Matthews C, Roberts AG & Caldwell CA (2012) Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory, *Evolution and Human Behavior*, 33 (6), pp. 759-770.

This is the peer reviewed version of this article

Matthews, C., Roberts, G., & Caldwell, C. A. (2012). Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior, 33*, 759-770.

Opportunity to Assimilate and Pressure to Discriminate can Generate Cultural Divergence in the Laboratory

Cristina Matthews
Gareth Roberts
Christine Caldwell

Psychology, School of Natural Sciences, University of Stirling

Formal models of cultural evolution have illustrated circumstances under which behavioral traits that have no inherent advantage over others can undergo positive selection pressure. One situation in which this may occur is when the behavior functions as a social marker, and there is pressure to identify oneself as a member of a particular group. Our aim in the current study was to determine whether participants organized into subpopulations could effectively exploit variation in a completely novel behavior to advertise themselves as belonging to a particular subpopulation, such that discrimination between in-group and out-group members was possible and subpopulations exhibited increasing distinctiveness. Eighty participants took part, organized into four subpopulations, each comprised of five four-member generations. They each completed a tower-building task, used in previous experimental studies of cultural evolution. An incentive payment structure was imposed with the aim of motivating participants to advertise themselves as belonging to a particular subpopulation, and to distinguish in-group members from members of other subpopulations. The first generation were exposed to photographs of randomly-assigned "seed" towers, and later generations were exposed to photographs of the towers built by the members of the previous generation of their own subpopulation. Participants were able to discriminate towers built by in-group members of the same generation, from towers built by out-group members. Over generations, tower designs evolved such that they were increasingly identifiable as belonging to a particular subpopulation. Arbitrary traits, which had no prior advantage, became associated with group membership, providing empirical support for theoretical models.

1. Introduction

Individuals have a range of motivations for preferentially interacting with those that share similar social backgrounds to themselves. These include potential protection against socially mobile free riders (e.g. Enquist & Leimar, 1993), as well as an increased chance of smooth and mutually beneficial interactions due to shared social norms (e.g. McElreath et al., 2003). In humans, culturally transmitted behaviors can potentially provide reliable information about the social background of the individuals displaying them, and hence we should expect people to be relatively adept when it comes to exploiting regularities, similarities, and variation, in behavior for the purpose of making such distinctions (Roberts, 2008). The result of this can be a change in the selective environment favoring the cultural (and possibly also biological) evolution of traits that were not previously advantageous (Efferson et al., 2008; Laland et al., 2000; Richerson & Boyd, 2005).

Previous research has identified language as particularly well suited for this purpose (e.g. Nettle & Dunbar, 1997; Roberts, 2008). In Nettle and Dunbar's (1997) model, individuals can both give and receive, with benefits to receivers outweighing costs to givers in accordance with the classic prisoner's dilemma. Although cheats (who receive but never give) can invade a population of cooperators (who both give and receive), introducing dialects can help to stabilize cooperation. In the model, dialect learning cooperators ("polyglots") give only to those with a similar dialect to their own, and modify their dialect on receipt of a gift to match their benefactor. The dialects also have a mutation rate which prevents them from stabilizing completely within an exchange group. Dialect learning cheats ("mimics") also copy dialects on receipt of a gift, but never reciprocate. However, since these mimics have more restricted learning opportunities (due to being unable to engage in repeat interactions with previously exploited individuals) they cannot keep up with dialect changes as effectively as the cooperative polyglots, and therefore receive fewer gifts.

Roberts (2008; 2010) carried out experimental work inspired by Nettle and Dunbar's (1997) simulations. In Roberts' studies, participants were assigned a teammate and engaged in an exchange with the goal of maximizing the score of their team. Participants were thus be motivated to exchange as much as possible with teammates, whilst attempting to benefit from their interactions with competing team members without giving anything. Participants were required to communicate using an invented (and highly restricted) language provided by the experimenter, which ensured that they were given no indication of who they were interacting with at any given point, and also that their only means of signaling their identity to their teammate was through non-standard use of the provided language. Over a series of rounds, distinctive dialects of the artificial language were developed such that teammates could effectively identify one another, and engage in preferential exchange together in order to increase their team's score.

The role of linguistic cues as potential social markers is therefore relatively uncontroversial, but in theory any marker that has the potential to become associated with group membership can be used for this purpose. Indeed, given the seemingly boundless range of human cultural behavior, it is highly likely that many observable traits are readily exploited (either explicitly or implicitly) as cues to social background, and hence also likely loyalties and social norms. McElreath et al. (2003) describe a model designed to investigate the capacity for two alternative markers to become associated with norms of interaction, such that they might be used to predict the outcome of a coordination game in which payoffs were highest for pairs of players that selected the same behavior. In their model there were only two possible markers, representing a much simpler and potentially more general example of behavioral variation, compared with Nettle and Dunbar's (1997) model of malleable and mutating dialects. In McElreath et al.'s model, mixing between populations (which may have settled on different coordination optima) can generate covariation between markers and behavior, potentially providing a useful cue to likely coordination or miscoordination. In their simulations they found that such associations between marker and behavior thus tended to increase over time due to the higher payoffs earned by individuals with the most common combinations of marker and behavior, and selective copying of successful individuals.

Efferson et al. (2008) followed up this model with an experimental study in the laboratory. Two subpopulations of participants were formed, with each placed in separate virtual environments, which had differing optimal coordination strategies. Within these subpopulations, participants learned to coordinate on the more beneficial behavior for their environment. However, participants were periodically swapped between subpopulations, resulting in their expectations about behavior conflicting with the local norms. Participants were also required to display one of two alternative (completely arbitrary) markers, and could decide whether or not to preferentially interact with those displaying the same marker as themselves. Although markers were initially randomly selected, and payoff-irrelevant, covariation between markers and behaviors accumulated over the course of the game such that it eventually became a very reliable predictor of behavior, and participants thus were inclined to choose to interact with those displaying the same marker.

Our aim in the current study was to use a novel non-linguistic task with an open-ended continuum of possible solutions, in order to investigate whether this could be exploited as a social marker, and if so, whether such use would result in increasing distinctiveness of the marker between groups. The task used was the spaghetti tower building task used by Caldwell & Millen (2008a; 2010), which has previously been shown to generate spontaneous variation in solution types across laboratory subpopulations of participants, even when participants are given an objective goal (maximize height), rather than a social one, as their target. These previous studies (Caldwell & Millen, 2008; 2010) also indicate that solutions to the task show an accumulation of modifications over experimental generations of participants, characteristic of cultural evolution. The task is therefore well suited to potential exploitation as a marker of group membership, and also as a candidate for cultural divergence between groups as a result of selection for (and possible exaggeration of) arbitrary traits.

