Applying the cultural ratchet to a social artefact: The cumulative cultural evolution of a language game.

Nicolas Fay[1], T Mark Ellison[2], Kristian Tylén[3,4], Riccardo Fusaroli[3,4] & Simon Garrod[5]

[1]School of Psychology, University of Western Australia, Perth, WA, 6009, Australia

[2]School of Culture, History and Language, College of Asia & the Pacific, Australian National University, Canberra, ACT, 0200, Australia

[3]Interacting Minds Centre, Aarhus University, Jens Chr. Skous Vej 4, 8000 Aarhus C, Denmark

[4]Center for Semiotics, Aarhus University, Jens Chr. Skous Vej 2, 8000 Aarhus C, Denmark

[5]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

Keywords: Functionalism; Language Use; Human Communication; Language Evolution; Language Game; Map Task; Cumulative Cultural Evolution

Corresponding author:

Nicolas Fay, School of Psychology, University of Western Australia

35 Stirling Highway, Crawley, WA 6009 Australia

Email: nicolas.fay@gmail.com; Tel: +61 (0)8 6488 2688; Fax: +61 (0)8 6488 1006

Word count (excluding title page, abstract and references): ~6,500 words

ABSTRACT

Material artefacts evolve by cumulative cultural evolution - the accumulation of adaptive modification over time - but it is not known if this process also applies to social artefacts. Using a large scale experimentally study (N=408) that simulates intergenerational transmission, we tested for the cumulative cultural evolution of a social artefact. Specifically, we demonstrate the cumulative cultural evolution of what *Wittgenstein* calls a 'language game'; in the context of our study, the use of an increasingly effective subset of language for conveying routes on a map. Over experimental generations, participants progressively reproduced the routes with higher precision. This progressive improvement in route reproduction precision was driven by cultural inheritance plus the accumulation of adaptive modifications to the language game. Observation-based social learning was sufficient for the cumulative cultural evolution of the language game, but the precision of the language game was enhanced by teaching, a collaborative form of social learning.

INTRODUCTION

Over time material cultural has become better adapted for purpose. This is seen in the progressive improvement of the hammer, evolving from a crudely shaped pounding stone to a stone-and-stick composite to its modern-day metal and mechanical forms[1]. This historical process is known as cumulative cultural evolution; the accumulation of adaptive modifications over time[2,3]. Cumulative cultural evolution has also been demonstrated under controlled laboratory conditions[4–6]. However, these studies have been exclusively concerned with the cultural evolution of material artefacts. This paper adopts a similar experimental methodology to study the cumulative cultural evolution of a social artefact. More specifically, we demonstrate the cumulative cultural evolution of a 'language game'[7]; the use of an increasingly effective subset of language for conveying routes on a map.

First, we review the literature on the cumulative cultural evolution of material and social artefacts. Next, we consider the social learning strategies important to cumulative cultural evolution, and the laboratory experiments designed to test their influence. We then explain the present study and report its findings.

Cumulative Cultural Evolution of Material and Social Artefacts

Cumulative cultural evolution occurs when the achievements of prior generations are built upon by subsequent generations to create complex products that no single individual could invent on their own. While great apes have cultural traditions, these have remained simple, and lack the progressive adaptation that is pervasive across the human lineage[8,9]. This suggests that the capacity to accumulate culture relies on social learning mechanisms that are uniquely human[3,10]. Cumulative cultural evolution is thought to rely on innovation together with the faithful transmission of the cultural artefact[2,3]. High fidelity transmission is essential to ensure that artefacts (material or social) survive long enough to acquire further adaptive modifications, or 'ratchet up'[11].

The literature on cumulative cultural evolution has tended to focus on the functional adaptation of material artefacts, such as clothing[12], stone tools[13] and kayaks[14]. This process has been simulated under laboratory conditions[4–6]. In one study human participants tried to build a paper plane that flew as far as possible, or a spaghetti tower that was as tall as possible[4]. Participants were arranged into transmission chains, where each participant learned from observing the previous participant's artefact (paper plane or spaghetti tower). Simulating intergenerational transmission in this way is frequently done to study cultural transmission[15,16]. Cumulative cultural evolution was observed; over generations the paper planes flew further, and the spaghetti towers became taller. The progressive improvement in the performance of the material artefacts was driven by faithful cultural transmission plus adaptive modification across the experimental generations.

Less attention has been paid to the cumulative cultural evolution of social artefacts such as customs and language, despite the importance of precise communication to the functional adaptation of material artifacts[6,17,18]. For example, it is not clear if religious practices or monetary systems have evolved by cumulative cultural evolution. Arguably, language is another type of social artifact that is subject to cumulative cultural evolution[19]. Rather than concentrate on the language system as a whole, we concentrate instead on language use within the context of a specific activity; what *Wittgenstein* calls a language game[7,20]. Example language games include narrating a sports event (a monologue), engaging in a debate (a competitive dialogue) or teaching (a cooperative dialogue).

In the present study the 'language game' relates to describing a route on a map to another person, such that the route can be accurately reproduced. As we shall see, there is considerable variation in the precision of the route description language games. This can be explained in terms of differences in the content and process used to convey the routes. For example, participants can employ a variety of different description types to convey a route, including those referring to distance, direction and landmarks. While there is evidence that human communication systems become more efficient through use[21–23], there is little evidence to suggest that they undergo cumulative cultural evolution (although experimental-semiotic studies with artificial languages provide some support[24–27]). If the language game evolves by cumulative cultural evolution this will be reflected by a progressive improvement in the precision with which the routes are reproduced across the experimental generations.

Social Learning Mechanisms for Cumulative Cultural Evolution

Two broad types of social learning are considered in the present study: observation learning and teaching. Observation learning, by emulation or by imitation[28,29], is social learning that is acquired by observing the behaviour of a model[30]. This is an individual-level social learning strategy. Teaching is when two (or more) people collaborate to ensure the privileged information held by the teacher is successfully transferred, or donated, to the student[31]. Compared to observation learning, teaching is a more sophisticated inter-individual social learning strategy[32]. Whereas observation learning is sufficient for the cumulative cultural evolution of material artefacts[33], teaching, especially with language, may be essential to the cumulative cultural evolution of more complex material artifacts[6].

In the context of referential communication games, teaching trumps observation; communication success is higher among active conversationalists compared to passive observers of the same conversation[34]. It is argued that active participation during information exchange allows shared meanings to be mutually agreed, or grounded[35]. A similar pattern is observed in a multi-generational version of this task[36]. Here, a narrative text was reproduced with higher fidelity across 4-person transmission chains when participants could directly interact with the adjacent chain member compared to when they could not. These studies indicate that social interaction is important to transmission fidelity, a crucial contributor to cumulative cultural evolution[11].

In the context of the present study, social interaction allows the student to negotiate a route description language game with the teacher. For example, a student might indicate uncertainty about the teacher's route description by asking a question or requesting a clarification. The teacher can then revise their earlier route description, or generate an alternative description, and this process can be repeated until the route description is accepted by the student. Interactive negotiation such as this offers a mechanism through which innovations arise and are selected. In our example, the student elicits route description modifications and innovations by querying the teacher's description, and cultural selection occurs when the student accepts a route description. This negotiation process also acts a mechanism to eliminate non-adaptive route descriptions; route descriptions that are not accepted by the student are discarded. Importantly, this collaborative selection process is not available to passive observers, who must rely solely on individual inference. We predict that on complex tasks, active participation during information exchange will advantage teaching over observation learning.

The Current Study

We present a large-scale experimental study that tests for the cumulative cultural evolution of a language game (N=408). Participants' task was to precisely communicate a route on a map to their partner (referred to as the Instruction-Giver and Instruction-Follower), who tried to reproduce the route on their map. Participants were either allocated to an 8-person transmission chain in which they could directly interact with the Instruction-Giver (Teaching condition), or an 8-person chain where they could observe the route descriptions produced by the Instruction-Giver, but they could not directly interact with the Instruction-Giver (Observation condition). If observation learning is sufficient for the cumulative cultural evolution of the language game, then route reproduction precision will improve across generations in both conditions. If teaching is important, then a cumulative improvement in route reproduction precision will only be seen in the Teaching condition, or route reproduction precision will be higher in this condition compared to the Observation condition. We predict that a cumulative improvement in route reproduction precision will be observed, but will be modulated by condition, such that teaching will confer an advantage over observation learning.

