Regular rates of popular culture change reflect random copying

R. Alexander Bentley\textsuperscript{a,}\textsuperscript{*}, Carl P. Lipo\textsuperscript{b}, Harold A. Herzog\textsuperscript{c}, Matthew W. Hahn\textsuperscript{d}

\textsuperscript{a}Department of Anthropology and the AHRC Centre for the Evolution of Cultural Diversity, University of Durham, 43 Old Elvet Road, Durham, DH1 3HN, UK
\textsuperscript{b}Department of Anthropology and the Institute for Integrated Research in Materials, Environments and Society (IIRMES), 1250 Bellflower Blvd., California State University Long Beach, Long Beach, CA 90840, USA
\textsuperscript{c}Department of Psychology, Western Carolina University, Cullowhee, NC 28723, USA
\textsuperscript{d}Department of Biology and School of Informatics, Indiana University, 1001 East Third Street, Bloomington, IN 47405, USA

Initial receipt 3 July 2006; final revision received 30 October 2006

Abstract

Almost by definition, “popular culture” reflects the effects of most people imitating those around them. At the same time, trends and fashions are constantly changing, with future outcomes potentially irrational and nearly impossible to predict. A simple null model, which captures these seemingly conflicting tendencies of conformity and change, involves the random copying of cultural variants between individuals, with occasional innovation. Here, we show that the random-copying model predicts a continual flux of initially obscure new ideas (analogous to mutations) becoming highly popular by chance alone, such that the turnover rate on a list of most popular variants depends on the list size and the amount of innovation but not on population size. We also present evidence for remarkably regular turnover on “pop charts”—including the most popular music, first names, and dog breeds in 20th-century United States—which fits this expectation. By predicting parametric effects on the turnover of popular fashion, the random-copying model provides an additional means of characterizing collective copying behavior in culture evolution.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Neutral theory; Random genetic drift; Pop music; Markets; Cultural evolution; Baby names; Cultural transmission; Power law; Fashion; Purebred dogs; Random copying

1. Introduction

As Boyd and Richerson (1985, p. 33) defined over 20 years ago, “culture is information capable of affecting individuals’ phenotypes which they acquire from other conspecifics by teaching or imitation.” Imitation is arguably the simplest form of culture transmission, termed unbiased transmission by Boyd and Richerson, which occurs when each individual acquires his or her behavior simply by copying from another individual within the population. Copying is a predominant human behavior (e.g., Gergely, Bekkering, & Király, 2004; Iacoboni et al., 1999) and is shared among primates (cf. Subiaul, Cantlon, Holloway, & Terrace, 2004). It can, thus, be useful to assume, as a null hypothesis in certain instances of social choice, that people simply copy each other at random. In cases where choices have intrinsic value with respect to one another, it makes more sense to assume that cost–benefit decisions are made independently, with conformity potentially among the biases in making those decisions (e.g., Boyd & Richerson, 1985, 2005; Gintis, in press; Henrich, 2001, 2004; Henrich & Boyd, 2001; McElreath, Boyd, & Richerson, 2003; Shennan, 2002). This distinction is crucial to the nature of collective human behavior, in anything from voting to corporate boardrooms, to deciding on a hunting strategy, as copying can tend toward baseles decisions, whereas independent decision making may lead to a rational, collective “wisdom” of a group (Surowiecki, 2004) and/or optimal solutions through a process analogous to natural selection (e.g., Crow & Aoki, 1982a, 1982b, 1984; Henrich, 2004). While there is a fairly large body of literature on group norms that arise as a consequence of identifiable costs and benefits of cultural traits, quantitative models of random copying of neutral cultural traits are relatively less well developed (see reviews by Eerkens & Lipo, 2005; Mesoudi,
Whiten, & Laland, 2006). Here, we focus on a particular prediction of the random-copying model, not to explain all human behavior, but to help identify it when it arises and further characterize the consequences of copying in collective behavior.