In the current study therefore we wished to create a situation in which participants would be motivated to show favoritism to in-group members (of experimentally-generated subpopulations). The benefits of in-group favoritism should generate a motivation to attempt to advertise group membership through one's own tower design, as well as a motivation to discriminate towers built by those with a similar social background to oneself (i.e. those with shared experience of exposure to the same set of previous designs in this case), compared with those from a different social background (exposed to a different set of towers).

Our design therefore has much in common with classic studies of intergroup processes from the social psychology literature, which have experimentally created temporary, and anonymous, social groupings based on arbitrary distinctions (Tajfel's famous "minimal group" paradigm, e.g. Tajfel, 1970). Studies carried out by Tajfel and others (see Brewer, 1979, for a review) indicate that participants categorized in this way show reliable tendencies to favor the in-group and discriminate against the out-group. Therefore, our intention is not to test such effects, but to pose the question of the likely population-level effects of such in-group favoritism over multiple learner generations. Mesoudi (2009) has drawn attention to the ways in which the two research fields of cultural evolution and social psychology can inform one another, and notes that while social psychology provides insights into cultural micro-evolutionary processes, cultural evolution contributes population thinking, linking these small-scale processes to macro-evolutionary patterns of variation in time and space. Mesoudi (2009) also notes that multi-generational experimental designs (to date relatively underused within the field of social psychology) present a valid method by which the population-level effects of individual social psychological processes can be investigated.

Therefore, unlike the previous experimental studies of divergence in group markers by Roberts (2008; 2010) and Efferson et al. (2008), our groups consisted of multiple (experimental) generations of participants. Such designs have been variously referred to in the literature as either "microcultures" (Gerard et al., 1956), or "microsocieties" (Baum et al., 2004; Caldwell & Millen, 2008b). In such studies, participants typically complete the same task in succession, with the opportunity to either interact with, or observe the solutions of, their immediate predecessors (i.e. the previous experimental generation). We were therefore looking for increasing distinctiveness in behavior over generations, rather than over successive interactions between the same individuals as in the previous experimental work. Thus, if increasing distinctiveness was observed in our experiment, it would be attributable to an emergent outcome of individual copying decisions, making it more analogous to real world cultural evolution.

We expected that participants would be able to effectively signal group membership using their own tower design, following exposure to the previous generation's towers, and that participants would therefore perform above chance when attempting to distinguish towers built by fellow in-group members (exposed to the same previous generation towers as themselves) from those built by out-group members (exposed to a different set of previous generation towers). It was also predicted that, over generations, group-specific designs would develop such that towers built by members of the same generation of the same subpopulation would be significantly more similar to each other than they were to the towers built by members of the same generation of the other subpopulations.

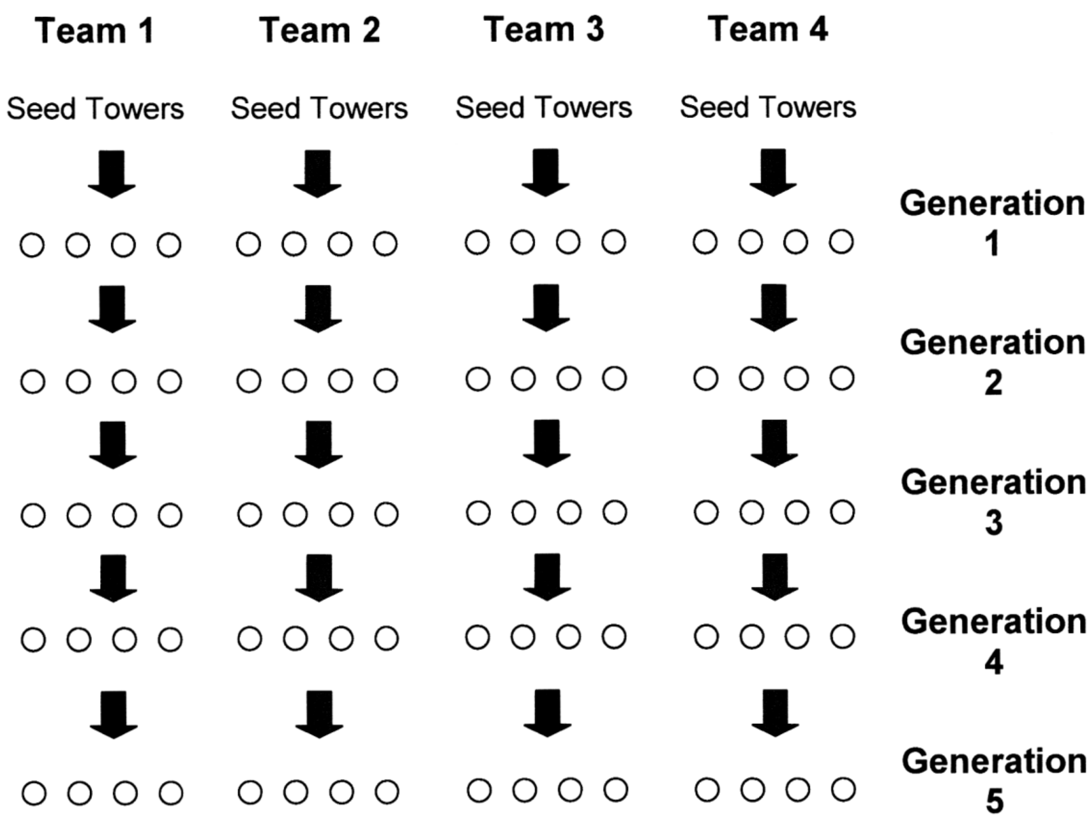
2. Method

2.1 Participants

Eighty participants were recruited from the University of Stirling and the University of Edinburgh. All participants took part in both sessions. Forty-six participants were females. The majority of participants (54) were in the age range 18-24. Nineteen were aged 25-34, and seven were aged 35-44. Participant rewards were paid in cash. Participants who were Psychology students at the University of Stirling could exchange £2.50 of their reward (or if the reward obtained was less, the entire reward) for 1 experimental participation credit (one quarter of their full requirement for the semester). The rest of the reward was paid in cash. Ethical approval for the study was granted by the University of Stirling Psychology Ethics Committee. All participants gave written consent prior to taking part in the experiment.

Participants were assigned to a “subpopulation” and “generation” (see Fig. 1). There were five generations of participants, and four subpopulations. Within each subpopulation, there were four members within a generation.