We then consider the cumulative cultural evolution of the language game that gives rise to the progressive improvement in route reproduction precision across the experimental generations and between the experimental conditions. It has been shown that the use of some terms positively affect performance on a joint task, whereas other terms negatively affect performance[37]. We first examine the content of the route descriptions, and the extent to which description terms that positively contribute to route reproduction precision accumulate, and the extent to which terms that negatively affect route reproduction precision are eliminated across the experimental generations. A benefit of teaching over observation learning is that the Instruction-Giver and -Follower can negotiate route description terms, and this may act as a selection mechanism (in addition to Instruction-Follower inference). If correct, more positive route description terms and fewer negative terms will be evident in the Teaching condition compared to the Observation condition. Next we examine the change in the process used to communicate the routes (e.g., how information is organized by Instruction-Givers). Conversation is the basic medium of human communication[35,38], so we were interested in how Instruction-Givers in the Observation condition might adapt the route description process to the non-interactive context. Our basic prediction is that route reproduction precision will be contingent upon the cumulative cultural evolution of the language game.

Our findings indicate that, like material artefacts, the language game evolved by cumulative cultural evolution; functional adaptation of the 'language-technology' led to progressive improvement in route reproduction precision across the experimental generations (both conditions). However, teaching improved route reproduction precision beyond observation. While the conditions were similar with regard to the adaptive change in the content of the language game, they differed in the process used to convey the routes.

METHOD

The study received approval from the University of Western Australia Ethics Committee. Participants viewed an information sheet before giving written consent to take part in the study. The information sheet and consent form were both approved by the Ethics Committee.

PARTICIPANTS

Four hundred and eight undergraduate psychology students (287 females) participated in exchange for partial course credit. Participants ranged in age from 18 to 57 years (M= 21.56, SD= 4.41).

TASK

The Map Task is an unscripted cooperative task that typically involves two persons[39]. Participants' task is to precisely communicate a route on a map to their partner. The Instruction-Giver's map contains a route and between 11 to 13 landmarks (simple line drawings of lakes, cottages, telephone boxes etc.). The Instruction-Follower's map contains only the landmarks, and no route. All routes begin with a start and end point that is marked on the Instruction-Giver and Instruction-Follower's map. In prior studies using the Map Task the landmarks were labeled with their name, and the number of shared landmarks was varied. In our study, the landmarks were not named, and both participants' maps contained the same set of landmarks. Removing the landmark labels was done to increase task difficulty (in addition to communicating the route, participants would also have to agree on how label the landmarks), and having participants share the same set of landmarks was done to ensure the task did not advantage participants in the Teaching condition over those in the Observation condition. Participants were told their task was to precisely reproduce the Instruction-Giver's route on the Instruction-Follower's map.

PROCEDURE

Fifty-one eight-person transmission chains were tested. Participants were randomly assigned to a condition (Teaching, Observation; 25 and 26 transmission chains respectively) and to a position in the transmission chain (1-8).

In each condition the task was done in pairs. Each participant was seated at a different computer terminal and all communication took place using an Internet text-based chat program (http://xchat.org/). Route description information appeared on the Instruction-Follower's screen after the Instruction-Giver pressed the return key (at which point the typed instructions were sent to the Instruction-Follower). In the Teaching condition the Instruction-Follower could interact with the Instruction-Giver. Using the text-chat tool they could ask questions, seek clarification of route descriptions etc. In the Observation condition the Instruction-Follower was not permitted to use the text-chat tool; they were restricted to passively processing the Instruction-Giver's instructions as they appeared on their screen (i.e., each time the Instruction-Giver sent a 'packet' of information). Thus, participants in Observation condition could not affect the route descriptions they received. The Instruction-Giver and Instruction-Follower's map was presented alongside the text-chat tool (to the right of the text-chat interface). The maps were presented as consistently-sized .gif images using the Microsoft Paint program. Using the computer mouse, the Instruction-Follower drew the route described by the Instruction-Giver onto their map. Instruction-Followers were told not to delete any part of the drawn route; if they thought they had made a mistake they were told to continue from that point.

Each pair of participants was given 10 minutes to complete the task, after which the reproduced routes and the text chat was saved to the local computer. The Instruction-Giver then left the experiment, the Instruction-Follower became the Instruction-Giver and a new Instruction-Follower was added. The new Instruction-Giver then communicated a new route to the new Instruction-Follower. This process was recycled across each 8-person transmission chain. To minimise practice effects, and promote the evolution of a general route description language game, eight different routes were used, and routes were randomly sampled (without replacement) and randomly assigned to a position in the transmission chain (see Supplementary Materials 1 - SM1 - for the 8 Instruction-Giver and Instruction-Follower maps).

RESULTS

The data was analyzed using linear mixed effects modeling. All analyses were conducted in R[40] and models were estimated using the lmer() function of lme4[41]. P-values were obtained by likelihood ratio tests of the full model against the model without the predictors. All predictors were centered prior to analysis. A maximal random effects structure was specified where possible[42]. The variance accounted for by the best-fitting model was calculated using r.squaredGLMM() function in MuMIn[43].

Section 1 tests for the progressive improvement in route reproduction precision across the experimental generations, and between the Teaching and Observation conditions. Section 2 tests for the cumulative cultural evolution of the language game that gives rise to the progressive improvement in route reproduction precision.

Section 1. Progressive Improvement in Route Reproduction Precision

We first quantified how far the Instruction-Follower's reproduced route deviated from the Instruction-Giver's route[44]. The deviation score in the present study is the number of pixels between the two routes. We then subtracted the deviation score from the total pixel count for the image to give a performance measure (expressed as a percentage of pixels) where a higher value reflects a higher precision route reproduction (see **Figure 1**).

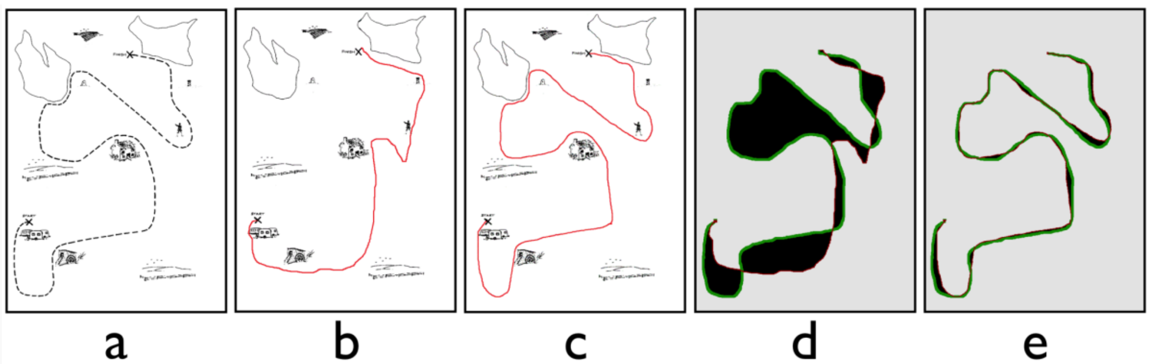

Figure 1. Example Instruction-Giver Map (Panel a) and the reproduced route by an Instruction-Follower at Generation 1-2 (Panel b) and at Generation 7-8 (Panel c) across two separate transmission chains (from the Observation condition). The Instruction-Follower's reproduced route (in red) was superimposed onto the Instruction-Giver's route (transformed to a solid green line). The deviation score (in pixels) was calculated (black area) and subtracted from the total number of pixels to give a route reproduction precision score (grey area; see Panel d and e). A lower deviation score returned a higher precision score (expressed as a percentage of total pixels). In this example, route reproduction precision was higher at Generation 7-8 than at Generation 1-2.

