A formal analysis of cultural evolution by replacement

Jing Xu (jing.xu@berkeley.edu)
Florencia Reali (floreali@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, 3210 Tolman Hall
Berkeley, CA 94720 USA

Abstract

Social psychologists have used the replacement method to simulate cultural evolution in the laboratory, studying what happens when a group of people performing a task are gradually replaced by new members. We provide a formal analysis of the dynamics of cultural evolution by replacement under the assumption that the people involved are Bayesian agents. We use a connection to a statistical inference algorithm – Gibbs sampling – to characterize the outcome of this process, and show that this analysis can account for both laboratory data and population-level trends resulting from cultural transmission.

Keywords: cultural evolution; cultural transmission; replacement; Gibbs sampling;

The science of cultural evolution studies the origin, persistence, and change of the body of information shared by the members of a society, such as knowledge, beliefs, and norms. Research in this field has traditionally been done by anthropologists and sociologists, although there is a long tradition in social psychology of reproducing processes of cultural evolution in the laboratory (Mesoudi, 2007). Recently, researchers have begun to apply methods developed in research on biological evolution to the study of cultural change (Mesoudi, Whiten, & Laland, 2006; Boyd & Richerson, 1985). In particular, mathematical models and computer simulations have proven to be powerful tools that can be used to uncover the dynamics of cultural evolution. For example, Bentley, Hahn, and Shennan (2004) argued that a random genetic drift model can be used to explain cultural phenomena in changes of baby names, prehistoric pottery styles and technology patents. However, there is still a significant gap between these formal models, which are applied at the level of whole populations, and the data that has been produced through the experimental investigation of cultural transmission. In this paper, we formally analyze one of the methods that has been used to simulate cultural transmission in the laboratory, and show that the resulting model can account for both experimental data and trends at the population level.

Social psychologists have studied cultural evolution using several strictly controlled laboratory paradigms, such as transmission chains, the replacement method, and the constant group method (Mesoudi, 2007). Transmission chains were originally used by Bartlett (1932) in his "serial reproduction" experiments. In these experiments, people were asked to reconstruct a stimulus from memory and the reconstructions produced by one person was used as the stimulus seen by the next. Over time, the original stimulus became distorted. Bartlett interpreted these results as indicating that people are biased by their pre-existing knowledge when they reconstruct information from memory, and that this bias became exaggerated through serial reproduction. Since Bartlett, other researchers have also used transmission chains to investigate how cultural biases affect the information being transmitted (see Mesoudi, 2007, for a review). Recently, Griffiths and Kalish (2007) formally analyzed this process under the assumption that the people in the chain are Bayesian agents. This analysis showed that the initial information dissipates over time, with the outcome of transmission being determined purely by the biases of the learners. Xu and Griffiths (2007) confirmed this prediction empirically for serial reproduction, using a strictly controlled set of transmission chains.

The existence of a formal analysis of transmission chains raises the possibility that similar analyses can be performed for the other methods used to simulate cultural evolution in the laboratory. Our focus here is on the replacement method, in which a group of people perform a task together, and the members of the group are replaced one at a time (Gerard, Kluckhohn, & Rapoport, 1956). The replacement method provides a complement to transmission chains, emphasizing transmission of information within a group rather than across generations. The method has been used to simulate patterns of persistence and change in group-based decision making, problem solving, and social organization (Mesoudi, 2007). By assuming that the people in the group are Bayesian agents, we are able to characterize the asymptotic consequences and dynamics of cultural evolution by replacement. As with transmission chains (Griffiths & Kalish, 2007), these results exploit a connection between cultural transmission and a statistical inference algorithm known as Gibbs sampling. We use this analysis to make quantitative predictions for laboratory data, capturing the way that confederates introduced to a group influence the naïve participants (Jacobs & Campbell, 1961), and for trends produced by cultural transmission in entire populations (Bentley et al., 2004).