Figure 1. The structure of the laboratory population. Each circle represents a participant. Participants were organized into four subpopulations, comprised of five generations. In session 1 (Assimiliation Phase) participants were shown the four towers built by members of the previous generation of their subpopulation. In session 2 (Discrimination Phase) they were shown the three towers built by the three members of the same generation of their own subpopulation, plus one tower built by a member of the same generation of one of the other subpopulations.



2.2 Stimuli: Seed Towers

The experimenters built the seed spaghetti towers which were shown to participants in the first generation. The seed towers were each built to one of four different pre-specified designs, and four towers were built of each type, so that there were 16 towers in total. The 16 towers were then randomly distributed across subpopulations, with four seed towers for each subpopulation.

2.3 Presentation of Stimuli

A script which selected the towers that each participant would see in each session of the experiment was written by author GR in Python 2.7. The scripts that stored the data and collected the token allocations were also written in Python 2.7. The user interfaces were created using the Tkinter module.

2.4 Incentive Payment Structure

The incentive payment structure was devised primarily to motivate participants to discriminate between the towers built by their in-group members (who had seen the same set of previous towers as themselves), from those built by out-group members (participants that had seen a different set of towers). It was also designed to motivate participants to build their own towers such that their fellow in-group members would be able to identify them in exactly this way. Payments were determined by the outcome of a follow-up session (the “Discrimination Phase”, see section 2.5) which occurred after participants had seen the previous generation’s towers and built their own (during the “Assimilation Phase”). In the follow-up session, participants were shown photographs of the towers built by the three members of the same generation of their own subpopulation, plus one photograph of a tower built by a member of the same generation from a different subpopulation (all previously unseen by this participant). Participants were asked to allocate nine tokens (each worth 50 pence) to these towers in any distribution they wanted, on the understanding that if they allocated a token to a in-group member both the participant and the recipient would get a “share” of 25 pence each. Alternatively, if they allocated any tokens to the out-group member this resulted in a direct transfer of the 50 pence to that recipient. The payment structure was devised in this way in order to simulate a situation in which favors towards members of one’s own community were likely to be reciprocated during future encounters (hence the payment share), whereas members of different communities who were simply passing through would be likely to move on before favors were repaid (hence the direct transfer of the full value). Although participants could earn the full value of tokens through being mistaken for an in-group member by an out-group member, each participant only had the opportunity to earn tokens this way from one participant, whereas their tower photograph was shown to three in-group members. Again, this was intended to simulate a situation in which one was more likely to encounter members of one’s own community, and only more rarely venture into other communities.

Therefore (in addition to a £1 base fee), all participants had the potential to earn up to £2.25 through their own choices (by allocating their tokens only to in-group members), and up to £6.75 through the choices of their fellow in-group members (through being allocated the maximum number of tokens by all three of these participants), and up to £4.50 through the choice of an out-group member (should they be allocated the maximum number of tokens by this participant). The minimum and maximum theoretically possible earnings were £1 and £14.50 respectively. In the actual experiment, the smallest reward received was £2.75 and the largest was £10.25. The mean reward was £5.50.