Instruction-Giver routes were progressively reproduced with higher precision by the Instruction-Followers across the experimental generations in both conditions. In addition, the Instruction-Giver routes were reproduced with higher precision in the Teaching condition compared to the Observation condition (see **Figure 2**).

Route reproduction precision scores exhibited strong negative skew (-2.260) that was normalized by logit transformation (skew= -0.058)[45,46]. Condition (Teaching, Observation) and Generation (1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8) with interaction were entered as fixed effects (see SM2 for the full model specification and model output). The best fitting model specified Condition and Generation as fixed effects without interaction, accounting for 28.29% of the variance in the route reproduction precision scores. Model comparison confirmed that allowing Condition and Generation to interact did not improve model fit, $\chi^2(1)= .105$, $p= .746$. By contrast, removing the fixed effect of Condition or the fixed effect of Generation reduced model fit, $\chi^2(1)= 8.978$, $p= .003$ and $\chi^2(1)= 24.377$, $p< .001$. The same pattern of results was returned by the raw, untransformed route reproduction precision scores (see SM3).

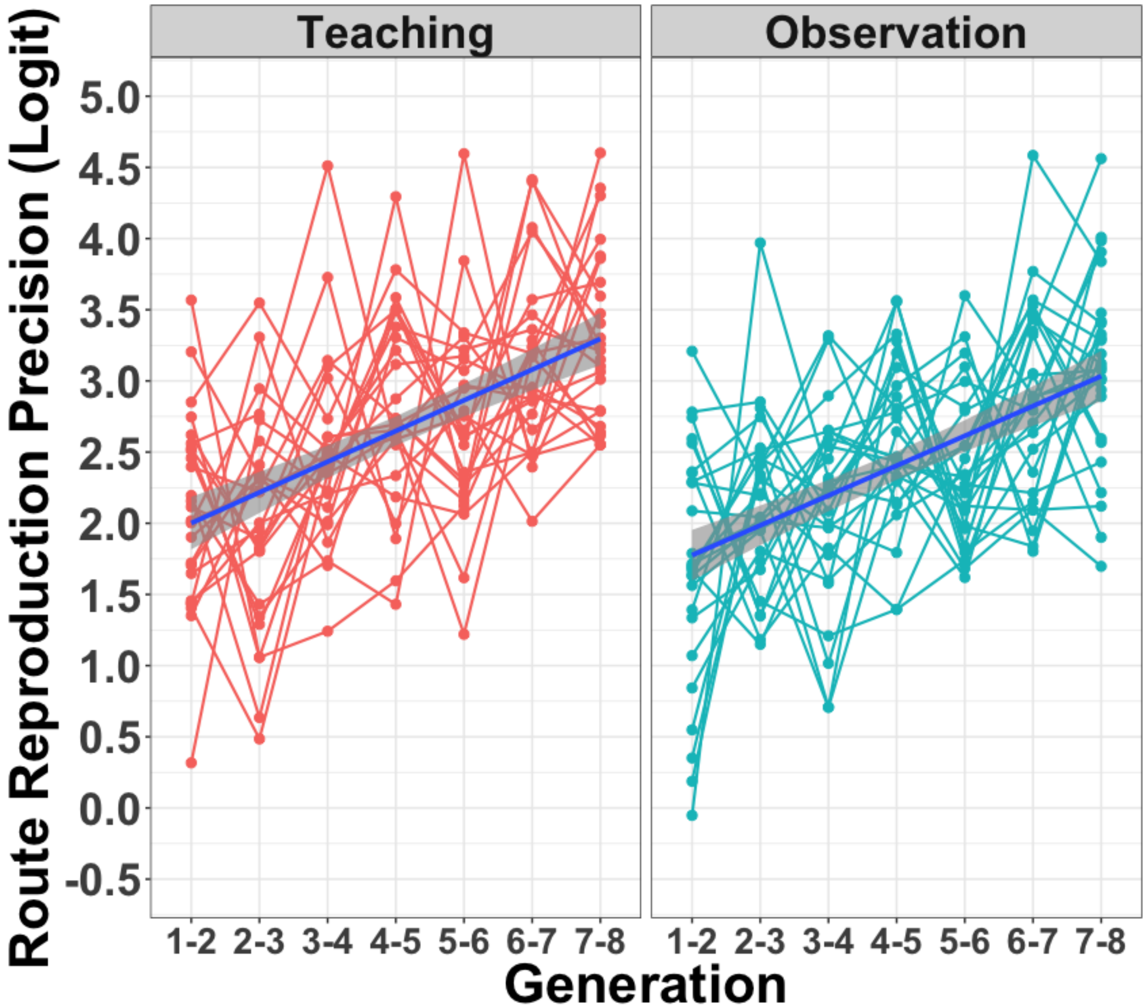

Figure 2. Change in route reproduction precision (expressed as Logit scores, and plotted for each transmission chain) across the experimental generations in the Teaching condition and the Observation condition. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

Cultural Inheritance of Route Reproduction Precision. For the language game to survive long enough to acquire further adaptive modifications, it must be faithfully transmitted between the experimental generations. If route reproduction precision at Generation

N-plus-1 is predicted by route reproduction precision at Generation N, this would be evidence of the cultural inheritance of the language game.

The route reproduction precision scores at Generation N-plus-1 was our dependent variable. Condition (Teaching, Observation) and the route reproduction precision scores at Generation N were entered as fixed effects with interaction (see SM4 for the full model specification and model output). The best fitting model specified Condition and Generation N Performance (route reproduction precision) as fixed effects without interaction, accounting for 6.05% of the variance in route reproduction precision scores at Generation N-plus-1. Model comparison confirmed that allowing Condition and Generation N Performance to interact did not improve model fit, $\chi^2(1)= .409$, $p= .522$. By contrast, removing the fixed effect of Condition or the fixed effect of Generation N Performance reduced the model fit, $\chi^2(1)= 5.347$, $p= .021$ and $\chi^2(1)= 13.945$, $p< .001$. As before, route reproduction precision was higher in the Teaching condition compared to the Observation condition, and route reproduction precision was predicted by the route reproduction precision of the prior generation, demonstrating cultural inheritance.

Section 2. The Cumulative Cultural Evolution of the Language Game

We now examine the adaptive changes in the language game that gives rise to the progressive improvement in route reproduction precision. Two mechanisms are considered: the content of what was communicated by the Instruction-Givers (Section 2a) and the process used to communicate the route information (Section 2b).

Section 2a: Adaptive Change in the Content Communicated by Instruction-Givers

The Instruction-Givers used a variety of description types to convey the route to the Instruction-Follower: distance (e.g., cm, close), direction (e.g., left/right, towards), shape (e.g., curve, slant), position (e.g., outer, alongside), preposition (e.g., onto, underneath), landmark (e.g., bridge, pond) and interactive feedback (e.g., ok, done). Some terms may have positively contributed to performance whereas others may have negatively affected performance. If the language game evolves by cumulative cultural evolution, then terms that positively contributed to route reproduction precision will have accumulated and terms that negatively affected route reproduction precision will have been eliminated.

The determination of which terms had a positive impact on route reproduction precision required three steps. First, the precision score was non-linearly scaled to fit a Normal distribution $N(0,1)$. Second, we assessed the performance of any term to be the log Bayes' Factor (lBF) assessing how much better the distribution $N(0.1,1)$ (predicting higher route reproduction precision) predicted the performance of pairs using the term than did the distribution $N(0,1)$ over all scores. Any term with lBF > 0.2 was regarded as a positive term, as pairs using it achieved better-than-average route reproduction precision. In the same manner, lBF of $N(-0.1,1)$ over $N(0,1)$ was used to identify negative terms (see SM5 for details on identifying term polarity). Using Bayes' Factor allowed us to incorporate both the number of cases where a word co-occurs with a positive or negative impact and the size of its impact in a single measure.