The plan of the paper is as follows. The next section presents our formal analysis of the cultural evolution by replacement in its full generality, outlining the connection between this process and Gibbs sampling. We then give a detailed analysis of the predictions of the model using two case studies, starting with laboratory data and then turning to population-level trends. We close by discussing the implications of our approach and future directions.

Analyzing the replacement method

The basic structure of the replacement method is shown in Figure 1. Each generation is composed of a fixed number of
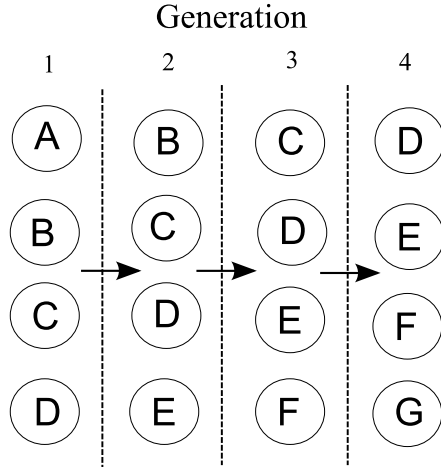
Generation

Figure 1: The replacement method. Each circle represents a single person, with identity expressed by a letter. In each generation, one member of the group is replaced. After four generations, no original members of the group remain.

individuals (in this case, four), and each individual performs a certain task, either together with the other group members or just in awareness of the responses of the rest of the group. After one iteration through the entire group, one member is replaced by a newcomer who has the opportunity to observe the actions of the other members of the group. Each replacement represents one generation of cultural transmission, as the newcomer is now exposed to the norms established by the group. Using this method, researchers can observe how much and to what degree old information is preserved and new information is introduced in the group across generations.

While a variety of tasks have been studied using the replacement method, we will focus on tasks that can be formulated as an informed choice or inductive inference. In such a task, a person receives some data x and has to infer the process z responsible for producing those data. For example, in one of the cases considered below, x is an observation of a visual stimulus, and z is the inferred extent of movement of that stimulus. In a group of n people, person i not only receives the data x_i, but also has the opportunity to observe the responses of the other people in the group, $\mathbf{z}_{-i} = (z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n)$, using this information to guide their response z_i.

The optimal solution to a problem of this kind is provided by Bayesian inference. A single person acting alone could express her expectations about the value of z before seeing x through a *prior* distribution $p(z)$, and her beliefs about the probability of seeing a particular value of x for each value of z through the *likelihood* $p(x|z)$. Her expectations about z after seeing x are expressed in the *posterior* distribution $p(z|x)$, obtained via Bayes' rule

$$p(z|x) = \frac{p(x|z)p(z)}{\int p(x|z)p(z)\,dz} \quad (1)$$

where the denominator simply ensures that the posterior probabilities sum to one. In a group setting, the expectations of each person before seeing data will be influenced by the responses of the other people, with the prior depended on these responses, $p(z_i|\mathbf{z}_{-i})$. Bayes' rule can still be applied in this case, with the posterior distribution being

$$p(z_i|x_i, \mathbf{z}_{-i}) = \frac{p(x_i|z_i)p(z_i|\mathbf{z}_{-i})}{\int p(x_i|z_i)p(z_i|\mathbf{z}_{-i})\,dz_i} \quad (2)$$

which is just Equation 1 with the appropriate prior.

With this formulation of the task, we can now analyze the replacement method. We can describe the state of a group with n members as a collection of n variables, $\mathbf{z} = (z_1, \ldots z_n)$, reflecting the responses of the people in the group. The process illustrated in Figure 1 can be viewed as successive resampling of the values of these variables as the person represented by each variable is replaced. Assuming that each person selects a response on the task by sampling from the posterior distribution conditioned on the responses of the other group members (Equation 2), the replacement method simply cycles through the following procedure:

sample $z_1^{(t+1)}$ from $p(z_1|x_1^{(t+1)}, z_2^{(t)}, \ldots, z_n^{(t)})$
sample $z_2^{(t+1)}$ from $p(z_2|x_2^{(t+1)}, z_1^{(t)}, z_3^{(t)}, \ldots, z_n^{(t)})$
$$\ldots$$
sample $z_n^{(t+1)}$ from $p(z_n|x_n^{(t+1)}, z_1^{(t)}, \ldots, z_{n-1}^{(t)})$

where the superscript (t) indexes the generation, and a new value for a variable, $z_i^{(t+1)}$, is used in sampling the other variables immediately after being generated.