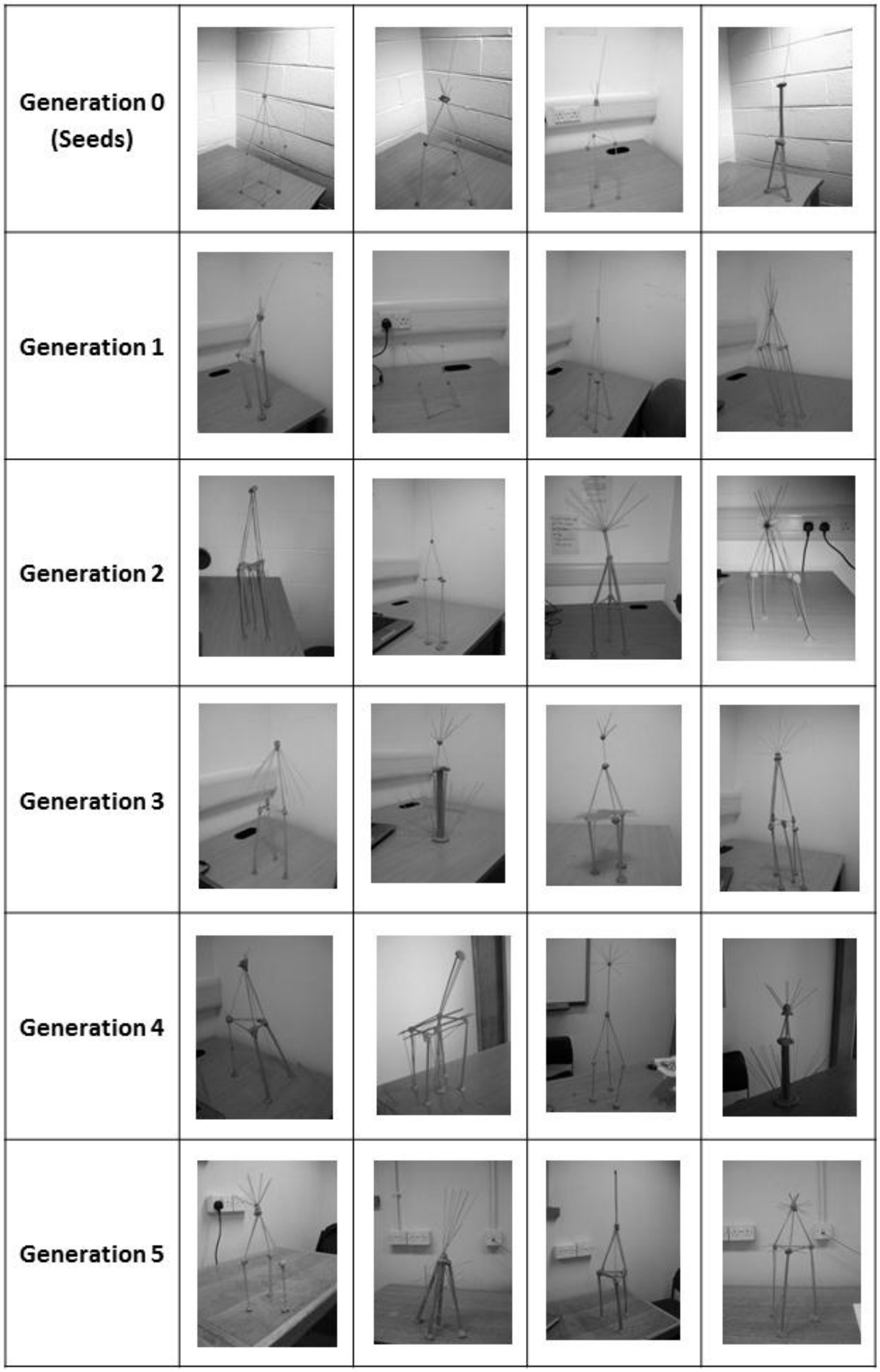
2.5 Procedure

All participants took part in two sessions. In the first session (here termed the “Assimilation Phase”), they were presented with photographs of four spaghetti towers and asked to build one themselves. It was explained to participants that the towers in the photographs were those built by the four participants in the preceding generation of their own team (i.e. subpopulation). Under each photograph was stated the total number of tokens that the tower had been allocated. The first generation saw seed towers that were built by the experimenters. As these towers had not been rated, no information about tokens was shown. The token allocations, which would take place in the second session, and the reward system, were explained to them. It was explained to participants that in the follow-up session they could maximize their earnings by selecting towers built by members of their own team, who had been shown the same towers as themselves, from the previous generation. Participants were encouraged to build a tower that their fellow participants would be able to identify as belonging to their own team. The only means by which participants might identify each other as fellow teammates was through their choice of design in the spaghetti tower building task. Participants were asked to read the information sheet and the experimenters ensured that participants understood all parts of the experiment – particularly the payment structure, due to its complicated nature - before proceeding. Participants were each provided with one 500g packet of raw spaghetti, plus around 200g of red Newplast™ modeling material for the purpose of building their own tower. They were given ten minutes in which to inspect the previous generation towers and build their own. Examples of the towers produced by the participants of one subpopulation are shown in Figure 2, and all towers produced as part of the experiment are provided as supplementary online material (see Fig. S1, available on the journal's website at www.ehbonline.org).

Participants returned for a second session (here termed the "Discrimination Phase"), typically between one and two weeks after they completed the Assimilation Phase. At the start of the session, each participant was given a brief reminder of the payment structure and the previous generation towers which they had seen during the Assimilation Phase. Participants were then shown a different set of four photographs of towers. Three of these towers had been built by the participant's in-group members of the same generation and one had been built by a participant from a different subpopulation from the same generation. Participants were asked to allocate their nine tokens (see section 2.4 above) to these towers in any distribution they wanted. Tokens were allocated by the participant entering a numeral in a box positioned below the photograph of the tower (a box could be left blank if no tokens were to be allocated, or a zero could be entered). When they had allocated all nine tokens and made their final decision, they clicked 'OK' and the information was stored. The program would not store the data if the total number of tokens was not equal to nine, and in this case the participant would be warned by a message appearing on screen stating, for example, "You still have 3 tokens to give away!". If the tokens allocated did add up to nine, the message would simply say, "Thanks!". Once all the participants in a generation had taken part in both sessions of the experiment, the rewards were calculated and awarded.

Matthews, C., Roberts, G., & Caldwell, C. A. (2012). Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior, 33*, 759-770.

Figure 2. Tower designs from one complete subpopulation (of a total of four).



In both sessions, the four photographs were presented in a random order. All randomizations were done by computer using Python 2.7 scripts. The particular out-group tower that a participant saw in the second session was selected at random, such that one randomly chosen member of subpopulation 1 would see a randomly chosen tower from subpopulation 2, another member of the same group would see a tower from subpopulation 3, and another would see a tower from subpopulation 4. The remaining member would see a tower chosen at random from subpopulation 2, 3, or 4. Any given tower was shown to only one out-group member. Since every tower was also shown to the other three members of the same subpopulation, every tower would be seen (and could potentially be awarded tokens) four times: three times by in-group members and once by an out-group member.

2.6 Similarity Ratings: Method

When data collection was complete, all resulting tower photographs (including seed towers, so 96 photographs in total) were collated, and two raters, blind to the subpopulation membership of the towers, coded the similarity of the tower designs. Raters were asked to judge how similar pairs of towers were to each other, using seven point Likert-type rating scale, where 1 meant not at all similar and 7 meant very similar. Each tower was rated for its similarity to every other tower from the same generation, and every other tower from the previous and subsequent generations. So, each tower in generation 1 was compared to all other towers in generation 1, as well as all towers in generation 0, i.e. the seed towers, as well as all of those from generation 2.

A new script was written in order to obtain the similarity ratings, again in Python 2.7 and using the Tkinter module. The interface presented pairs of spaghetti towers. Under the photographs were seven buttons with the numerals from 1 to 7. Pairs of towers were presented in a random order, and it was also randomly determined which tower of each pair appeared on the left and which on the right. Every given pair was rated twice, and when the same pair was shown for the second time, the left/right presentation of that particular pair was reversed. Raters were simply required to click one of the buttons, at which point the rating was stored, and a new pair of towers was presented. There was also a button allowing the rater to go back to the previous pairing if a mistake was made. The rater could also quit at any time and resume the task from the point at which it had been left. Since each pair of towers was compared twice by each rater, it was possible to calculate intra-rater reliability by correlating the resulting pairs of similarity ratings comparing the same two towers, as well as inter-rater reliability by correlating the mean similarity scores given to each pair by the two raters. Intra-rater reliability was high (Rater 1 Spearman’s Rho: $r = 0.857$, $N = 2000$, $p < 0.0005$; Rater 2 Spearman’s Rho: $r = 0.864$, $N = 2000$, $p < 0.0005$). Inter-rater reliability was lower, but ratings were nonetheless still highly significantly correlated (Spearman’s Rho: $r = 0.653$, $N = 2000$, $p < 0.0005$). The mean of the two raters’ similarity ratings for any given pair of towers was used in the subsequent analyses.

2.7 Analysis

Analyses were carried out using both the token allocations data, and the similarity ratings data. Both datasets were non-normally distributed, so nonparametric statistical tests were applied where possible.

3. Results

3.1 Success in Identifying Group Members: Token Allocations

Participants were required to divide their nine tokens between the four tower photographs they were shown during session 2, three of which were built by their in-group members of the same generation, and one of which was built by a member of a another subpopulation. A chance level allocation to another individual was therefore 9/4 (2.25 tokens). Data from participants' token allocations showed that they were able to identify in-group members. Across all participants, the mean allocation made to each individual in-group member was 2.58 tokens (SD = 0.60), and the mean allocation to out-group members was 1.25 tokens (SD = 1.79). A Wilcoxon test showed that participants were allocating significantly more of their tokens to individual in-group members, compared with out-group members: $Z = 5.372$, $N = 80$, $p < 0.0005$.