As we have shown in previous studies (Hahn & Bentley, 2003; Herzog, Bentley, & Hahn, 2004), a highly useful null hypothesis for popular culture change can be a process of random copying between individuals, akin to the process of random genetic drift in population genetics. With its great potential for future modification and development, there are many ways in which the random-copying model, with the resources of population genetics theory to support and develop it, can make substantial contributions to social science. Large-scale shifts in popular preferences (e.g., fashions) offer insight into general mechanisms of cultural change (Lieberson, 2000). Whereas the collective effect of independent decisions may be a sensible equilibrium, random copying is unpredictable, with no tendency toward an optimum. For example, a recent Internet-based sociological experiment (Salganik, Dodds, & Watts, 2006) demonstrated that popular success in music markets is as much a matter of social influence as of quality. A model that has proven surprisingly robust in explaining shifts in tastes assumes simply that the majority of individuals randomly copy the choices of others, with occasional innovation (Bentley, Hahn, & Shennan, 2004). In population genetics, a formal model of random copying between generations with mutation is called the neutral model (Kimura & Crow, 1964). While developed to explain genetic variability, the neutral model has been effectively applied to ecological and cultural phenomena (e.g., Cavalli-Sforza & Feldman, 1981; Dunnell, 1978; Hubbell, 2001; Lang & Barlow, 1997; Lipo, Madsen, Dunnell, & Hunt, 1997; Neiman, 1995). It predicts that, inevitably, some variants will become highly popular simply due to imitation, not because they are in some way “better” than other variants. We have found that the assumption of random copying provides realistic predictions of the frequency distribution and change in frequency over time of such diverse phenomena as Neolithic pottery decorations (Bentley & Shennan, 2003), baby names (Hahn & Bentley, 2003), and dog breeds (Herzog et al., 2004).

The random-copying model assumes that there are $N$ individuals, each characterized by a behavioral/stylistic variant (Fig. 1). At each time step, we refresh the population with $N$ new individuals, and each is assigned a new variant by either (a) copying a randomly selected individual from the previous time step, with equal probability of choosing any individual, or (b) inventing a new variant (gray lightning bolts) with probability $\mu$, the fraction of innovators among the $N$ individuals.

**Fig. 1.** A simple representation of the neutral-trait model. Shown are five individuals for three successive time steps. At each time step, we refresh the population with new individuals, and each is given a new copy of a variant (represented by numbers inside the circles). Each variant is assigned a new value by either (a) copying a randomly selected individual from the previous time step, with equal probability of choosing any individual, or (b) inventing a new variant (gray lightning bolts) with probability $\mu$, the fraction of innovators among the $N$ individuals.
We are not aware of any direct analytical solution to this problem since the sample includes only the most-frequent variants, which means that we cannot simply assume, as for an entire population at equilibrium, that the innovation rate balances the loss rate. Our approach, therefore, was to use computer simulation (Bentley et al., 2004; Hahn & Bentley, 2003), by which we run the random-copying model using different numbers of individuals, $N$, and innovation fractions, $\mu$. We then compared our simulation results to real-world data sets involving pop music, baby names, and dog breeds in the 20th-century United States.

2. Methods

As described in detail previously (Bentley et al., 2004; Hahn & Bentley, 2003), we used a simple computer simulation of the neutral model written in a Java-based simulation package called RePast (v 2.0, http://repast.sourceforge.net/).
As represented schematically in Fig. 1, the simulation begins with \( N \) individuals that are assigned \( N \) different variants, which are then subject to repeated copying and innovation (cf. mutation). The simulation records the occurrence of every variant to appear in the population throughout the run. At every time step, the \( N \) individuals are replaced with \( N \) new individuals, the majority of which receives a variant copied at random from the previous time step, while the remaining minority (\( N\mu \), where \( \mu \) is the innovation fraction) invents a novel variant (“innovation”). After running the simulation for 250 time steps to reach a quasi-equilibrium state, we recorded all the variants present in the population and their frequencies, for every other time step until time step 300 (25 total samples). At each sampling point, we recorded all the variants present and their frequencies. We then created Top \( y \) charts of varying sizes \( (y=5, 10, 20, 30, 40) \) for each of these samples. To determine the turnover rate among the Top \( y \) most-frequent variants, we ranked the variants by abundance for each sampled interval and then tabulated the number of new variants to enter the Top \( y \) chart relative to the previously sampled interval.

