

Random drift and large shifts in popularity of dog breeds

Harold A. Herzog¹, R. Alexander Bentley^{2*} and Matthew W. Hahn³

¹Department of Psychology, Western Carolina University, Cullowhee, NC 28723, USA (herzog@email.wcu.edu)

²AHRB Centre for the Evolutionary Analysis of Cultural Behaviour, University College London, 31–34 Gordon Square, London WC1H 0PY, UK

³Center for Population Biology, University of California, Davis, CA 95616, USA (mwhahn@ucdavis.edu)

* Author for correspondence (r.bentley@ucl.ac.uk).

Recd 15.01.04; Acctpd 16.02.04; Published online

A simple model of random copying among individuals, similar to the population genetic model of random drift, can predict the variability in the popularity of cultural variants. Here, we show that random drift also explains a biologically relevant cultural phenomenon—changes in the distributions of popularity of dog breeds in the United States in each of the past 50 years. There are, however, interesting deviations from the model that involve large changes in the popularity of certain breeds. By identifying meaningful departures from our null model, we show how it can serve as a foundation for studying culture change quantitatively, using the tools of population genetics.

Keywords: cultural transmission; cultural evolution; random genetic drift; power laws

1. INTRODUCTION

A remarkably useful way to study cultural change is to assume that individuals, confronted with many different choices, simply copy other individuals rather than make ‘optimal’ or ‘rational’ decisions. In general, this process relates to a classic evolutionary phenomenon called random genetic drift, for which quantitative models originally developed in population genetics have proven useful in identifying mechanisms of cultural change (Cavalli-Sforza & Feldman 1981). In the neutral model of random drift, individuals randomly copy existing cultural variants from other individuals, with some occasionally inventing new variants (Kimura & Crow 1964; Hahn & Bentley 2003). This model predicts the frequency distributions of first names over the past century (Hahn & Bentley 2003), decorations on prehistoric pottery (Neiman 1995; Bentley & Shennan 2003) and journal citations of scientific authors (Simkin & Roychowdhury 2003).

We report that the neutral model of random drift explains changes in distributions of the popularity of purebred dogs in the United States in each of the past 50 years. By using random drift as the null hypothesis, we demonstrate how it can serve as a foundation for studying culture change quantitatively. Concerning culture change in general, it is meaningful to identify departures from the null model, which in this case involve large changes in the

popularity of certain breeds. Concerning the particular case of dogs in North America, our identification of random drift among dogs in this century elicits the interesting question of when drift began to predominate in their coevolution with human culture.

2. DATA COLLECTION

Founded in 1884, the American Kennel Club (AKC) maintains the world’s largest registry of purebred dogs (O’Neill 1985), with purebreds comprising *ca.* 50% of dogs in households in the United States (New *et al.* 2000). AKC registrations are filed for every generation with the registration of a litter. Under AKC Rules, any person who sells dogs represented as ‘AKC registerable’ must maintain records that make it possible to give full identifying information with every dog delivered.

Dogs registered with the AKC are identified with date of birth, name of breeder and date sold or delivered. In addition to the breed, each AKC registration certificate identifies the dog by names and numbers of sire (father) and dam (mother), litter number, date of birth, sex, colour and markings.

We obtained from the AKC the annual number of new puppy registrations for all recognized breeds between 1946 (107 breeds) and 2001 (150 breeds). Although not all purebred dogs are registered by their owners, and not all dogs are AKC purebreds, these data represent a large (total of $n = 42\,280\,840$) and highly accurate index of the relative popularity of purebred dog breeds in the United States over the past five decades.

In studying the transmission of culture, it is important that, over the period of study, the variants defined as the units of transmission be discrete and copied accurately from one individual to the next, such that mutations are obvious (Dawkins 1976; Lipo 2001; Gabora 2004). This is an issue in the study of the copying of songs among birds (Slater & Ince 1979; Lynch & Baker 1994), for example, and we believe that shifts in preferences for dog breeds fit these criteria.