Because the total number of words used to communicate the different routes was higher for Instruction-Givers in the Teaching condition compared to the Observation condition ($M_{Teaching}$= 251.49 words, *SD*= 96.61 and $M_{Observation}$= 200.23 words, *SD*= 69.16; see SM6 for data visualization and statistical analysis), we computed the density of positive and negatively-biased tokens (positive and negatively-biased words, including repetitions, as a percentage of total Instruction-Giver words) for each Instruction-Giver at each generation. We then compared the change in the density of positive- and negatively-biased route description tokens over generations in the Teaching and Observation conditions.

In each condition the density of positively-biased route description tokens increased over generations, and the density of negatively-biased route description tokens decreased over generations. Instruction-Giver route descriptions in the Teaching condition were characterized by a higher density of positively-biased tokens, and a lower density of negatively-biased tokens compared to Instruction-Givers in the Observation condition (see **Error! Reference source not found.**).

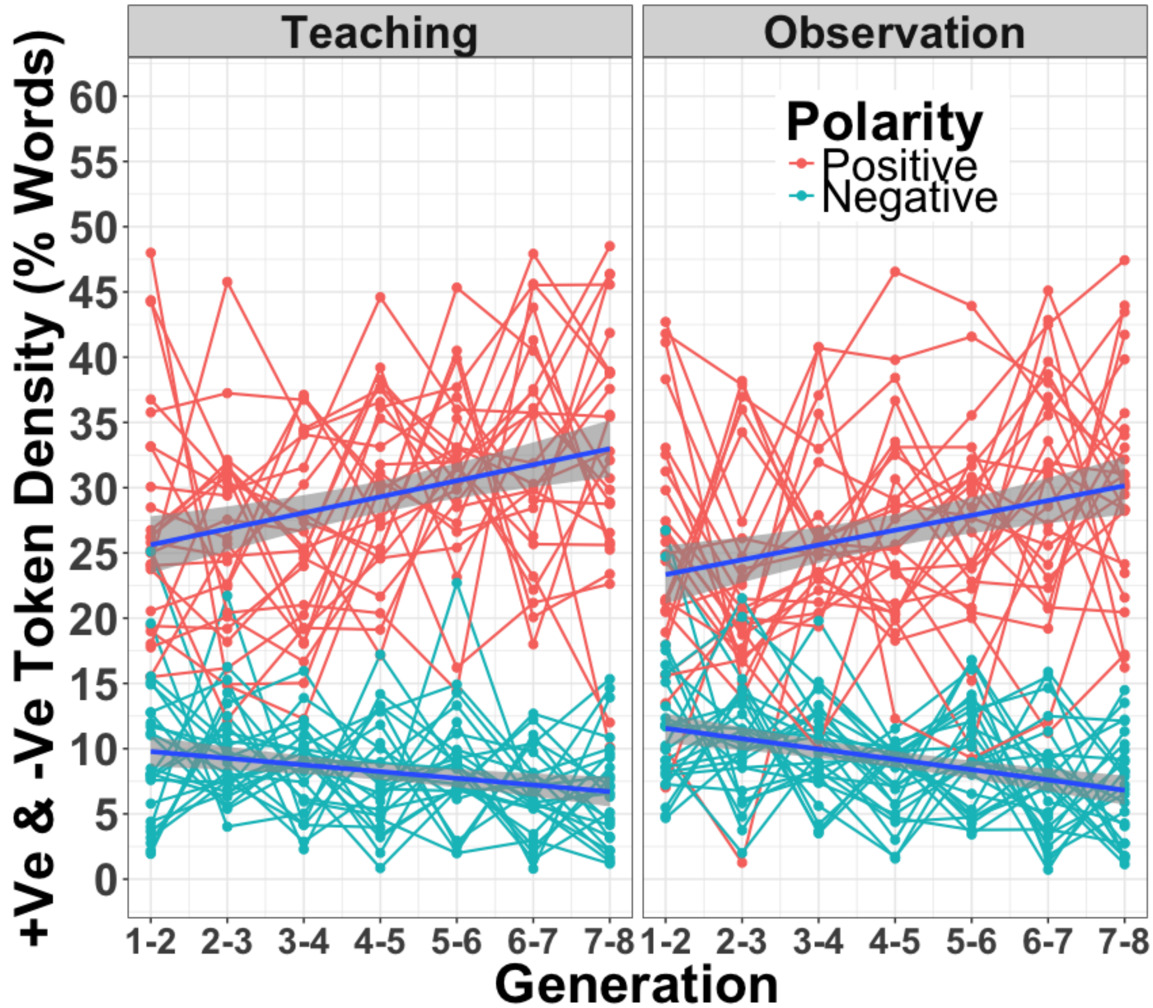

Figure 3. Change in the density of Positive- and Negatively-biased route description tokens (as a % of total tokens/words, and plotted for each transmission chain) produced by Instruction-Givers across Generations in the Teaching and Observation conditions. The blue straight line is the linear model fit and the grey shaded area is the 95% confidence interval.

Condition (Teaching, Observation), Token Polarity (Positive, Negative) and Generation (1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8) with interaction were entered as fixed effects (see SM7 for the full model specification and model output). The best fitting model specified Token Polarity and Generation as fixed effects with interaction plus Condition and Token Polarity as fixed effects with interaction, accounting for 69.57% of the variance in token density. Model comparison confirmed that allowing Condition, Polarity and Generation to interact did not improve model fit, $\chi^2(2)= 0.75$, $p= .689$. By contrast, removing the interaction between either pair of fixed effects reduced model fit, $\chi^2(1)= 25.13$, $p< .001$ and $\chi^2(1)= 4.73$, $p= .029$. The Token Polarity by Generation interaction is explained by the increase in the density of positively-biased route description tokens, and the decrease in the density of negatively-biased route description tokens over generations in each condition. The Condition by Polarity interaction is explained by the higher density of positively-biased tokens and the lower density of negatively-biased tokens in the Teaching condition compared to the Observation condition (see SM8).

Cultural Inheritance of Positively- and Negatively-Biased Route Description Tokens. If positively-biased route description tokens are inherited by the next generation and negatively-biased tokens are not, this would indicate the language game evolved by cultural selection.

Generation N Token Density, Condition and Token Polarity were entered as fixed effects with interaction (see SM9 for the full model specification and model output). The best fitting model specified Generation N Token Density and Token Polarity with interaction, accounting for 70.83% of the variance in token density at Generation N-plus-1. Model comparison confirmed that allowing Generation N Token Density, Condition and Token Polarity to interact did not improve model fit, $\chi^2(4)$= 5.508, p= .239. By contrast, removing the Generation N Token Density by Token Polarity interaction reduced model fit, $\chi^2(1)$= 7.015, p= .008. Further analysis indicated that positively-biased token density at Generation N predicted positively-biased token density at Generation N-plus-1, but not negatively-biased token density (see SM10).

In both conditions, positively-biased route description tokens were inherited from the prior experimental generation, but negatively-biased route description tokens were not. This pattern of results suggests that cultural selection was operating on the language game.

Profile of Positive and Negative Route Description terms. What constitutes a positively-biased and negatively-biased route description, and do they differ between the Teaching and Observation conditions? **Figure 4** illustrates the change in positive- and negatively-biased route description terms over generations in the different conditions, in the context of the route description types described earlier, e.g., direction, distance, shape. Also included is an ‘other’ category to capture terms that fell outside these categories. NF and TME independently coded the different route description terms into one of eight categories. Substantial inter-coder agreement was observed (K= .663, k= 2, N= 551)[47]. The conditions exhibited a similar profile of positive- and negatively-biased route description terms, but with a higher density of positively-biased terms in the Teaching condition and a lower density of negatively-biased terms in this condition (compared to the Observation condition). This suggests that the difference between the conditions is quantitative rather than qualitative.