3. Results

The simulations all show that, after the transient phase of the first 250 time steps, the turnover rate \( z_y \) of a Top \( y \) list (the number of new variants to enter the Top \( y \) chart relative to the previously sampled interval) finds a steady state, where it fluctuates around a nominal average. In this steady state, as Fig. 2A shows, the average turnover is linearly proportional to the size of the list \( y \) \( (r^2=.970) \). There is, therefore, a strong dependence of turnover rate on the length \( y \) of the Top \( y \) list.

Our simulation results also showed that under the assumption of random copying, the turnover rate is predicted to be independent of the population size \( N \), as varying \( N \) from 250 to 2000 in the simulations (Fig. 2A) had no significant effect on the turnover rate \( (r^2=.003) \). In sum, the average turnover in the simulated random-copying model can be described simply as

\[
    z_y = Ay,
\]

where \( A \) is a constant, such that \( z_y \) is largely independent of \( N \). At the same time, the simulation results show that \( z_y \)
does depend on the innovation fraction $\mu$ (Fig. 2B). In fact, as Fig. 2C shows, the constant $A$ is simply proportional to $\mu$, with an $r^2$ of .991 ($p<.005$). Hence, we find the simple relation:

$$z_y = y\sqrt{\mu}.$$  

An example that appears to exhibit characteristics of the random-copying model involves the Billboard "Top 200" Pop Chart (Bentley & Maschner, 1999; Whitburn, 1985)—hereafter, the "Pop Chart"—a mainstream record of popular music in the United States. The number of new albums (or, equivalently, the number of exiting albums) per week on the Pop Chart jumped erratically about a nominal average—rather like the "El Farol" problem of how many people visit a bar from night to night (Arthur, 1999)—of around 7–8 albums per week from 1963 to 1967 when the Pop Chart had 150 albums. This average then increased to around 11 albums per week when the Pop Chart was expanded from 150 to 200 in mid-1967. Given that the population of the United States increased almost 70% during this time, from about 150 million in 1960 to about 250 million in 1990, and assuming a commensurate increase in albums sold, it appears that the steady turnover on the Pop Chart was independent of $N$, in line with the random-copying model.

Our next real-world example involves popular baby names, as recorded by the U.S. Social Security Administration (http://www.ssa.gov/OACT/babynames/) by ranking the 1000 most common boys' and girls' names in each decade of the 20th century. The rates at which new names appeared on the list averaged 182 female names and 133 male names per decade. Taking these turnover values as measures of $z_y$, we find the turnover rate for that the fraction of Pop Chart turnover $z_y/y$, in terms of the percentage change in the composition of the Pop Chart per week, will be constant. Indeed, the Pop Chart turnover was remarkably steady at 5.6% per week for over 20 years (Fig. 3B). Unfortunately, we cannot systematically test different values of $y$ because our source (Whitburn, 1985) only gives the date each album entered the Pop Chart and the date it exited, rather than the position of each album week to week. However, as Fig. 3B shows, there is no visible change in the fractional turnover rate, $z_y/y$, when the Pop Chart was expanded from 150 to 200 in mid-1967. Given that the population of the United States increased almost 70% during this time, from about 150 million in 1960 to about 250 million in 1990, and assuming a commensurate increase in albums sold, it appears that the steady turnover on the Pop Chart was independent of $N$, in line with the random-copying model.

Fig. 4. Turnover rate of Top $y$ charts, plotted against $y$, the size of the list, for (A) boys' names (filled circles) and girl's names (open circles) and for (B) dog breeds. For baby names, the turnover rates are per decade and averaged over the 20th century. For dog breeds, the turnover rates were calculated based on 4-year intervals. Error bars show $\pm 1\sigma$. 
female names to be 1.37 times (182/133) that for male names. The difference appears to be real and not by statistical chance, as the rate for female names was higher for every decade of the 20th century (Hahn & Bentley, 2003). The higher turnover rate for female names in each decade implies more innovation in naming girls, as is clear in other studies (see Fryer & Levitt, 2004; Levitt & Dubner, 2005, pp. 179–204; Lieberson, 2000).