The procedure described in the previous paragraph is likely to be familiar to readers familiar with Monte Carlo algorithms. In statistics, a method known as Gibbs sampling (Geman & Geman, 1984) is a popular tool for generating samples from the joint distribution of several random variables. Each step of Gibbs sampling replaces the value of one of the variables by drawing a sample from the distribution of this variable conditioned on all the remaining variables. For example, if we have variables $\mathbf{z} = (z_1, \ldots, z_n)$, the joint distribution is $p(\mathbf{z}) = p(z_1, z_2, \ldots, z_n)$. In each iteration, the algorithm cycles through the variables and draws a new value for each variable z_i from the conditional distribution $p(z_i|\mathbf{z}_{-i})$. This process constructs a Markov chain that converges to the target distribution $p(\mathbf{z})$. By iterating the process, the values of $\mathbf{z}$ ultimately behave as if they were sampled from $p(\mathbf{z})$.

Under the analysis given above, the replacement method implements a Gibbs sampling algorithm for the distribution

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{z})\prod_i p(x_i|z_i)}{\int p(\mathbf{z})\prod_i p(x_i|z_i)\,d\mathbf{z}} \quad (3)$$

where $p(\mathbf{z})$ is a joint distribution on $\mathbf{z}$ consistent with the set of conditional distributions $p(z_i|\mathbf{z}_{-i})$, and we assume that the set of observations $\mathbf{x} = (x_1, \ldots, x_n)$ is constant over time (ie. everybody receives the same observations). More precisely, if there exists a joint distribution $p(\mathbf{z})$ such that $p(z_i|x_i, \mathbf{z}_{-i}) \propto$

$p(x_i|z_i)p(\mathbf{z})$ for all i, then the replacement method defines a Markov chain on $\mathbf{z}$ with stationary distribution $p(\mathbf{z}|\mathbf{x})$ as defined in Equation 3. Cultural evolution by replacement will result in convergence to an equilibrium characterized by this distribution regardless of the initial value of $\mathbf{z}$.

The requirement that the distributions $p(z_i|\mathbf{z}_{-i})$ be consistent can be satisfied by assuming that each person estimates an appropriate prior from the responses of the the members of the group. If each person has a prior $p(z|\theta)$ with parameters θ, then learner i can estimate θ from $\mathbf{z}_{-i}$ by using Bayes' rule

$$p(\theta|\mathbf{z}_{-i}) = \frac{p(\theta)\prod_{j\neq i}p(z_j|\theta)}{\int p(\theta)\prod_{j\neq i}p(z_j|\theta)\,d\theta} \tag{4}$$

and then integrate over θ to obtain the prior

$$p(z_i|\mathbf{z}_{-i}) = \int p(z_i|\theta)p(\theta|\mathbf{z}_{-i})\,d\theta \tag{5}$$

which can be used as in Equation 2. The resulting set of distributions $p(z_i|\mathbf{z}_{-i})$ are consistent, with corresponding joint distribution $p(\mathbf{z}) = \int p(\theta)\prod_i p(z_i|\theta)\,d\theta$.