3. RESULTS AND DISCUSSION

Figure 1*a* shows that the number of new registrations within each breed obeys a steeply descending distribution with a long extended ‘tail’, such that while most breeds include a small number of new purebred puppies registered each year, a few breeds have a very large number of new registrants. When the relative frequencies of dogs in each breed are plotted as a function of their expected number (the caption to figure 1*b* describes the plotting method) the data clearly form a power-law probability distribution, $E(\nu) \sim C/\nu^\alpha$, where C is a constant and α represents the slope on a log–log plot. In every year from 1946 to 2001, the power-law fit yields an r^2 value of 0.88 or better, with values of α always close to 1.0 (table 1). Figure 1*b* shows distributions from representative years, which extend over almost five orders of magnitude in terms of the frequencies of breeds. These power-law distributions are stable over the past 56 years even though individual breeds have changed in frequency, and have been introduced and lost from the population.

These distributions are well explained by the infinite allele model of population genetics with random drift (Kimura & Crow 1964). In this model, each new dog is either randomly chosen from an existing breed (sampling)

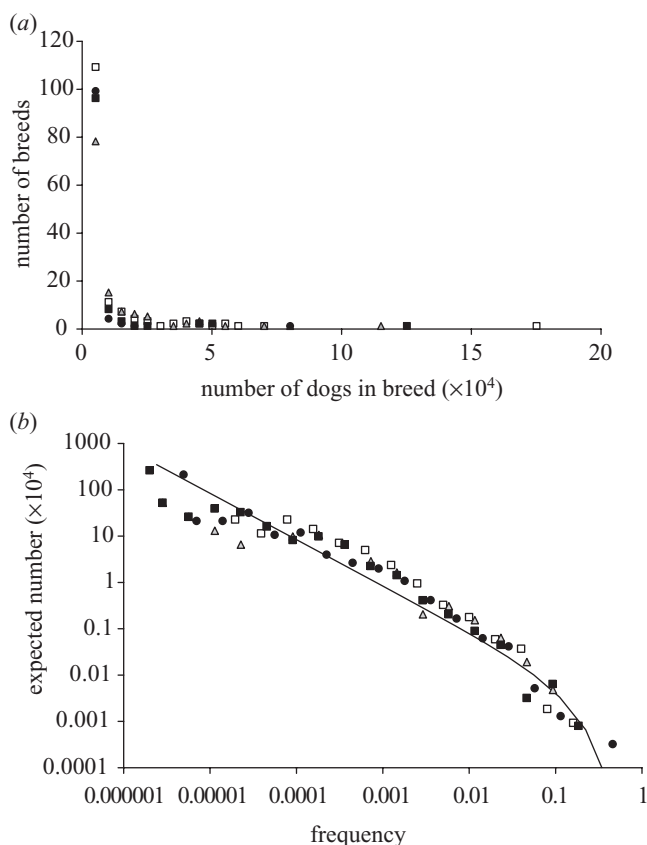


Figure 1. (a) Histogram of the numbers of dogs registered in each breed, for selected years. The bin size is 5000 dogs, and the many empty bins on the high end are not plotted. (b) Frequencies of dog breeds in the United States, for several selected years, on a log-log plot. For this plot, ‘frequency’ represents the number of dogs in a breed divided by the total number of all dogs registered in that year, and ‘expected number’ is the number of breeds falling within the histogram bin divided by the total number of registered breeds in that year. As usual for such log-log plots (11), one avoids empty bins by using bin sizes that increase in powers of 2 (0.0001–0.0002, 0.0002–0.0004, 0.0004–0.0008...), data are plotted at the middle of each bin, and probabilities are normalized for the increasing bin sizes. The line shows the prediction of the neutral-trait model using equation (3.1), with $\theta = 8$. Years: 2000 (open squares); 1977 (triangles); 1962 (filled squares); 1946 (filled circles).

or is chosen from a novel breed (mutation). The equilibrium frequency distribution of variants in a finite population is a balance between sampling and mutation (Kimura & Crow 1964):

$$\phi(\nu) = \theta\nu^{-1}(1 - \nu)^{\theta-1}. \tag{3.1}$$

Here, ν is the frequency of a given variant ($1/N \leq \nu \leq 1$, where N is the population size), $\phi(\nu)$ is the expected number of variants at frequency ν and θ is a measure of the variability in a population ($\theta = 2N\mu$ for the haploid case, where μ is the mutation rate). The ν^{-1} term means that $\phi(\nu)$ plots nearly as a power law with slope of -1 , which is similar to the slopes that we observe in the data (table 1).