Table 1 displays the descriptive statistics for token allocations over generations. The maximum possible average allocation to any individual in-group member was 3 tokens (i.e. if all nine tokens were allocated to in-group members and none to the out-group member). Participants were therefore performing close to ceiling level across all generations. The allocations of each generation were analyzed individually, and all but the very first generation were allocating significantly more tokens to in-group members, compared with out-group members. For generation 1, $Z = 1.560$, $N = 16$, $p = 0.119$; for generation 2, $Z = 2.801$, $N = 16$, $p = 0.005$; for generation 3, $Z = 3.318$, $N = 16$, $p = 0.001$; for generation 4, $Z = 2.419$, $N = 16$, $p = 0.016$; and for generation 5, $Z = 2.102$, $N = 16$, $p = 0.036$.

Table 1. Mean allocation of tokens to individual in-group and out-group members, during session 2 (Discrimination Phase) of the experiment, over generations 1-5. Standard deviations are given in parentheses. A flat distribution of tokens to participants would generate an average of 2.25 tokens, and the maximum possible average in-group allocation was 3 tokens (9 tokens all allocated to the three in-group members, and none to the out-group member).

	Gen 1	**Gen 2**	**Gen 3**	**Gen 4**	**Gen 5**	**All Gens**
Average allocation to individual in-group members	2.35 (0.74)	2.67 (0.42)	2.79 (0.38)	2.63 (0.51)	2.48 (0.79)	2.58 (0.60)
Average allocation to out-group member	1.94 (2.21)	1.00 (1.26)	0.63 (1.15)	1.13 (1.54)	1.56 (2.37)	1.25 (1.79)

3.2 Success in Being Selected: Tokens Received

Due to the payment structure, the mean number of tokens earned by participants was exactly nine. However, there was considerable variability in the number of tokens received by individual participants (min = 0, max = 21, *SD* = 4.99). The mean number of tokens earned by individual participants as a result of the total allocations by their three in-group members was 7.75 (min = 0, max = 19, *SD* = 0.50), whereas the mean number earned as a result of the allocation of the out-group member was 1.25 (min = 0, max = 9, *SD* = 1.79). Consequently, although tokens awarded by out-group members were worth twice as much to a recipient (due to the share-or-seize payment structure) participants nonetheless earned considerably more as a result of allocations split with in-group members (mean = £1.94), compared with those seized from out-group members (mean = £0.63).

3.3 Similarity Ratings: Results

Using the similarity ratings (see section 2.6) scores could be calculated for each participant for how similar her tower was to other towers from: a) the same generation of their own subpopulation; b) the previous generation of their own subpopulation; c) the same generation of the other subpopulations; and d) the previous generation of the other subpopulations. Descriptive statistics for these scores are displayed in Table 2. These scores could be used to determine the predictors of participants' success in being selected during the Discrimination Phase. Table 3 displays the matrix of correlations (Spearman's Rho due to the non-normal distributions involved) for the similarity rating scores, and the allocations received from in-group members and out-group members. As one might expect, the best predictor of allocations from in-group members, and total earnings, appeared to be the similarity between the tower in question and the towers from the previous generation of the same subpopulation (i.e. those towers that both recipient and in-group donors had had the opportunity to view during session 1, the Assimilation Phase).

Although all similarity scores (whether measuring the similarity to towers from the same or previous generations, from the same or different subpopulations) appeared somewhat related to earnings, it should be noted that the similarity scores themselves were all highly intercorrelated (Table 3). Despite the non-normal distribution of the dataset partial correlations were also carried out, in order to determine whether any of the similarity measures (other than similarity to towers from the previous generation of the same subpopulation) independently predicted total earnings. These reinforced the picture presented by the alpha-corrected correlation matrix in Table 3. When holding constant the similarity to towers from the previous generation of the same subpopulation, similarity to the previous generation of the other subpopulations remained significant ($r = 0.300$, $df = 77$, $p = 0.007$) but the other relationships did not (for similarity to the same generation of other subpopulations, $r = 0.177$, $df = 77$, $p = 0.119$; and for similarity to the same generation of the same subpopulation, $r = -0.014$, $df = 77$, $p = 0.903$). Figure 3 displays the mean similarity scores for participants according to brackets of earnings from the allocations of other participants.

Table 2. Mean similarity ratings for towers, when compared to towers of the same and previous generations, from the same or different subpopulations, over generations 0-5. Standard deviations given in parentheses.

	Gen 0 (seeds)	**Gen 1**	**Gen 2**	**Gen 3**	**Gen 4**	**Gen 5**
In-Group Same Generation	3.39 (1.89)	2.91 (1.73)	2.98 (1.43)	3.43 (1.76)	3.13 (1.52)	3.56 (1.38)
In-Group Previous Generation (viewed during session1: Assimilation Phase)	--- ---	3.29 (1.92)	3.44 (1.95)	3.60 (1.98)	3.51 (1.80)	3.51 (1.51)
Out-Group Same Generation	3.71 (1.90	2.73 (1.45)	2.53 (1.28)	2.58 (1.45)	2.76 (1.41)	2.80 (1.23)
Out-Group Previous Generation	--- ---	2.97 (1.64)	2.59 (1.34)	2.52 (1.32)	2.67 (1.45)	2.74 (1.33)

Figure 3. Total earnings from other participants' allocations in relation to the tower's similarity to others from the same and previous generations, from their own and other subpopulations (mean +/- *SD*).

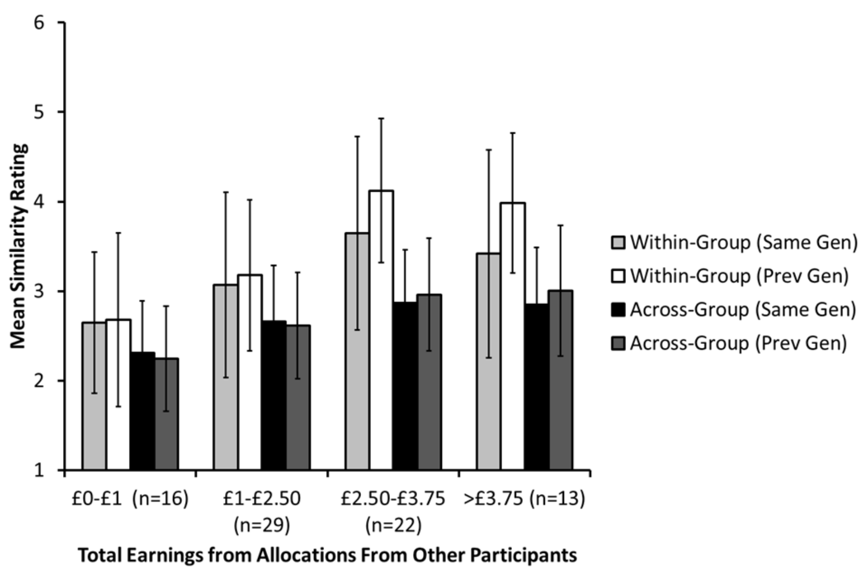


Table 3. Correlation matrix (Spearman's Rho) for measures of similarity and earnings from allocations from in-group members (IGMs), out-group members (OGMs), as well as total earnings from allocations (N = 80 for all). Results in bold remained significant following correction for multiple comparisons by using an adjusted alpha level of p = 0.0024.