These results give us a simple characterization of the asymptotic consequences of cultural evolution by replacement. First, the population will converge to the same equilibrium, the distribution $p(\mathbf{z}|\mathbf{x})$, regardless of the initial value of $\mathbf{z}$. Second, we can say how the performance of the group will relate to the performance of individuals at this equilibrium. When $\mathbf{x}$ consists of a set of observations and $p(\mathbf{z})$ is the result of a parameterized prior as outlined above, the conclusions reached by a group will be more accurate than those reached by an individual who sees only one observation. Specifically, the marginal distribution on z_i at equilibrium is $p(z_i|\mathbf{x}) = \sum_{\mathbf{z}_{-i}} p(\mathbf{z}|\mathbf{x})$, as compared to $p(z_i|x_i)$ for an individual acting alone. This means that person i has the opportunity to benefit from the observations of all of the other people in the group, although the extent to which these observations increase accuracy will depend on how $p(\mathbf{z})$ is defined. Similar results also hold for variants of the replacement method: using different schemes for determining the people to be replaced or selecting them at random result in the same stationary distribution (e.g., Liu, Wong, & Kong, 1995).

An important special case of our analysis is that where there are no observations $\mathbf{x}$. In this case, each person simply responds based on the responses of the others, sampling from $p(z_i|\mathbf{z}_{-i})$. The stationary distribution of this process is $p(\mathbf{z})$, and in this case the group ultimately provides no advantage to the individual: the marginal distribution for each individual at equilibrium is the same regardless of the size of the group. However, the responses of individuals will be dependent, meaning that the population as a whole can manifest trends that are determined by the joint distribution $p(\mathbf{z})$.

In the remainder of the paper, we show that this analysis provides not just general results concerning the consequences of cultural evolution by replacement, but also quantitative predictions of specific phenomena. We present two case studies that illustrate how the model can be used to account for both laboratory data and population-level trends.

The first is a study in which group influence on people's judgments was investigated using a perceptual illusion (Jacobs & Campbell, 1961). The results showed that biases were easily introduced to the group by confederates, but would also disappear after several generations as those confederates were replaced by new naïve subjects. The second case study shows that the model can produce power-law distributions in choice frequencies, using the example of baby names in the United States (Bentley et al., 2004; Hahn & Bentley, 2003).

Predicting laboratory data

One of the classic social psychology studies on group influence was Sherif's (1936) experiment, in which participants were asked to estimate how far a lighted dot moved in the darkness when the dot actually did not move at all. The "auto-kinetic effect" refers to the reliable visual illusion that the dot was moving. Sherif's classic finding was that when group members who were confederates gave exaggerated estimates of the amount of movement, naïve participants were strongly influenced and reported similar estimates. Jacobs and Campbell (1961) studied the persistence of this bias using the replacement method. The groups initially consisted of all naïve participants or a number of confederates plus one naïve participant. After each generation the "oldest" group member was replaced by a new naïve participant. The biases introduced by the confederates decreased with each generation and entirely disappeared after six generations had passed since the removal of the last confederate. Figure 2 (a-c) shows the data for four-person groups with three initial confederates, and three-person groups with two and zero confederates.

Gibbs sampling with Gaussian distributions

We model the auto-kinetic effect experiment as a special case of the framework introduced above, where the distributions involved are Gaussian. Using Gaussian distributions simplifies the model and also allows us to analytically compute the probabilities required to simulate the process of replacement. In this task, x is the observed apparent motion of the stimulus, and z is the estimate of the amount that it moves. We assume that each participant has a prior distribution $p(z|\theta)$ that is Gaussian with mean μ and variance σ_z^2, where $\theta = \{\mu, \sigma_z\}$ are the parameters of the prior, and that the likelihood $p(x|z)$ is also Gaussian, with mean z and variance σ_x^2. Intuitively, μ and σ_z express each person's expectations about the amount of motion they are likely to see for this stimulus, while σ_x reflects how much noise is assumed to be in the perceptual observations. By inferring μ and σ_z from the responses of the other people in the group, people can hope to form a better estimate of the amount of motion, but they also become susceptible to biases introduced by confederates.