The overall distribution of dog breeds can be described by a random drift process, using equation (3.1) with $\theta = 8$ (figure 1b). If we estimate the average N to be approximately one million dogs registered per year over the past 50 years (table 1), this value of θ implies that

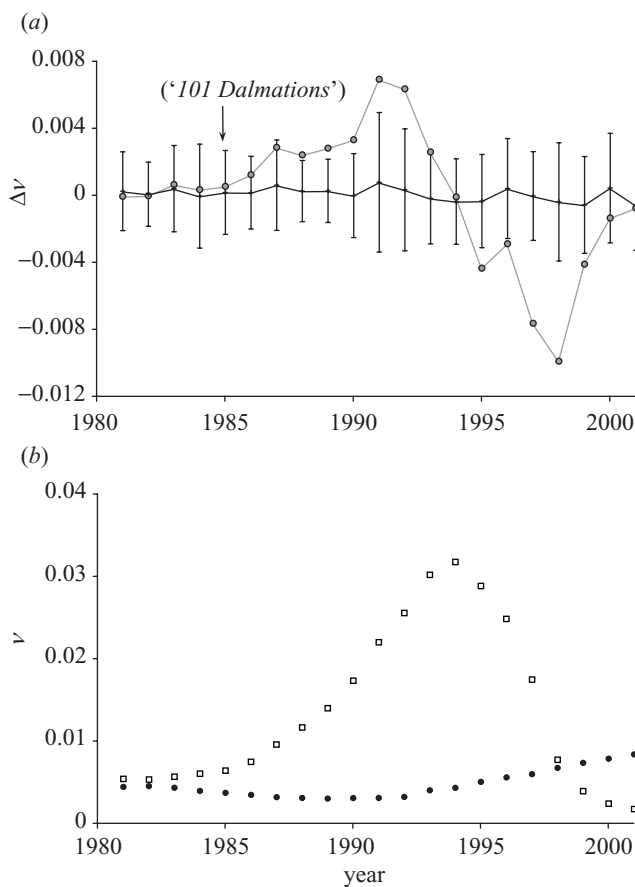


Figure 2. (a) Tracking $\Delta\nu$, the yearly change in the number of dogs in each breed divided by number of dogs in all breeds, for Dalmatians (grey circles) alone versus the average $\pm 2\sigma$ of all other breeds (crosses) that were present in 1980. (b) Yearly frequencies of Dalmatians (squares) and Weimaraners (circles), with the expected (very small) $\pm 2\sigma$ per generation as calculated by equation (3.2), where $\sigma = \sqrt{V}$.

$\mu = 0.000\ 004$, or one novel breed per quarter million registrations. In fact, our data show 43 new breeds among the 42 280 840 new registrations, or a mutation rate of 0.000 001.

This match between theory and data strongly suggests that the frequency of any particular dog breed occurs largely by random chance. This view suggests that some widely held beliefs about the causes of changes in a cultural variant such as breed popularity are incorrect. For example, it is commonly believed that taking the Best in Show Award at the annual Westminster Kennel Club Dog Show usually results in a substantial increase in the popularity of the winning breed. In reality, the rate of increase in the number of new registrations for Westminster Best in Show breeds over the 5 years following their win does not significantly differ from their rate of increase over the 5 years prior to winning (Herzog & Elias 2004).

In some cases, however, there are dramatic fluctuations in the popularities of individual breeds that do not fit this simple null hypothesis. For example, after the release of the 1985 version of the Disney movie *101 Dalmatians*, new Dalmatian registrations increased 6.2-fold, from 6880 registrations in 1985 to 42 816 registrations in 1993, followed by a precipitous fall to 4652 registrations 6 years later. Although frequency changes are expected with

Table 1. Yearly numbers of purebred dogs registered with the AKC, together with the exponent α and r^2 values for a power law fit to the distribution of breed frequencies.

