, Rendell L, Laland KN, Webster M, Jones PL. 2018 Social learning strategies: bridge-building between fields. *Trends Cogn. Sci.* **22**, 651–665. (doi:10.1016/j.tics.2018.04.003)

43. Acerbi A, Mesoudi A. 2015 If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution. *Biol. Philos.* **30**, 481–503. (doi:10.1007/s10539-015-9490-2)

44. Thornton A, Raihani NJ. 2008 The evolution of teaching. *Anim. Behav.* **75**, 1823–1836.

45. Laland KN. 2017 The origins of language in teaching. *Psychon. Bull. Rev.* **24**, 225–231. (doi:10.3758/s13423-016-1077-7)

46. Kolodny O, Edelman S. 2018 The evolution of the capacity for language: the ecological context and adaptive value of a process of cognitive hijacking. *Phil. Trans.*

R. Soc. B **373**, 20170052. (doi:10.1098/rstb.2017.0052)

47. Panger MA, Brooks AS, Richmond BG, Wood B. 2003 Older than the Oldowan? Rethinking the emergence of hominin tool use. *Evol. Anthropol.* **11**, 235–245. (doi:10.1002/evan.10094)

48. Derex M, Mesoudi A. 2020 Cumulative cultural evolution within evolving population structures. *Trends Cogn. Sci.* **24**, 654–667. (doi:10.1016/j.tics.2020.04.005)

49. Collard M, Buchanan B, O’Brien MJ, Scholnick J. 2013 Risk, mobility or population size? Drivers of technological richness among contact-period western North American hunter-gatherers. *Phil. Trans. R. Soc. B* **368**, 20120412. (doi:10.1098/rstb.2012.0412)

50. Biro D, Haslam M, Rutz C. 2013 Tool use as adaptation. *Phil. Trans. R. Soc. B Biol. Sci.* **368**, 20120408. (doi:10.1098/rstb.2012.0408)