		Earnings			Similarity Ratings		
		from IGMs	from OGMs	Total	In-Group Same Gen	In-Group Prev Gen	Out-Group Same Gen
Earnings	from OGMs	*r*=.120 *p*=.288					
	Total	***r*=.860 *p*<.001**	***r*=.582 *p*<.001**				
Similarity Ratings	In-Group Same Gen	*r*=.265 *p*=.018	*r*=.219 *p*=.051	*r*=.322 *p*=.004			
	In-Group Prev Gen	***r*=.608 *p*<.001**	*r*=.167 *p*=.139	***r*=.578 *p*<.001**	***r*=.587 *p*<.001**		
	Out-Group Same Gen	*r*=.250 *p*=.025	*r*=.193 *p*=.086	*r*=.312 *p*=.005	***r*=.338 *p*=.002**	*r*=.245 *p*=.029	
	Out-Group Prev Gen	***r*=.349 *p*=.001**	***r*=.336 *p*=.002**	***r*=.461 *p*<.001**	***r*=.399 *p*<.001**	***r*=.448 *p*<.001**	***r*=.722 *p*<.001**

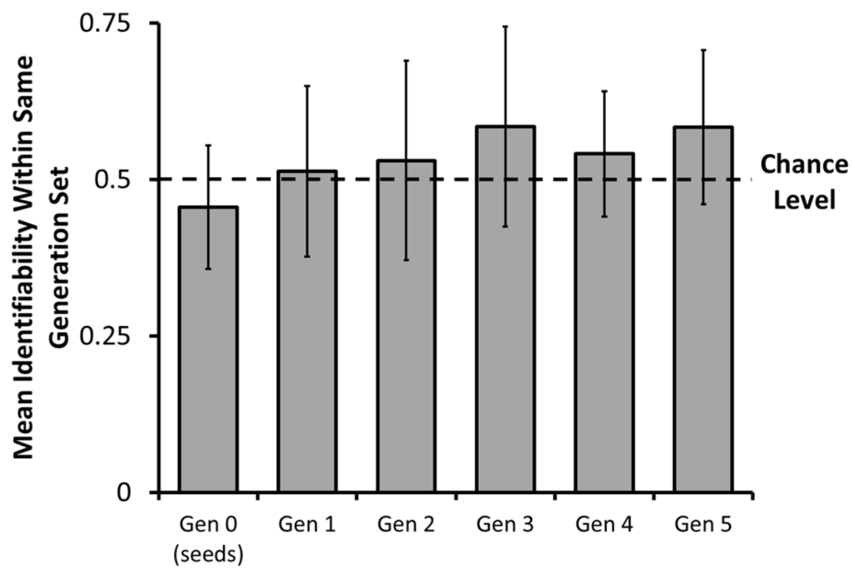
3.4 Cultural Divergence

We expected that, over generations, participants' towers would become increasingly easy to identify as belonging to a particular subpopulation, as groups developed distinguishing features. We therefore predicted that the similarity ratings comparing pairs of towers from the same generation of the same subpopulation would increase over generations, and that an objective measure of identifiability, calculated from the within-group and across-group similarities for each tower (as in previous studies, e.g. Caldwell & Millen, 2010) would also increase over generations.

The mean similarity rating for all towers built by participants, when compared with towers from the same generation of their own subpopulation, was 3.20 (SD = 1.07), whereas the mean similarity rating when being compared with towers from the same generation of other subpopulations was 2.68 (SD = 0.63). The within-group similarity ratings were significantly higher than the across-group similarity ratings (Wilcoxon test: Z = 3.940, N = 80, p < 0.0005).

We calculated the identifiability of towers as in Caldwell & Millen (2010). Similarity ratings were given on a scale of 1-7 (as detailed in section 2.6) so these were first transformed by subtracting 1 from each to give a scale of 0-6. This scale was then used to calculate each tower's similarity to others from the same subpopulation in relation to its similarity to towers from other subpopulations. A proportion [within-group similarity / (within-group similarity + across-group similarity)] was calculated, such that the range of possible values ran from 0 to 1, with values greater than 0.5 indicating higher within-group similarity ratings, and values less than 0.5 indicating higher across-group similarity ratings. This score was calculated for each individual tower, the resulting value providing an indication of the likelihood of identifying that particular tower as coming from that particular chain. Figure 4 displays descriptive statistics for this measure of identifiability across all generations, including the seed towers. It should be noted that the seed towers, which were randomly assigned to generation 0 of each chain, were in fact rather more *dissimilar* to others from the same subpopulation than would be expected by chance (also apparent from the separate within-group and across-group mean similarity ratings as displayed in Table 2). However, consistent with the direct comparisons between within- and across-group similarity, the measures of identifiability for all towers built by participants were significantly higher than an expected chance level of 0.5 (One-sample Wilcoxon test: $Z = 3.190$, $N = 80$, $p = 0.0005$, one-tailed).

Figure 4. Mean identifiability (+/- *SD*) of towers from their own generation set (calculated from same generation within-group and across-group similarity scores for each tower).



Each generation was also analyzed individually, including the seed generation, to determine whether that generation showed significant within-group similarity in relation to across-group similarity. Whether testing this by comparing within-group to across-group similarity measures, or using the identifiability measure discussed above, made no difference to the significance of any of the tests. Consequently, only the results using the identifiability measure are reported here. For the seed towers, there was a trend towards the towers being somewhat *less* identifiable than would be expected by chance (One-sample Wilcoxon test: Z = -1.655, N = 16, p = 0.098, two-tailed). For towers built by participants however, all generations showed above chance-level identifiability, although this was not significant for every generation. For generations 1, 2 and 4, identifiability was not significantly higher than would be expected by chance (Generation 1: Z = 0.724, N = 16, p = 0.235; Generation 2: Z = 0.621, N = 16, p = 0.268; Generation 4: Z = 1.448, N = 16, p = 0.074, all one-tailed). However, for Generation 3 and Generation 5, identifiability was significantly higher than would be expected by chance (Generation 3: Z = 1.965, N = 16, p = 0.025; Generation 5: Z = 2.379, N = 16, p = 0.009, both one-tailed). The result for Generation 5 also remains significant following Bonferroni correction for multiple comparisons, using an adjusted alpha level of 0.01 (0.05/5), although the Generation 3 result does not, so the Generation 3 result should be interpreted more cautiously.