year	N	α	r^2
1946	205 211	1.04	0.946
1947	233 661	1.03	0.936
1948	225 794	1.02	0.938
1949	239 397	1.02	0.938
1950	248 953	1.01	0.945
1951	261 614	1.01	0.951
1952	286 396	0.93	0.930
1953	310 617	0.99	0.968
1954	337 931	0.94	0.945
1955	351 512	0.97	0.929
1956	421 718	0.94	0.946
1957	427 698	0.94	0.953
1958	437 234	0.99	0.960
1959	453 738	0.93	0.946
1960	435 839	0.96	0.947
1961	485 445	0.94	0.945
1962	508 368	0.92	0.934
1963	559 227	0.88	0.936
1964	630 587	0.85	0.925
1965	712 616	0.91	0.942
1966	793 440	0.89	1.940
1967	874 869	0.86	0.943
1968	898 508	0.83	0.934
1969	961 975	0.82	0.927
1970	1 045 555	0.83	0.937
1971	1 119 212	0.85	0.958
1972	1 093 384	0.83	0.942
1973	1 092 577	0.87	0.963
1974	1 095 733	0.80	0.933
1975	1 017 806	0.79	0.891
1976	1 043 975	0.85	0.916
1977	1 009 396	0.74	0.900
1978	976 339	0.78	0.909
1979	961 488	0.77	0.882
1980	1004933	0.76	0.915
1981	1 030 017	0.75	0.902
1982	1 033 334	0.83	0.911
1983	1 081 369	0.93	0.943
1984	1 067 653	0.92	0.937
1985	1 088 655	0.87	0.899
1986	1 106 399	0.79	0.886
1987	1 187 400	0.86	0.936
1988	1 220 500	0.87	0.934
1989	1 257 700	0.86	0.936
1990	1 253 214	0.86	0.935
1991	1 379 544	0.83	0.900
1992	1 528 392	0.94	0.952
1993	1 422 559	0.81	0.905
1994	1 345 941	0.81	0.895
1995	1 277 039	0.87	0.928
1996	1 333 581	0.87	0.935
1997	1 307 362	0.97	0.959
1998	1 220 982	0.87	0.915
1999	1 119 700	0.89	0.931
2000	1 175 473	0.90	0.935
2001	1 081 335	1.01	0.966

random drift, by 1988 the yearly changes in Dalmatians exceed 95% of other breeds (figure 2).

In the neutral-trait model, the only source of variance in allele frequencies over time, V , is random sampling, as

$$V = \frac{\nu(1-\nu)}{2N}. \quad (3.2)$$

Figure 2 shows that after 1985, dalmatian frequencies changed much more each year than would be predicted by the random drift prediction of equation (3.2), whereas weimaraner frequencies, as representative of a drifting breed, changed more on a par with the predictions.

The predominant effects of drift in the popularity of modern American dog breeds suggest a recent shift between Darwinian and cultural selection mechanisms during the long interaction between humans and dogs. Evidence from mitochondrial DNA (mtDNA) indicates that early North American dogs evolved from wolves in east Asia *ca.* 15 000 years BP (Savolainen *et al.* 2002). Domestic dogs have the ability to respond to human behaviour more skilfully than wolves or even chimpanzees (Hare *et al.* 2002), which suggests that early dogs were genetically selected for this ability. However, mtDNA evidence also suggests that, once domesticated, the dog quickly spread into the New World as an element of culture, as part of the Late Pleistocene human colonization across the Bering Strait (Leonard *et al.* 2002). Our findings indicate that while the domestication of dogs was originally propelled by selective processes of evolution, human cultural change in the form of random drift now predominates.

However, while the neutral model is unarguably the simplest model that can explain the observed patterns of cultural variants, predictions made by models of natural selection can be indistinguishable from a neutral model (Gillespie 1977, 1991; Kimura 1983; Hubbell 2001; McGill 2003; Volkov *et al.* 2003). While we do not claim to resolve the debate over drift versus selection, we favour the neutral model because it is the simplest null model for quantitatively studying cultural change (Crow & Kimura 1970; Neiman 1995; Hartl & Clark 1997; Lipo *et al.* 1997; Lipo 2001). The frequency distribution of variants is a power law whose slope can be predicted from the neutral model if just two quantities are known: the effective population size and the mutation rate (Bentley *et al.* 2004). In previous studies (Bentley & Shennan 2003; Hahn & Bentley 2003; Bentley *et al.* 2004), we have found this prediction method to work well with baby names, pottery decorations and patents, giving further evidence that they change neutrally.

In summary, the stochastic, multiplicative nature of random drift explains why a few cultural variants can be expected to become highly popular owing to chance alone—not necessarily because they are somehow better or more pleasing to people. In certain instances, however, such as after highly successful films, the popularity of variants can change more dramatically through non-neutral processes if choices are no longer being made randomly. Testing against random drift as a null hypothesis can identify such cases of extraordinary rapid cultural change.