Following the framework outlined above, evaluating $p(z_i|\mathbf{z}_{-i})$ requires computing the posterior distribution on μ and σ_z^2 given $\mathbf{z}_{-i}$, which in turn requires specifying a prior on μ and σ_z^2. We used an uninformative prior, making minimal *a priori* commitments to the values of μ and σ_z^2. Under this prior, $p(z_i|\mathbf{z}_{-i})$ is a Student t distribution, with a mean

Human Data
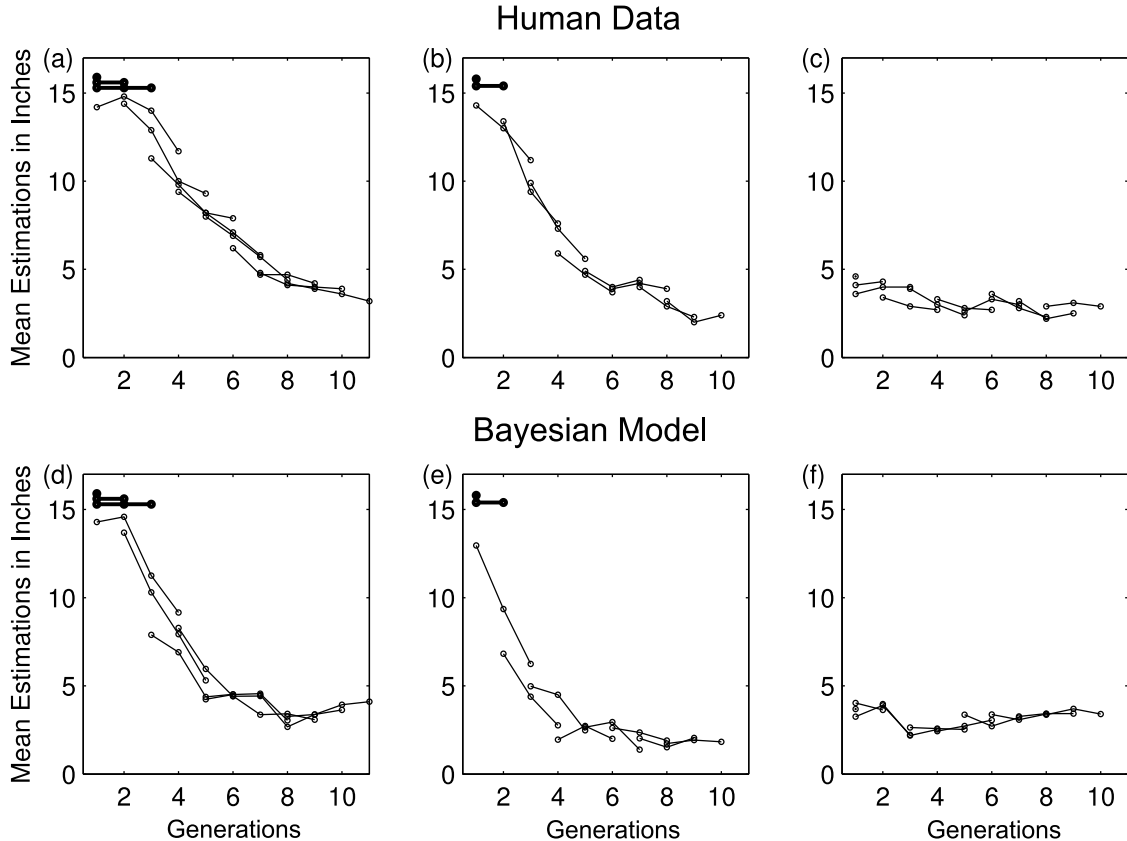

Bayesian Model

Figure 2: (a-c) Data from Jacobs and Campbell (1961). The heavy lines in the top left corners of the figures indicate the biased responses of confederates. The lighter lines are the responses of naïve participants, each of whom stayed in the group for four generations. (a) The experimental condition of 4-person groups with 3 initial confederates. (b) The experimental condition of 3-person groups with 2 initial confederates. (c) The control condition with 3-person groups with no confederates. (d-f) Model simulations for each condition. The correlation between data and simulation is 0.89.