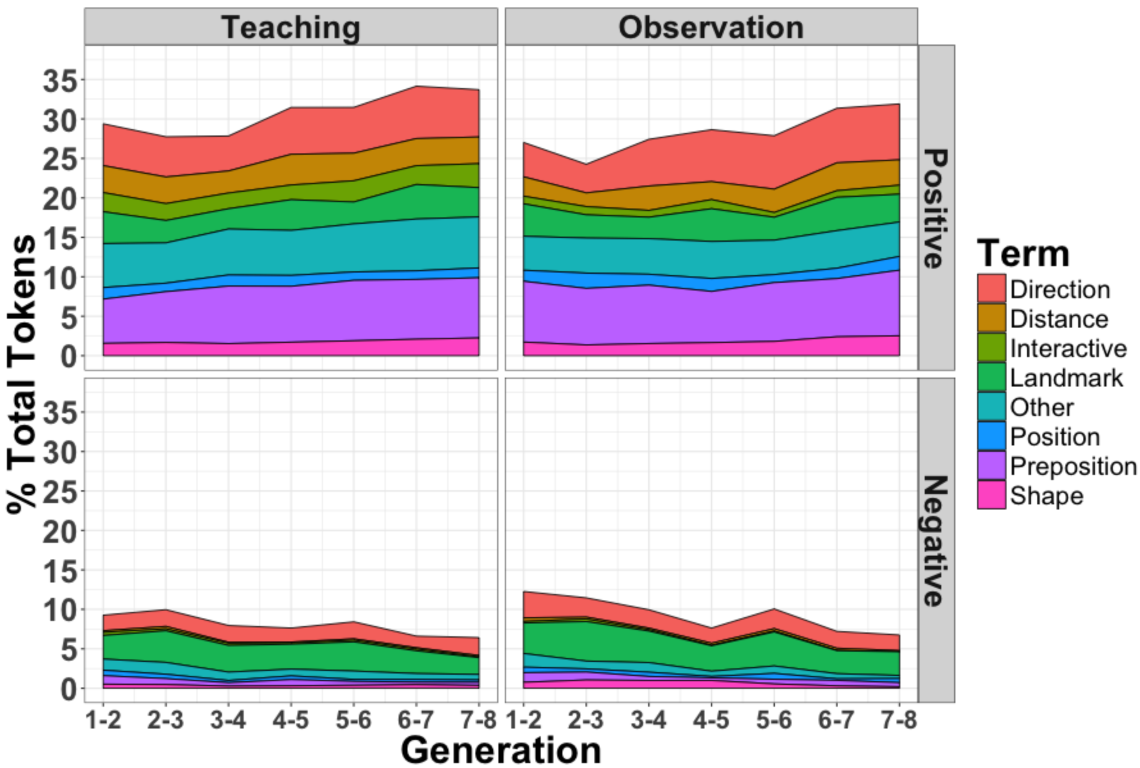

Figure 4. Change in the mean density of Positive- and Negatively-biased route description terms (as a % of total tokens/words) for each category of description type, produced by Instruction-Givers across Generations in the Teaching and Observation conditions.

Section 2b: Adaptive Change in the Communication Process

First we examine the adaptive influence of social interaction on route reproduction precision, and whether social interaction became a cultural tradition in the Teaching condition. Next we examine how Instruction-Givers packaged their route descriptions for the Instruction-Follower, and if this changed over generations. We then examine the adaptive influence of packet size (mean number of words per information transmission episode by Instruction-Givers) on route reproduction precision, and the extent to which packet size became a cultural tradition.

Social Interaction. Communication between Instruction-Givers and Instruction-Followers in the Teaching condition was was tightly coupled; the mean number of information packets sent was 20.80 for Instruction-Givers and 12.18 for Instruction-Followers, and these were strongly correlated, $r(173)= .65$, $p< .001$.

Was more frequent social interaction associated with higher route reproduction precision? To answer this, we computed the ratio of Instruction-Giver to Instruction-Follower packets sent and entered this as a fixed effect in a mixed effects model where route reproduction precision was the dependent variable (see SM11 for the full model specification and model output). Compared against a null model, including the ratio of Instruction-Giver to Instruction-Follower social interaction improved model fit, $\chi^2(1)= 6.942$, $p= .008$, accounting for 3.21% of the variance in the route reproduction precision scores. Despite the benefit of social interaction to route reproduction precision, there was no evidence of a statistical change in Instruction-Giver and Instruction-Follower interactivity over generations (Ratio of Instruction-Giver to Instruction-Follower packets sent: $M_{\text{Generation1-2}}= 0.59$, $SD= 0.32$ and $M_{\text{Generation7-8}}= 0.58$, $SD= 0.27$; see SM12 for data visualization and statistical analysis). This suggests that interactivity is a trait, either at the level of the Instruction-Follower or at the level of the dyad, and may not be amenable to social learning.

Route Information Packaging. Communication in the Teaching condition was characterized by short Instruction-Giver route descriptions (M= 20.80 packets, with M= 13.71 words per packet) accompanied by shorter Instruction-Follower feedback (M= 12.18 packets, with M= 4.06 words per packet). This conversation-style was consistent across transmission chains and generations in the Teaching condition (see **Figure 5**). By contrast, Instruction-Givers in the Observation condition did not receive Instruction-Follower feedback. In this condition there was a systematic change in the process used to transmit the route descriptions; from a conversation-style to a narrative-style. Initially short route descriptions were given by Instruction-Givers ($M_{\text{Generation1-2}}$= 13.81 packets, with M= 13.35 words per packet), but over generations the number of route description packets decreased and the size of each packet increased ($M_{\text{Generation7-8}}$= 5.00 packets, with M= 49.49 words per packet). This change in the route description process was consistent across Instruction-Givers in the Observation condition (see **Figure 5**).

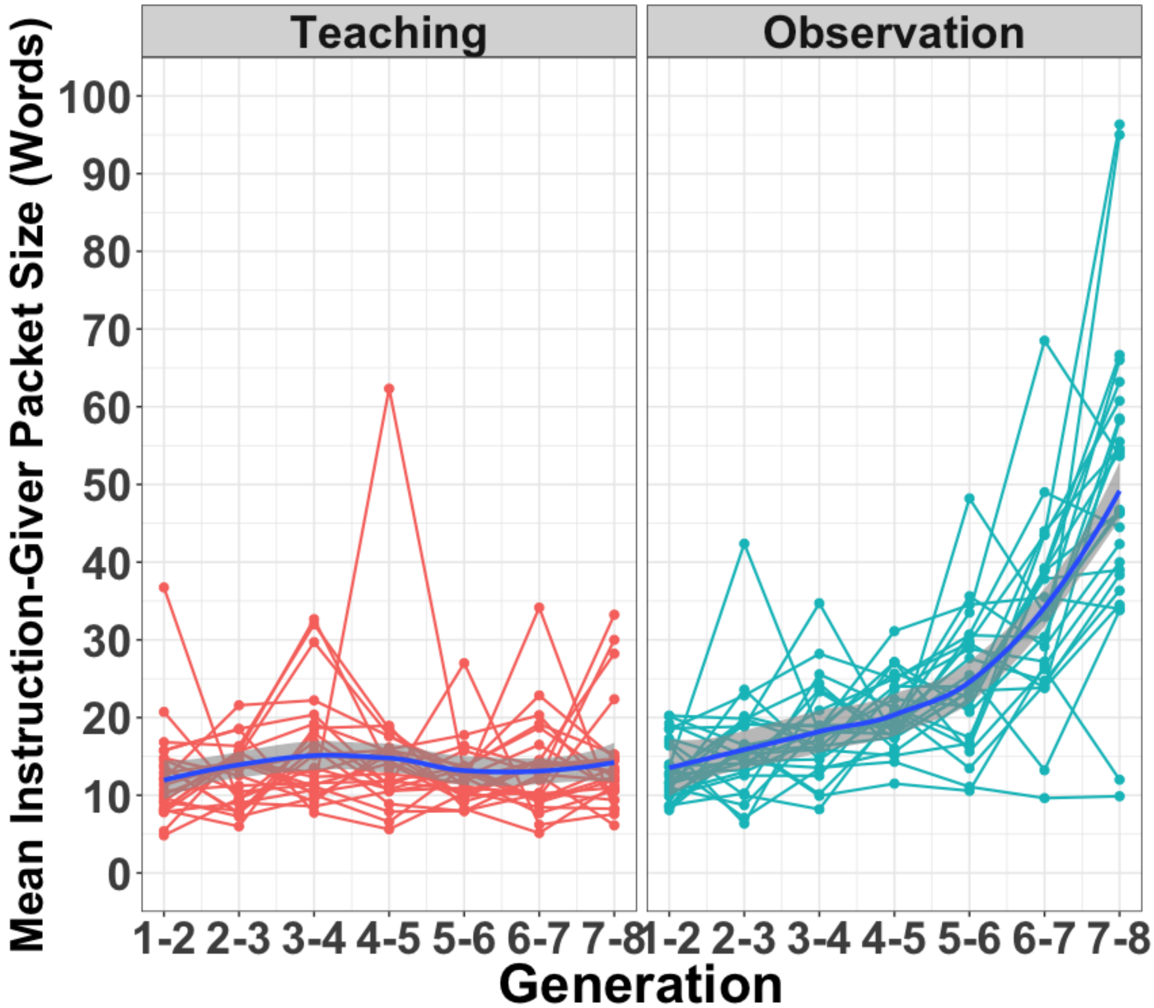

Figure 5. Change in Instruction-Giver mean packet size (in words) across generations in the Teaching and Observation conditions (plotted for each transmission chain). The blue line is the smoothed model fit and the grey shaded area is the 95% confidence interval.