Acknowledgements

The authors thank Robiyn Mimms for help with data collection, the American Kennel Club for providing access to the data, and Carl Lipo for helpful comments on an earlier draft.

Bentley, R. A. & Shennan, S. J. 2003 Cultural transmission and stochastic network growth. *Am. Antiquity* **68**, 459–485.

Bentley, R. A., Hahn, M. W. & Shennan, S. J. 2004 Random drift and culture change. *Proc. R. Soc. Lond. B* **271**. (In the press.) (DOI 10.1098/rspb.2004.2746.)

- Cavalli-Sforza, L. L. & Feldman, M. W. 1981 *Cultural transmission and evolution*. Princeton University Press.
- Crow, J. F. & Kimura, M. 1970 *An introduction to population genetics theory*. New York: Harper & Row.
- Dawkins, R. 1976 *The selfish gene*. Oxford University Press.
- Gabora, L. 2004 Ideas are not replicators but minds are. *Biol. Phil.* **19**, 127–143.
- Gillespie, J. H. 1977 Sampling theory for alleles in a random environment. *Nature* **266**, 443–445.
- Gillespie, J. H. 1991 *The causes of molecular evolution*. Oxford University Press.
- Hahn, M. W. & Bentley, R. A. 2003 Drift as a mechanism for cultural change: an example from baby names. *Proc. R. Soc. Lond. B* **270**(Suppl. 1), S1–S4. (DOI:10.1098/rsbl.2003.0045.)
- Hare, B., Brown, M., Williamson, C. & Tomasello, M. 2002 The domestication of social cognition in dogs. *Science* **298**, 1634–1636.
- Hartl, D. L. & Clark, A. G. 1997 *Principles of population genetics*. Sunderland, MA: Sinauer.
- Herzog, H. A. & Elias, S. M. 2004 The effects of winning Westminster on dog breed popularity. *J. Am. Vet. Med. Assoc.* (In the press.)
- Hubbell, S. P. 2001 *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
- Kimura, M. 1983 *The neutral theory of molecular evolution*. Cambridge University Press.
- Kimura, M. & Crow, J. F. 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.
- Leonard, J. A., Wayne, R. K., Wheeler, J., Valadez, R., Guillén, S. & Vilà, C. 2002 Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**, 1613–1617.
- Lipo, C.P. 2001 *Science, style and the study of community structure: an example from the central Mississippi river valley*. British Archaeological Reports International Series 918, Oxford.
- Lipo, C. P., Madsen, M. E., Dunnell, R. C. & Hunt, T. 1997 Population structure, cultural transmission and frequency seriation. *J. Anthropol. Archaeol.* **16**, 301–333.
- Lynch, A. & Baker, A. J. 1994 A population memetics approach to cultural evolution in chaffinch song: differentiation among populations. *Evolution* **48**, 351–359.
- McGill, B. J. 2003 A test of the unified neutral theory of biodiversity. *Nature* **422**, 881–885.
- Neiman, F. D. 1995 Stylistic variation in evolutionary perspective. *Am. Antiquity* **60**, 7–36.
- New, J. C., Salman, M. D., King, M., Scarlett, J. M., Kass, P. H. & Hutchinson, J. M. 2000 Characteristics of shelter-relinquished animals and their owners compared with animals and their owners in U.S. pet-owning households. *J. Appl. Anim. Welfare Sci.* **3**, 179–201.
- O'Neill, C. A. T. 1985 *The American Kennel Club 1884–1984: a source book*. New York: Howell Book House.
- Savolainen, P., Zhang, Y., Luo, J., Lundeberg, J. & Leitner, T. 2002 Genetic evidence for an east Asian origin of domestic dogs. *Science* **298**, 1610–1613.
- Simkin, M.V. & Roychowdhury, V.P. 2003 Copied citations create renowned papers? arXiv.org e-Print archive: cond-mat/0305150.
- Slater, P. J. B. & Ince, S. A. 1979 Cultural evolution in chaffinch song. *Behaviour* **71**, 146–166.
- Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. 2003 Neutral theory and the relative species abundance in ecology. *Nature* **424**, 1035–1037.