that is the average of $\mathbf{z}_{-i}$, scale parameter $(1 + 1/n)^{1/2}$ times the standard deviation of $\mathbf{z}_{-i}$, and $(n-1)$ degrees of freedom (Gelman, Carlin, Stern, & Rubin, 1995). The resulting posterior required for Gibbs sampling, $p(z_i|x_i, \mathbf{z}_{-i})$ has no simple form, so we used an importance sampler to generate from this distribution (see, e.g. Neal, 1993). Specifically, we drew 1000 samples from the prior $p(z_i|\mathbf{z}_{-i})$ and weighted these samples by the Gaussian likelihood $p(x_i|z_i)$ to construct an approximation to the posterior $p(z_i|x_i, \mathbf{z}_{-i})$.

Simulating the replacement method We simulated the study conducted by Jacobs and Campbell (1961) using the Gibbs sampling model presented in the previous section. We assumed that each participant forms a single estimate of the amount of movement at each generation, sampling this value from their posterior distribution, and that all participants receive the same perceptual observation x. The value of x was set to 2.9 inches, the ultimate estimate arrived at by individuals exposed to the autokinetic effect. The only free parameter in the model was σ_x, which was set at 4.0 inches for all three conditions. Figure 2 (d-f) shows the results of the simulations for these three conditions. Each point is the mean of ten

replications for each condition.

These simulations show that the Markov chain converges faster with smaller group sizes and fewer confederates. When more naïve participants with their unbiased priors and new observations of the state of the world are introduced in the group over several iterations, the Markov chain decreases the initial influence of the confederates. These predictions are consistent with the findings of Jacobs and Campbell (1961), capturing both the qualitative decrease in biased responding and the quantitative rate at which this decrease occurs, with the correlation between simulation and empirical data being 0.89.

Predicting population-level trends

While the replacement method has largely been used to study cultural evolution in the laboratory, the underlying process provides a plausible model of cultural transmission in large populations. The idea that each person in the population is eventually replaced by a new person provides a useful simplification of the true process of cultural evolution. Indeed, exactly the same assumption is made in the Moran model, which is widely used to simulate biological evolution in fi-

Table 1: Power-law distributions of baby names through the 20th century. Exponents γ for power-law distributions were fit to log-log plots of probability vs. frequency for the top 1000 baby names (from Hahn & Bentley, 2003).

Years	γ (males)	γ (females)
1900-1909	1.79	1.88
1910-1919	1.79	1.86
1920-1929	1.69	1.86
1930-1939	1.77	1.86
1940-1949	1.64	1.86
1950-1959	1.63	1.71
1960-1969	1.64	1.88
1970-1979	1.70	1.93
1980-1989	1.71	1.84
1990-1999	1.61	1.75

nite populations (see Ewens, 2004). In this section we show that this kind of model can produce power-law distributions in the frequencies with which people make particular choices, a phenomenon that has been used to support the application of models of biological evolution to cultural evolution (Bentley et al., 2004). While this phenomenon has been noted in several datasets, we will focus on the case of baby names.

The name given to newborn babies is an example of a cultural trait that changes over time as a result of cultural transmission. Most parents pick a pre-existing name for their child, although some parents choose to invent a new one. Over generations, some names become more frequent and others decline in prevalence. Assuming names to be equivalent in desirability, Hahn and Bentley (2003) argued that the process by which names are transmitted is analogous to the biological mechanism of random drift. The basis for this argument was the observation that the frequencies with which names appear in the population follow a power-law distribution: the probability that a name appears with frequency k is roughly $p(k) \propto k^{-\gamma}$ for some γ, meaning that most names are infrequent but a few names have very high frequencies. They reported that for each decade of the twentieth century, the frequency distributions of baby names used in the United States obey power-law distributions with slopes that are consistent throughout the century (Table 1). Hahn and Bentley then showed that a model of genetic drift in a finite population could produce similar trends, provided the frequencies of names were accumulated over a number of generations.