Condition (Teaching, Observation) and Generation (1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8) with interaction were entered as fixed effects (see SM13 for the model specification and model output). The best fitting model specified Condition and Generation as fixed effects with interaction, accounting for 51.88% of the variance in Instruction-Giver packet size. Removing the interaction between Condition and Generation reduced the model fit, $\chi^2(1)= 49.36$, $p< .001$. To understand the interaction effect we analyzed each condition separately. In the Teaching condition the full model did not differ from the null model, $\chi^2(1)= 0.25$, $p= .618$, indicating no evidence for a statistical change in Instruction-Giver packet size over generations. By contrast, the Observation condition differed from the null model, $\chi^2(1)= 27.87$, $p< .001$; in this condition the mean Instruction-Giver packet size significantly increased over generations. **Figure 5** suggests the increase in Instruction-Giver packet size is non-linear, with the slope of the regression line becoming steeper from Generation 5-6. Growth curve analysis[48] indicated that adding a quadratic term for generation to the model improved model fit over a simple linear model, $\chi^2(1)= 43.68$, $p< .001$ (see SM14 for the model specification and model output).

Cultural Inheritance of Instruction-Giver Packet Size in the Observation Condition. If Instruction-Giver packet size at Generation N-plus-1 is predicted by Instruction-Giver packet size at Generation N, this would be evidence of the cultural inheritance of the language game.

Instruction-Giver packet size at Generation N-plus-1 was our dependent variable, and Instruction-Giver packet size at Generation N was entered as a fixed effect (see SM15 for the full model specification and model output). Compared against a null model, including Packet Size at Generation N improved model fit, $\chi^2(1)= 78.838$, $p< .001$, accounting for 39.92% of the variance in Instruction-Giver packet size at Generation N-plus-1, demonstrating cultural inheritance.

Instruction-Giver Packet Size and Route Reproduction Precision. Does the increase in Instruction-Giver packet size over generations in the Observation condition reflect a functional adaptation to the non-interactive context? To test this, we entered Packet Size as a fixed effect in a linear mixed effect model where route reproduction precision was the dependent variable (see SM16 for the full model specification and model output). Compared against a null model, including Packet Size improved model fit, $\chi^2(1)= 51.245$, $p< .001$, accounting for 21.13% of the variance in the route reproduction precision scores (see **Figure 6**). To examine the causal influence of packet size on route reproduction precision we tested if Packet Size at Generation N predicted route reproduction precision at Generation N-plus-1 (where Instruction-Giver packet size preceded task performance; see SM17 for the full model specification and model output). Compared against a null model, including Packet Size at Generation N improved model fit, $\chi^2(1)= 26.386$, $p< .001$, accounting for 13.34% of the variance in the route reproduction precision scores at Generation N-plus-1. The increase in Instruction-Giver packet size over generations in the Observation condition reflects a functional adaptation to the non-interactive context.

Finally, we tested if packet size positively contributed to route reproduction precision over and above the content communicated by Instruction-Givers in the Observation condition. First, the influence of the density of positive and negatively-biased route description tokens (as a percentage of total tokens) on the route reproduction precision scores in the Observation condition was modeled to provide a baseline. Positive and Negative route description token density were entered as fixed effects with interaction (see SM18 for the model specification and model output). The best fitting model specified Positive and Negative token density as fixed effects with interaction, accounting for 38.14% of the variance in the route reproduction precision scores. Removing the interaction between Positive and Negative token density reduced the model fit, $\chi^2(1)= 4.89$, $p= .027$. Next, Packet Size (at Generation N) was added to the model as a fixed effect. This improved the model fit over the model with Positive and Negative route description token density as fixed effects with interaction, $\chi^2(1)= 31.39$, $p< .001$, accounting for 47.05% of the variance in the route reproduction precision scores (i.e., explaining an additional 8.91% of the variance in route reproduction precision).

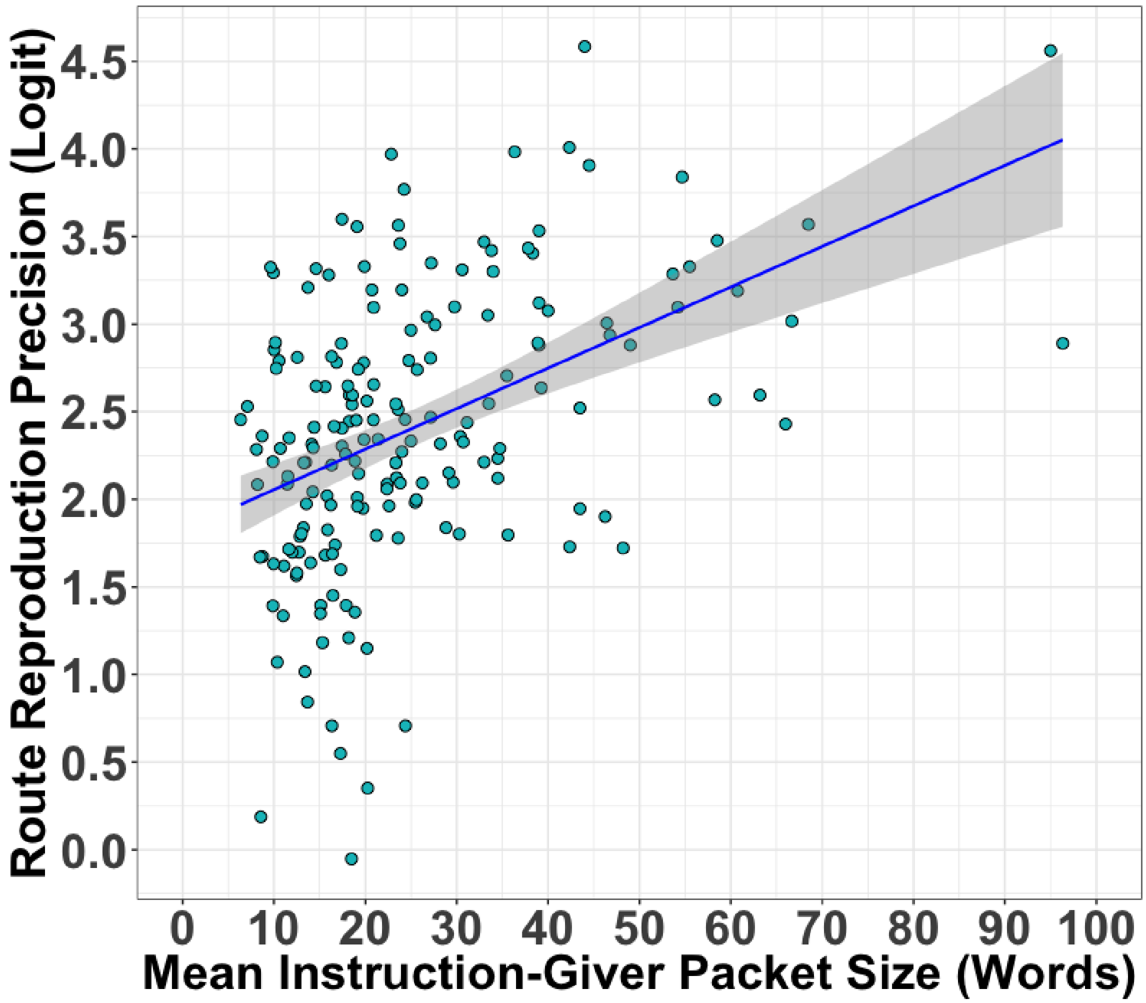

Figure 6. Positive Correlation between the Instruction-Giver mean packet size (in words) and Instruction-Follower route reproduction precision (Logit transformed) in the Observation condition (r=0.447, p<.001). The blue line is the linear model fit and the grey shaded area is the 95% confidence interval.