We also used Page's L Trend Test, a nonparametric repeated-measures analysis predicting a monotonic relationship amongst treatment groups (Page, 1963), in line with our predictions of increasing identifiability and within-chain similarity over generations. For the current study, reducing the data to subpopulation-level averages produced a dataset consisting of only four replicates, giving limited statistical power, but doing so permitted repeated-measures analyses over generations in line with the predictions.

Using the subpopulation averages for identifiability, there was a significant tendency for increasing identifiability over generations, when including the seed generation towers (L = 330, m = 4, n = 6, p = 0.010), and this approached significance when excluding the seed generation (L = 195, m = 4, n = 5, p = 0.066). Within-chain similarity also showed an increase over generations. For this measure, this was not significant when including the seed generation, due to the high degree of similarity between *all* towers in this generation, including those from different subpopulations (see Table 2 for the relevant similarity measures): L = 305, m = 4, n = 6, p = 0.241. This was due to the way in which the towers for the seed generation were created (as detailed in section 2.2). However this was significant when considering only the towers built by participants: L = 197, m = 4, n = 5, p = 0.045).

4. Discussion

Consistent with our predictions, participants were able to build towers that could function as social markers, in terms of revealing which set of previous towers they had been exposed to. Also in line with predictions, this allowed participants to perform above chance when it came to picking out towers built by in-group members (who were exposed to the same set of previous towers as themselves) from those of out-group members (who saw a different set). The final predication was that, over generations, the designs used by the different subpopulations would differentiate (from a starting point of completely random variation between subpopulations), and this was also supported.

It is worth drawing particular attention to the fact that the measures used to determine increasing within-group similarity and distinctiveness over generations used the similarity ratings between pairs of towers from the *same* generation. Thus, the greater similarity to towers from the same subpopulation was not a consequence of direct exposure, since these participants did not see each other's towers until after they had built their own. Any within-group similarity between same generation towers arose as a consequence of the participants making similar decisions about what to copy from the previous generation's towers. Even early generation participants were able to build towers that were similar to those from the previous generation (to which they were exposed) such that their ingroup members could identify them (doing so at above chance levels from generation 2 onwards). It can also be seen that similarity to towers from the previous generation remained relatively stable over generations (Table 2). However, evidence of significant above-chance identifiability first emerged in generation 3, with a relatively robust above-chance effect arising in generation 5. Presumably in earlier generations participants were able to recognize some similarity between an in-group member's tower and the set of previous tower designs to which both had been exposed, without these participants necessarily building towers that were similar to each other's. It was only in later generations when the cumulative effect of similar copying decisions generated significant within-generation within-group similarity.

The within-generation similarity therefore appears to arise as an emergent consequence of the participants' attempts to model their designs on the previous generation. This occurred without any explicit attempt at coordination between in-group members, nor any opportunities for communication since participants never met one another. Nonetheless, there were good theoretical reasons for predicting this outcome, as it was expected that participants would exhibit similar biases in their copying of previous generation towers, and that this would generate the predicted differentiation between groups. Mathematical models of gene-culture coevolution (e.g. Boyd & Richerson, 1985) have illustrated that certain biases in social learning are highly adaptive, including positive frequency-dependent social learning, and model-based biases, both of which are relevant here. In positive frequency-dependent social learning (or conformity) the learner shows a disproportionate tendency to copy majority variants, and such a bias is capable of supporting within-group homogeneity and between-group variation (Efferson, Lalive, Richerson, McElreath & Lubell, 2008). If our participants were copying features of the previous generation towers which appeared to be most common, then over the generations this would result in greater homogeneity within the subpopulations.

Model-based biases could have a similar effect. Amongst potential model-based biases, particular attention has been given to the idea that learners might preferentially attend to, and learn from, successful individuals (e.g. Henrich & Gil-White 2001; Henrich & McElreath, 2007). Social learners who are selective in this way will tend to have an advantage over those who are less selective, as they have a greater chance of adopting behaviors associated with success. Since we provided participants with information about the tokens earned by each member of the previous generation, this made it possible to use a success-bias copying strategy, based on highest earnings. Again, this would be liable to generate greater similarity within subpopulations, assuming participants were using similar copying biases.

The details of our payment and population structure must of course be taken into consideration in interpreting the results, as it may well be the case that the outcome would be rather different should some of these details be altered. As noted in the methods, we made the decision that in session 2, the Discrimination Phase, participants would be asked to allocate their tokens between four towers, three of which were built by members of the same generation of their own subpopulation, and one of which was built by a member of one of the other subpopulations. This meant that to maximize earnings it was a good strategy to try to be identifiable as an in-group member by those from one's own subpopulation, even taking the share-or-seize payment structure into consideration. However, the average allocation per individual in-group member was 2.58 tokens (£0.65 when shared) and the average allocation per out-group member was 1.25 tokens (£0.63 seized), and therefore the actual earnings *per individual encounter* was roughly equivalent (see Table 1). As a consequence, had our game structure been different, with participants asked to choose between an equal number of in-group and out-group members, there might have been no particular advantage to those that were very similar to others from their own specific subpopulation compared with the others. All the same, in the absence of equivalent exposure to markers from any other subpopulation, explicitly attempting to pass oneself off as an out-group member does not seem viable as a strategy within our game structure, whereas trying hard to make oneself highly identifiable to the in-group would remain an option by which one could attempt to maximize earnings.

Interestingly, this point is related to another feature of our game structure, which, given the results, may provide an illuminating insight into the evolution of real world group markers. Since participants were required to allocate their tokens between a set of towers which included more than one (three, in this case) built by in-group members, this meant that for any given allocation there was not just competition between in-group and out-group members for the tokens, but also competition between in-group members. For the participant making the allocations, it was in their interests to allocate tokens in line with their confidence that a particular tower was built by a member of their own subpopulation. Furthermore, the zero-sum nature of our payment structure (since shared payments were exactly halved) meant that the development of group markers could not increase average earnings over generations since these were constant. This means that competition to be the most identifiable within one's own generation of a subpopulation is likely to be at least partly responsible for the divergence observed in our experiment. It is quite likely that such effects would extend to real world situations in which cultural evolution operates on group markers. When the behavior that functions as a marker shows continuous variation and has almost limitless possibilities for innovation, as in our experiment, then there is likely to be competition not just between those exhibiting the marker or not, but also between those who exhibit the marker to differing degrees. It is possible that this could lead to the increasing exaggeration of particular group-typical characteristics. Interestingly, such competition is also likely to prevent markers from becoming completely stabilized, maintaining the need for learning through exposure to recent exemplars (c.f. Nettle & Dunbar, 1997, who explicitly incorporate a mutation rate in their model of dialects in order to prevent stabilization).