Producing power-law distributions via replacement

The replacement method can produce power-law frequency distributions with exponents matching those found in human populations, without requiring accumulation of frequencies over multiple generations or any other additional mechanisms. Stochastic processes based on the principle of preferential attachment have been shown to produce outcome frequencies that obey power-law distributions (Mitzenmacher,

2003). The principle of preferential attachment is based on the assumption that the probability that a particular outcome takes on a particular value depends upon the frequency of that value in the choices of other people. If values that have been chosen by more people are more probable, then the high frequencies associated with a power-law distribution can be produced through the resulting "rich-get-richer" process.

We can define a simple rich-get-richer process that seems appropriate for selecting baby names. Let **z** be the set of names of a collection of n people. Assume that this set of names changes over time through the replacement of the oldest person with a new baby, and the name z_i for this new baby is chosen based on the names $\mathbf{z}_{-i}$ of all other members of the population. We can make popular names more probable, but still allow for the possibility of choosing a completely new name, by sampling a name according to the distribution

$$p(z_i = j|\mathbf{z}_{-i}) = \begin{cases} \frac{n_j^{\mathbf{z}_{-i}} - a}{n-1+b} & \text{if } n_j^{\mathbf{z}_{-i}} > 0 \\ \frac{K(\mathbf{z}_{-i})a+b}{n-1+b} & \text{otherwise} \end{cases} \quad (6)$$

where a and b are parameters of the process, $n_j^{\mathbf{z}_{-i}}$ is the number of babies with name j in $\mathbf{z}_{-i}$, and $K(\mathbf{z}_{-i})$ is the total number of name types. The parameters are constrained such that $0 \leq a < 1$ and $b \geq 0$, and varying these parameters varies the extent to which parents are likely to choose a new name.

While there are presumably many ways to define $p(z_i|\mathbf{z}_{-i})$ that might seem reasonable, using Equation 6 makes the outcome of cultural evolution by replacement particularly simple to analyze. This is the conditional distribution associated with a particular stochastic process known as the Pitman-Yor process (Pitman & Yor, 1997), and converges to

$$P(\mathbf{z}) = a^{K(\mathbf{z})} \frac{\Gamma(K(\mathbf{z}) + \frac{b}{a})}{\Gamma(\frac{b}{a})} \frac{\Gamma(b)}{\Gamma(n+b)} \prod_{k=1}^{K(\mathbf{z})} \frac{\Gamma(n_k^{\mathbf{z}} - a)}{\Gamma(1-a)} \quad (7)$$

where $\Gamma(\cdot)$ is the generalized factorial function (Boas, 1983). The induced distribution over the frequencies of different names follows a power-law distribution with slope value $\gamma = 1 + a$, provided n is large (Pitman, 2006). Thus, it is possible for cultural evolution by replacement to produce power-law frequency distributions that exhibit the same properties as human choice data, such as choices of baby names.

Simulating cultural evolution We implemented the procedure outlined above to examine the properties of the stationary distribution with large but finite n. The simulation was initialized by arbitrarily assigning numerical names to a population of $n = 10,000$ individuals. Each individual in the starting population had a unique name. At each step of the process, a new individual was assigned a name according to the probability $p(z_i = j|\mathbf{z}_{i-1})$ defined in Equation 6. We let the simulation run for 500,000 iterations. The parameters a and b were set as 0.7 and 4 respectively.