DISCUSSION

Instruction-Giver routes were reproduced with progressively higher precision by Instruction-Followers across the experimental generations. Cultural inheritance of route reproduction precision was also observed; route reproduction precision at Generation N predicted route reproduction precision at Generation N-plus-1 (Teaching and Observation conditions). While route reproduction precision improved across the experimental generations in both conditions, the routes communicated by Instruction-Givers in the Teaching condition were reproduced with higher precision compared to the routes communicated by Instruction-Givers in the Observation condition. This indicates that teaching, via active participation in the information exchange process, is important to cumulative cultural evolution.

The Cumulative Cultural Evolution of the Language Game

Next we examined the adaptive changes to the content of the language game that might explain the progressive improvement in route reproduction precision across the experimental generations, and between the Teaching and Observation conditions. Our key prediction was that teaching allows the teacher and student to negotiate route descriptions terms, and this acts as a cultural selection mechanism. This opportunity is not available to participants in the Observation condition, who must rely exclusively on individual inference. Route description terms that positively contributed to route reproduction precision accumulated, and terms that negatively affected route reproduction precision were progressively eliminated across the experimental generations (both conditions). The cultural inheritance of terms that positively contributed to route reproduction precision was observed; the density of positively-biased terms at Generation N predicted the density of positively-biased terms at Generation N-plus-1. By contrast, terms that negatively affected route reproduction precision were not inherited from the prior generation; instead they were progressively eliminated over the experimental generations. This pattern of results indicates that cultural selection was operating on the route description terms. As predicted, Teaching conferred a valuable cultural selection mechanism: the density of terms that positively contributed to route reproduction precision was higher, and the density of terms that negatively contributed to route reproduction precision was lower in the Teaching condition compared to the Observation condition. So, the cumulative improvement in route reproduction precision over the experimental generations was driven by the accumulation of adaptive modifications and innovations to the language game, and the elimination or non-adaptive content. This was most strongly seen in the Teaching condition.

The final set of analyses were exploratory, testing for adaptive changes to the language game process. Conversation is the basic medium of human communication[35,38], so we were interested in how participants in the Observation condition might adapt their communication to the non-interactive context. Instruction-Givers in the Observation condition adapted their communication to the non-interactive context: they changed their communication from a conversation-style, characterized by frequent and succinct information exchanges, to a narrative-style, characterized by infrequent but elaborate information exchanges. This process change, consistent across the transmission chains, was a powerful cultural tradition that was amplified over the experimental generations. Crucially, this process change was adaptive, being strongly associated with route reproduction precision. Furthermore, its positive influence on route reproduction precision extended beyond the adaptive changes to the route description content. By contrast, no process change was observed in the Teaching condition. In the Teaching condition a conversation-style between the Instruction-Givers and -Followers was maintained across the experimental generations.

The Observation Condition as a Basic Teaching Condition

Instruction-Givers in both conditions intentionally designed their route descriptions with their destination in mind (i.e., the Instruction-Follower)[49–51]. If observation learning is defined as learning from a model whose behaviour is unaffected by the actor observing them (citations, including social facilitation?), then the Observation condition in the present study could be argued to constitute a basic teaching condition. In the context of the present study, this would work against the experimental hypothesis; it would overestimate the benefit of observation learning, and underestimate the benefit of teaching to the cumulative cultural evolution of the language game.

We could have created an Observation condition in which the Instruction-Giver produced route descriptions from which s/he was told they would later try to reproduce the route, but have this done instead by the Instruction-Follower. This would create a *pure* observation learning condition, but it would also conflate communication intentionality with the conditions of primary interest. That is, it would be impossible to determine if the benefit of the Teaching condition over the Observation condition arose because of a difference in communication intentionality, or because of the opportunity Teaching affords the Instruction-Giver and -Follower to negotiate a mutually acceptable language game. Irrespective of how Observation learning is defined, the key finding from the present study stands; social collaboration between the Instruction-Giver and the Instruction-Follower proved important to the cumulative cultural evolution of the language game.

Transmission Chain Design Affects the Cultural Evolutionary Process

Experimental studies of cultural transmission among human participants have tended to focus on observation learning, so have relied on non-interactive transmission chain designs[19,52–57]. A benefit of using this design is that non-interactive transmission chains are easier to collect compared to interactive transmission chain designs. The transmission fidelity of a narrative text was enhanced when participants could directly interact with the adjacent chain member compared to when they could not[36]. This suggests that the difference between Teaching and Observation transmission chain designs on cultural transmission is quantitative, rather than qualitative. In the present study, a similar quantitative difference between the conditions was observed; the density of route description terms that positively contributed to route reproduction precision was higher in the Teaching condition, and the density of terms that negatively affected route reproduction precision was lower in this condition compared to the Observation condition. In addition, the profile of route description terms was similar across the Teaching and Observation conditions (see **Figure 4**).

By contrast, there was a qualitative difference between the conditions in the language game process. Whereas participants in the Teaching condition consistently adopted a conversation-style, the process adopted by participants in the Observation condition evolved from a conversation-style to a narrative-style over the experimental generations. Furthermore, this process change represented a functional adaptation to the non-interactive context. This qualitative difference between the Teaching and Observation conditions demonstrates that the transmission chain design can influence the cultural evolutionary process. Future studies using the transmission chain design should be aware of this, and use this information to make an informed decision about the transmission chain design that might best capture the cultural evolutionary process being simulated. Language evolution experimental simulations that have added social interaction to traditionally observation-based transmission chain designs have showed this to dramatically alter the nature of the evolved linguistic systems; from systems that are primarily driven by ease of acquisition, to systems that are both easy to acquire and functionally adapted for use[19,58].

CONCLUSION

We conducted a large-scale experimental study (N=408) that tested for the cumulative cultural evolution of a language game[7,20]; in our study, a subset of language used to communicate a route on a map to another person. The study also examined the influence of two social learning mechanisms to the evolution of the language game: Teaching and Observation learning. Using a transmission chain design that simulated intergenerational transmission, we demonstrated that the language game evolved by cumulative cultural evolution: the routes were reproduced with progressively higher precision over the experimental generations, and this was driven by adaptive modifications and innovations to the language game that accumulated over the experimental generations (both conditions). Like materials artifacts, the 'language technology' evolved by cumulative cultural evolution, a position that has been controversial in linguistics[59–61]. Our findings also support an important role for teaching to the cumulative cultural evolution of the language game[6,31]; route reproduction precision was higher in the Teaching condition compared to the Observation condition. Analysis of the changes to the language game indicated that teaching allowed the teacher and student to collaborate to find mutually acceptable adaptive route description terms. This social collaborative process operated as a second selection mechanism (to individual inference) that enhanced the cumulative cultural evolution of the language game.