As seen by the numbers of new names to enter the Top 20, Top 100, Top 500, and Top 1000 lists in each decade, the turnover rate clearly increases as $y$ increases, for both boys’ and girls’ names. Averaged over all decades of the century, the resulting linear relationship with $y$ (Fig. 4A) is as predicted by Eq. (1), for both girls’ ($r^2 = .96$) and boys’ ($r^2 = .99$) names. According to Eq. (2), each slope (0.259 for girls’ names, 0.176 for boys’ names) corresponds to the square root of the innovation rate $l$, yielding a century-averaged innovation parameter of 0.067 for girls’ names and 0.031 for boys’ names.

Finally, we obtained data from the American Kennel Club on the annual number of new puppy registrations in the United States for all recognized breeds (Herzog et al., 2004). This is a large (a total of 52,806,268 registrations from 1926 to 2004) and highly accurate index of the relative popularity of purebred dog breeds over the past five decades. Using these data, we created multiple Top $y$ lists for each year since 1926 by listing and comparing the set of top registered dog breeds in order of decreasing frequency for the year. For each year in the study, we used the total number of dogs registered as our measure of $N$. By sampling lists of different sizes, $y$, at regular time intervals, we then determined the turnover rate in each list per 4 years (sampling every year would leave too many zeros in the time series).

The number of different dog breeds registered in the United States has increased from 73 breeds in 1926 to 150 breeds in 2004. Nonetheless, throughout this time period, the pattern is as predicted by the random-copying model, which is an increase in turnover rate, $z_y$, as $y$ increases. As Fig. 4B shows, the turnover rate $z_y$ for dog breeds shows a convincing ($r^2 = .894$) linear relationship with $y$, and the slope of 0.124 corresponds via Eq. (2) to an innovation parameter $\mu$ of 0.015 for 4-year sampling.

4. Discussion

In accord with the evidence of copying behavior in downloading music (Salganik et al., 2006), the random-copying model provides a simple, parsimonious explanation for the steady turnover of modern baby names, dog breeds, and pop music albums over much of the 20th century. As is often noted in the social sciences, many models can fit the data, and we do not rule out the possibility that other models could be devised to fit the patterns we have shown. We advocate random copying as a null model firstly because it appears to be the absolute simplest model capable of replicating the data patterns at the societal scale and, secondly, because its two mechanisms, innovation and random copying, are the two most basic elements of unbiased culture transmission as defined by Boyd and Richerson (1985). In fact, there is growing evidence that in situations where cultural transmission occurs predominantly from one individual to another, a neutral or “random-copying” model is the best null model against which
real-world cultural variants can be compared (e.g., Bentley et al., 2004; Bentley & Shennan, 2003, 2005; Hahn & Bentley, 2003; Herzog et al., 2004; Lipo et al., 1997; Lynch, 1996; Neiman, 1995; Simkin & Roychowdhury, 2003). Use of this model in future studies will make it easier to test for more detailed effects, including those of race (e.g., Fryer & Levitt, 2004), geography, and/or class. A recent study within the publishing industry (Lulu.com, 2006), for example, shows a decreasing life expectancy of books on bestseller lists since the 1950s (equivalent to an increasing turnover rate), which, by comparison with the random-copying model, would suggest an increase in innovation, perhaps as books can be published more and more quickly in response to public topics and tastes.