Figure 3 is a log-log plot that shows the power-law distribution generated by Gibbs sampling, showing the frequencies
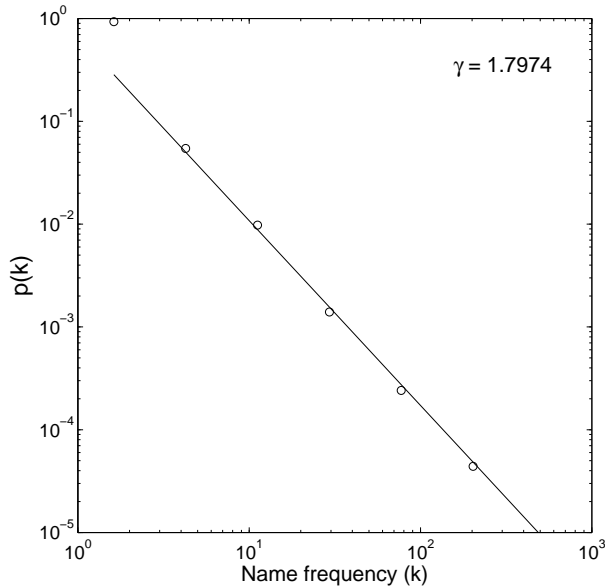
Figure 3: Power-law distribution generated by replacement. The horizontal axis represents the number of names of type k in the total sample of individuals, and the vertical axis represents the probability $P(k)$ that a certain name would fall within the bin at that frequency level. The exponent value that best fit the plot was 1.79, consistent with the slopes reported in Hahn and Bentley (2003).

of names in the final 10,000 samples. A power-law distribution will appear as a straight line in a log-log plot, since $\log p(k) = -\gamma \log k + C$. As can be seen from the figure, the distribution over names converges to a power-law distribution with exponent $\gamma = 1.79$, which is consistent with the values found in baby name data as reported by Hahn and Bentley (see Table 1). The results demonstrate that cultural evolution by replacement provides an alternative to the biological models used by Hahn and Bentley (2003) to explain the dynamics of cultural transmission of baby names. The model may provide a viable sampling scheme to explain the way individuals choose a name for their child. Moreover, the dynamics captured by the model may not be restricted to the case of baby names. Rather, the replacement method may help explain the emergence of power-law distributions in other cultural domains that have previously been modeled simply by applying ideas from biology (Bentley et al., 2004).

Conclusions

We have presented a formal analysis of the replacement method used to study cultural evolution study and have shown that this analysis can account for both empirical work, such as the auto-kinetic effect study, and real world data, such as changes in baby names. The replacement method is particularly useful in studying group dynamics across overlapping generations, and this analysis complements previous work examining the outcome of transmission chains. By connecting the process of cultural evolution to statistical inference algorithms such as Gibbs sampling, our results provide a link to a large literature in statistics and computer sciences that could provide further insight into the dynamics and consequences of the transmission of knowledge. Most significantly, they begin to lay the groundwork for an approach to cultural evolution that links laboratory data and population-level trends, and provides an alternative to simply applying models from the study of biological evolution to the literatures of social psychology and anthropology.

References

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

Bentley, R., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *27*, 1143-1150.

Boas, M. L. (1983). *Mathematical methods in the physical sciences* (2nd ed.). New York: Wiley.

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

Ewens, W. J. (2004). *Mathematical population genetics* (2nd ed.). New York: Springer-Verlag.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. New York: Chapman & Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

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

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*.

Hahn, M. W., & Bentley, R. (2003). Drift as a mechanism for cultural change: an example from baby names. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *270*, S120-S123.

Jacobs, R. C., & Campbell, D. T. (1961). The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. *Journal of Abnormal and Social Psychology*, *62*, 649-658.

Liu, J. S., Wong, W. H., & Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society B*, *57*, 157-169.

Mesoudi, A. (2007). Using the methods of experimental social psychology to study cultural evolution. *Journal of Social, Evolutionary, and Cultural Psychology*, *1*, 35-38.

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, *29*, 329-383.

Mitzenmacher, M. (2003). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, *1*(2), 226-251.

Neal, R. M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods* (Tech. Rep. No. CRG-TR-93-1). University of Toronto.

Pitman, J. (2006). *Combinatorial stochastic processes*. New York: Springer-Verlag.

Pitman, J., & Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, *25*, 855-900.

Sherif, M. (1936). *The psychology of social norms*. Oxford: Harper.

Xu, J., & Griffiths, T. L. (2007). *A Bayesian analysis of serial reproduction.* (Poster presented at the 29th Meeting of the Cognitive Science Society, Nashville, TN)