REFERENCES

1. Basalla, G. *The evolution of technology*. (Cambridge University Press, 1988).
2. Tennie, C., Call, J. & Tomasello, M. Ratcheting up the ratchet: on the evolution of cumulative culture. *Philos. Trans. R. Soc. B Biol. Sci.* **364,** 2405–2415 (2009).
3. Tomasello, M. *The cultural origins of human cognition*. (Harvard University Press, 1999).
4. Caldwell, C. A. & Millen, A. E. Experimental models for testing hypotheses about cumulative cultural evolution. *Evol. Hum. Behav.* 165–171 (2008).
5. Muthukrishna, M., Shulman, B. W., Vasilescu, V. & Henrich, J. Sociality influences cultural complexity. *Proc. R. Soc. Lond. B Biol. Sci.* **281,** 20132511 (2014).
6. Morgan, T. J. H. *et al.* Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat. Commun.* **6,** (2015).
7. Wittgenstein, L. *Philosophical investigations*. (Wiley-Blackwell, 2009).
8. Whiten, A. *et al.* Cultures in chimpanzees. *Nature* **399,** 682–685 (1999).
9. Whiten, A. The second inheritance system of chimpanzees and humans. *Nature* **437,** 52–55 (2005).
10. Boyd, R. & Richerson, P. J. Why Culture is Common but Cultural Evolution is Rare. *Proc. Br. Acad.* **88,** 77–93 (1996).
11. Lewis, H. M. & Laland, K. N. Transmission fidelity is the key to the build-up of cumulative culture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367,** 2171–2180 (2012).
12. Gilligan, I. The prehistoric development of clothing: archaeological implications of a thermal model. *J. Archaeol. Method Theory* **17,** 15–80 (2010).

13. Stout, D., Toth, N., Schick, K. & Chaminade, T. Neural Correlates of Early Stone Age Toolmaking: Technology, Language and Cognition in Human Evolution. *Philos. Trans. Biol. Sci.* **363,** 1939–1949 (2008).

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

15. Mesoudi, A. & Whiten, A. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philos. Trans. R. Soc. B-Biol. Sci.* **363,** 3489–3501 (2008).

16. Whiten, A. & Mesoudi, A. Establishing an experimental science of culture: animal social diffusion experiments. *Philos. Trans. R. Soc. B Biol. Sci.* **363,** 3477–3488 (2008).

17. Goren-Inbar, N. Culture and cognition in the Acheulian industry: a case study from Gesher Benot Yaʿaqov. *Philos. Trans. R. Soc. B Biol. Sci.* **366,** 1038–1049 (2011).

18. Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B. & Laland, K. N. Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture. *Science* **335,** 1114–1118 (2012).

19. Kirby, S., Cornish, H. & Smith, K. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci.* **105,** 10681–10686 (2008).

20. Levinson, S. Activity Types and Language. *Linguistics* **17,** 365–399 (1979).

21. Piantadosi, S. T., Tily, H. & Gibson, E. Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci.* **108,** 3526–3529 (2011).

22. Zipf, G. K. *Human Behavior and the Principle of Least Effort*. (Addison-Wesley, 1949).

23. Garrod, S., Fay, N., Lee, J., Oberlander, J. & MacLeod, T. Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cogn. Sci.* **31,** 961–987 (2007).

24. Fay, N., Garrod, S., Roberts, L. & Swoboda, N. The interactive evolution of human communication systems. *Cogn. Sci.* **34,** 351–386 (2010).

25. Tamariz, M., Ellison, T. M., Barr, D. J. & Fay, N. Cultural selection drives the evolution of human communication systems. *Proc. R. Soc. B Biol. Sci.* **281,** (2014).

26. Fay, N., Garrod, S. & Roberts, L. The fitness and functionality of culturally evolved communication systems. *Philos. Trans. R. Soc. B-Biol. Sci.* **363,** 3553–3561 (2008).

27. Fay, N. & Ellison, T. M. The Cultural Evolution of Human Communication Systems in Different Sized Populations: Usability Trumps Learnability. *PLoS ONE* **8,** e71781 (2013).

28. Whiten, A., McGuigan, N., Marshall-Pescini, S. & Hopper, L. M. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos. Trans. R. Soc. B Biol. Sci.* **364,** 2417–2428 (2009).

29. Whiten, A., Horner, V., Litchfield, C. A. & Marshall-Pescini, S. How do apes ape? *Anim. Learn. Behav.* **32,** 36–52 (2004).

30. Bandura, A. *Social learning theory*. (General Learning Press, 1977).

31. Fogarty, L., Strimling, P. & Laland, K. N. The Evolution of Teaching. *Evolution* **65,** 2760–2770 (2011).

32. Tomasello, M., Kruger, A. C. & Ratner, H. H. Cultural learning. *Behav. Brain Sci.* **16,** 495–511 (1993).

33. Caldwell, C. A. & Millen, A. E. Social Learning Mechanisms and Cumulative Cultural Evolution: Is Imitation Necessary? *Psychol. Sci.* **20,** 1478–1483 (2009).

34. Schober, M. F. & Clark, H. H. Understanding by Addressees and Overhearers. *Cognit. Psychol.* **21,** 211–232 (1989).

35. Clark, H. H. *Using language*. (Cambridge University Press, 1996).

36. Tan, R. & Fay, N. Cultural transmission in the laboratory: Agent interaction improves the intergenerational transfer of information. *Evol. Hum. Behav.* **32,** 399–406 (2011).

37. Fusaroli, R. *et al.* Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychol. Sci.* **23,** 931–939 (2012).

38. Pickering, M. J. & Garrod, S. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* **27,** 169–226 (2004).

39. Anderson, A. H. *et al.* The HCRC map task corpus. *Lang. Speech* **34,** 351–366 (1991).

40. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2013).

41. Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. *R Package Version* **1,** (2013).

42. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68,** 255–278 (2013).

43. Bartoń, K. MuMIn: multi-model inference. *R Package Version* **1,** (2013).

44. Anderson, A. H., Clark, A. & Mullin, J. Introducing information in dialogues: forms of introduction chosen by young speakers and the responses elicited from young listeners. *J. Child Lang.* **18,** 663–687 (1991).

45. Jaeger, T. F. Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *J. Mem. Lang.* **59,** 434–446 (2008).

46. Haldane, J. The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Hum. Genet.* **20,** 309–311 (1956).

47. AJ Viera & JM Garrett. Understanding interobserver agreement: the kappa statistic. *Fam. Med.* **37,** 360–363 (2005).

48. Winter, B. & Wieling, M. How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *J. Lang. Evol.* **1,** 7–18 (2016).

49. Bell, A. Language Style as Audience Design. *Lang. Soc.* **13,** 145–204 (1984).

50. Clark, H. H. & Murphy, G. L. Audience design in meaning and reference. *Adv. Psychol.* **9,** 287–299 (1982).

51. Fussell, S. R. & Krauss, R. M. Understanding friends and strangers: The effects of audience design on message comprehension. *Eur. J. Soc. Psychol.* **19,** 509–525 (1989).

52. Bebbington, K., MacLeod, C., Ellison, T. M. & Fay, N. The sky is falling: evidence of a negativity bias in the social transmission of information. *Evol. Hum. Behav.* **38,** 92–101 (2017).

53. Mesoudi, A. & Whiten, A. The hierarchical transformation of event knowledge in human cultural transmission. *J. Cogn. Cult.* **4,** 1–24 (2004).

54. Ravignani, A., Delgado, T. & Kirby, S. Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* **1,** 7 (2016).

55. Kalish, M. L., Griffiths, T. L. & Lewandowsky, S. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychon. Bull. Rev.* **14,** 288–294 (2007).

56. Tamariz, M. & Kirby, S. Culture: copying, compression, and conventionality. *Cogn. Sci.* **39,** 171–183 (2014).

57. Bartlett, F. C. *Remembering: a study in experimental and social psychology*. (Cambridge University Press, 1932).

58. Kirby, S., Tamariz, M., Cornish, H. & Smith, K. Compression and communication in the cultural evolution of linguistic structure. *Cognition* **141,** 87–102 (2015).

59. Berwick, R. C., Friederici, A. D., Chomsky, N. & Bolhuis, J. J. Evolution, brain, and the nature of language. *Trends Cogn. Sci.* **17,** 89–98 (2013).

60. Pinker, S. & Jackendoff, R. The faculty of language: what's special about it? *Cognition* **95,** 201–236 (2005).

61. Christiansen, M. H. & Chater, N. Language as shaped by the brain. *Behav. Brain Sci.* **31,** 489–509 (2008).