It should also be noted that, within our game structure, during the Assimilation Phase participants were only exposed to towers from their own subpopulation. The study was intentionally designed in this way in order to reflect a situation in which naïve members of a community have the opportunity to learn from more experienced members of that particular community, and do not have the opportunity to learn from members of other communities. This meant that, although participants could copy towers from the previous generation in such a way that might allow theirs to be identified as *similar* to those ones (perhaps by copying features which were most common in the set of previous generation towers, or by copying a particular tower design which they assumed to be quite distinctive), they could not actively attempt to produce something that was *dissimilar* to the out-group. We would expect that (assuming it was still most beneficial to be highly identifiable as an in-group member) exposure to out-group towers in the Assimilation Phase would actually increase the divergence observed, and/or cause it to happen more rapidly.

The payment structure itself could of course be responsible for the outcome. We used direct financial incentives to motivate participants to advertise their group membership, and to discriminate the in-group from the out-group. As we detailed in the introduction, we were taking these factors as basic assumptions about human intergroup psychology, and therefore wanted to ensure that our participants did indeed behave in this way within our laboratory situation, allowing us to study the population-level consequences of such behavior. However, these human tendencies for identifying with and favoring the in-group are now extremely well understood, and they appear to be remarkably persistent. Studies in the laboratory show that groups formed on even the most trivial and irrelevant grounds nonetheless reliably exhibit these biases (Tajfel, 1970). Our payment structure therefore may have been crucial to motivate our participants to behave in this way, but it is also entirely possible that they would have shown these same motivations and biases in the absence of the incentive payment structure.

In addition, it must be taken into account that our task (in the context of this particular study) functioned only as a marker of group membership. In the real world, traits that can function as group markers often perform other functions which place constraints on their form. It remains to be seen how flexible group markers can be when there are competing functional pressures on the form of the behavior.

Our task involved building a material artifact, whereas some social markers (e.g. linguistic cues like dialects, see introduction) involve behaviors that leave no physical trace. The consequence of this is that different social learning mechanisms may be implicated in the transmission of different types of social marker. Copying another's behavior is generally referred to as *imitation*, with copying of products (e.g. material artifacts) defined as *emulation* (e.g. Whiten, 2011). Reproduction of linguistic cues therefore requires imitative learning, whereas copying from the photograph of a finished product (as in our task in the current study) depends solely on emulation. Caldwell, Schillinger, Evans and Hopper (2012) have discussed the imitation/emulation distinction in greater depth, with regard to the spaghetti tower building task. With this task, participants readily copy from finished products (including photographs of other towers, see Caldwell et al., 2012), and so the current design was adequate to capture these effects. However, to enable the transmission of some other types of social marker, direct interaction with another individual may be necessary, and indeed such interaction might even enhance social learning in a task such as this one.

With regard to the specific features which may have functioned as group markers in our subpopulations, it should be noted that there are certain limitations to what our similarity ratings data can reveal. We cannot tell from the ratings which features, if any, participants might have exploited in order to either discern or advertise group membership. All the same, we can see informal evidence of at least one particular, relatively distinctive, feature acting as a probable marker in the towers built by one of our four subpopulations (see Fig. 2). In this subpopulation subtle (and likely unintentional) features of early generation towers appear to have been deliberately exaggerated in later generations for the purpose of communicating group membership, resulting in a distinctive fan-like feature that appeared in the majority of towers from generation 3 onwards.

In conclusion, we found that, in an experimental laboratory situation, human participants could readily exploit variation in a novel behavior to both effectively assimilate themselves to a particular group, and to discriminate between others on the basis of their efforts to assimilate to a particular group. The outcome of this, over several experimental generations of participants, was that particular designs arbitrarily became associated with group membership. This provides a valuable insight into the likely population-level effects of well recognized psychological processes relating to intergroup behavior. On the basis of our findings, we conclude that many socially transmitted behaviors (assuming functionality places no major constraints on the form) can potentially reveal information about the social background of the actor and hence be exploited as a social marker.

Acknowledgments

This research was funded by the Economic and Social Research Council (Research Grant RES-062-23-1634). We thank the University of Edinburgh Department of Psychology for allowing us to use their research cubicles during data collection.

References

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

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.

Brewer, M. (1979). In-group bias in the minimal group situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*, 307-32.

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

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

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

Caldwell, C. A., Schillinger, K., Evans, C. L. & Hopper, L. M. (2012). End state copying by humans (*Homo sapiens*): Implications for a comparative perspective on cumulative culture. *Journal of Comparative Psychology, 126*, 161-169.

Efferson, C., Lalive, R. & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science, 321*, 1844-1849 (doi: 10.1126/science.1155805).

Efferson, C., Lalive, R., Richerson, P. J., McElreath, R., Lubell, M. (2008) Conformists and mavericks: the empirics of frequency-dependent cultural transmission. *Evolution and Human Behavior*, *29*, 56-64.

Enquist, M. & Leimar, O. (1993). The evolution of cooperation in mobile organisms. *Animal Behaviour, 45,* 747-757.

Gerard, R. W., Kluckhohn, C. & Rapoport, A. (1956). Biological and cultural evolution: Some analogies and explorations. *Behavioral Science, 1*, 6-34.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, *22*, 165-196.

Henrich, J., & McElreath, R. (2007). Dual inheritance theory: The evolution of human cultural capacities and cultural evolution. In R. Dunbar & L. Barrett (Eds.), *Oxford Handbook of Evolutionary Psychology*. Oxford, UK: Oxford University Press.

Laland, K. N., Odling-Smee, J. & Feldman, M. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, *23,* 131–175.

McElreath, R., Boyd, R. & Richerson, P. J. (2003). Shared norms and the evolution of ethnic markers. *Current Anthropology, 44,* 122-129.

Mesoudi, A. (2009). How cultural evolutionary theory can inform social psychology and vice versa. *Psychological Review, 116*, 929–952.

Nettle, D. & Dunbar, R. I. M. (1997). Social markers and the evolution of reciprocal exchange. *Current Anthropology, 38,* 93-99.

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

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

Roberts, G. (2008). Language and the free-rider problem: An experimental paradigm. *Biological Theory, 3,* 174-183 (doi: 10.1162/biot.2008.3.2.174).

Roberts, G. (2010). An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies*, *11*, 138-159 (doi: 10.1075/is.11.1.06rob).

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223,* 96-102.

Whiten, A. (2011). The scope of culture in chimpanzees, humans and ancestral apes. *Philosophical Transactions of the Royal Society, 366*, 935-1187.