As Baum, Richerson, Efferson, and Paciotti (2004, p. 306) point out, while there is a wealth of research on individual-level mechanisms of social learning, there is a need for more discussion by social scientists of how cultural traditions change over time at the population level. Applied here on the societal scale, the random-copying model simply assumes a constant proportion of innovators in the population, as often assumed in epidemiological models applied to binary-choice culture change (e.g., Dodds & Watts, 2005; Watts, 2002). Of course, innovation, that is, creativity, is an enormous topic (e.g., Martindale, 1975, 1986, 1999) that we are not attempting to explain at the individual, psychological level. As we move outward in scale, however, new quantitative effects emerge (Anderson, 1972), particularly with regard to the collective effects of society (e.g., Ball, 2004; Barabási, 2005; Le Bon, 1896). Baum et al., for example, conducted an empirical investigation of "cultural microevolution" on the scales as small as four individuals, which is an important bridge to the societal scale we are investigating here (we discuss their results below). At the societal scale, creative innovation generates what is effectively "quasi-random" variation as Martindale (1986) described it.

An alternative to the random-copying model is clearly some form of selection or that people choose the variant with the highest payoff with respect to some benefit (e.g., Baum et al., 2004; Boyd & Richerson, 1985; Gintis, in press; Henrich, 2001, 2006; Henrich et al., 2006). In the case of the phenomena we have investigated here—baby names, pop music, and dog breeds—we see no evidence for inherent benefits of one variant over another. As Baum et al. (2004) recently showed through experiments on groups of four participants, the stronger the payoff is for choosing one particular variant over another, the stronger is the "tradition" that evolves in bringing that choice to be the most popular, which is passed on to the newcomers of each generation (Fig. 5). Henrich (2001) and many others (e.g., Boyd & Richerson, 1985; Dunnell, 1978; Gintis, in press) have made similar predictions: when variants are not neutral, then we expect the most beneficial choice to rise to popularity and remain until a superior choice becomes available. Hence, we would not expect constant, population-independent turnover among cultural variants if they were being selected according to intrinsic value, although a future study of Top y charts of nonneutral cultural variants could reveal intriguing and unexpected aspects of their turnover.

Finally, it might be argued that none of the variants we have studied is truly neutral for various reasons—certain famous pop artists have clear advantages over newcomers, for example, or people tire of old fashions in favor of new ones (cf. pro-novelty bias in Boyd & Richerson, 1985). This is true, and it is the reason Herzog et al. (2004) could identify Dalmatians as being selected among dog breeds against a background of neutral evolution. To say that things evolve neutrally means that, of the variants observed, all behave in a neutral fashion. If they were all positively selected compared to some unseen variant, they would still all behave neutrally with respect to one another because fitnesses are always relative. Hence, the inevitable fact that newcomers have a lower fitness is not inconsistent with the model. It would be interesting in the future to investigate the effects of people getting tired of old fashions by introducing some sort of pro-novelty bias among individuals (cf. Shennan & Wilkinson, 2001) or by adding intrinsic value to variants to introduce elements of selection, which could even decay with time (cf. Dorogovtsev & Mamedes, 2000). At this stage, however, the aim of the proponents of the neutral model (Bentley et al., 2004; Bentley & Shennan, 2005; Hahn & Bentley, 2003; Herzog et al., 2004; Hubbell, 2001; Lipo et al., 1997; Neiman, 1995; Shennan & Wilkinson, 2001) is still to establish the random-copying model as an appropriate basis for making these added alterations. For all its simplicity, the neutral model replicates a remarkable range of patterns of cultural transmission. In this study, our main aim is to show that while added rules can always be imposed to engineer the results, the random-copying model produces constant turnover on its own, which we find unexpected, somewhat counterintuitive, and, therefore, significant.

In conclusion, our simulations of the random-copying model indicate that the population size N should not significantly affect the turnover rate on the pop charts. The time-averaged turnover rate in Top y charts is linearly proportional to the chart size y and the square root of the innovation fraction μ. Hence, while prediction of the next big popular success may be impossible (Salganik et al., 2006), predicting the frequency distribution (Bentley et al., 2004; Hahn & Bentley, 2003) and turnover rate is relatively straightforward. Since such regular turnover is not necessarily expected when independent, rational decisions are made, random copying may be identifiable in research of markets and cultural change. The neutral model could also be useful for assessing situations where copying and continual, yet directionless, turnover may be undesirable, as in politics or academic publishing. For these reasons and more, further research on random copying should be of high priority for the study of culture evolution and collective behavior.