

*Comment*

## Population-level neutral model already explains linguistic patterns

In customizing the neutral model for language transmission, Reali & Griffiths [1] have added a new, ‘Bayesian’ learning interpretation to a neutral model that has been used in cultural evolution studies for some time (e.g. [2–6]). While Reali & Griffiths [1, p. 435] dismiss previous applications of this neutral model as being used merely ‘as a metaphor’, it has been explored in quantitative detail, and applied even to word frequencies [7]. We therefore question whether this fairly complex description of Bayesian learning is necessary, as it does not change the results of the model, and runs the risk of obscuring the advances made both through careful modifications and wider applications of this powerful model.

The Reali & Griffiths [1] version of the model is mathematically equivalent to previous versions, in representing a set of  $N$  variants per time step, with the next generation of  $N$  variants constructed by repeatedly sampling variants from the previous time step, with some probability of ‘mutation’, i.e. introduction of a variant of a unique new variant. Reali & Griffiths [1] apply the model in a novel sense to individual cognition, whereas it has previously been considered in terms of a population of fixed size  $N$ , which is replaced by  $N$  new agents in each time step. In this population version, each new individual either copies an existing variant from the last time step (with probability proportionate to the previous choices of agents), or chooses a new variant. With probability  $1 - \mu$ , an incoming agent copies its variant from that of an agent within the previous time step, or else with probability  $\mu$ , the agent innovates by choosing a unique new variant at random.

This model has already been shown to yield the inverse power law in the probability of variant frequencies that Reali and Griffiths report, from which the other inverse power law, in word frequency versus replacement rate, also follows. We demonstrate this briefly below, in order to show how the discussion of the model has advanced beyond this, including the incremental addition of a new extra parameter, ‘memory’ [8].

Before we consider the model predictions, however, we address the question of the psychological plausibility of the assumptions that are made about agent behaviour. The Bayesian learning approach assumes agents use ‘a rational procedure for belief updating that explicitly represents the expectations of learners’ (Reali & Griffiths [1, p. 430]). This raises a very important topic in social sciences. Standard economic theory, for example, assumes rational agents capable of obtaining all relevant

information and then processing this to arrive at an optimal decision. The less demanding postulate of bounded rationality restricts access to full information, but the hypothesis of optimal decisions is retained (subject to information constraints).

Individual rationality is not the only way, however, that learning can be represented (e.g. [9,10]), and alternative hypotheses have support both in particular applications and in the wider literature. At the other extreme, for instance, the assumption of literally ‘zero-intelligence’ agents, based on the particle model of physics, has provided a powerful explanation of many population-scale phenomena, such as financial asset price changes [11]. Daniel Kahneman [12] argued in his Nobel prize lecture, awarded for his work in economic psychology, that ‘the central characteristic of agents is not that they reason poorly, but that they often act intuitively. And the behaviour of these agents is not guided by what they are able to compute, but by what they happen to see at a given moment’. Heuristically, this seems perfectly consistent with a population-level view of learning, rather than the standard rational-individual model.

We might, therefore, alternatively assume that agents (including language speakers) observe the relative popularity of previous choices among other agents, and copy in proportion to these popularities, with a small probability of making an innovative choice. This population view, which is able to offer a good account of a wide range of cultural phenomena, assumes that agents are indifferent—‘neutral’—to the particular agent that they copy. The result of this neutrality is that agents tend to copy variants in proportion to the previous choices made by other agents; a more popular choice is more likely to be copied than a less popular one. In the special case where  $\mu = 0$ , the model is formally equivalent to a process of preferential attachment (e.g. [13]).

This population view also can provide extra flexibility for modifying the model. In the Bayesian learning representation, for instance, each individual commits previous experience to individual memory. In the population view, previous experience is copied directly from other individuals, and ‘memory’ becomes a useful added parameter. In our own version [8], this variable memory, designated as  $m$ , allows agents to copy from  $m$  previous time steps. This effectively means that the original variants can ‘linger’, because even if replaced in the current time step, they might still be ‘remembered’ again within  $m$  subsequent time steps.

Adding this memory parameter adds both flexibility and power to replicate real-world data patterns, including the three that Reali & Griffiths [1] report; in the inverse power law probability distribution of variant frequencies,

and in ‘S’ curves of replacement. Furthermore, by keeping the model simple in description, it is then easier to explore different phenomena, such as the flux of variants on a ranked popularity list [8,14]. We can visit each of the three main results of Reali & Griffiths [1] to show the advantages.

Firstly, Reali & Griffiths [1] fit their results to a power law probability distribution with exponent  $-1.7$ . Among the many processes that can produce Zipf’s laws [15], the simpler neutral model yields this very same power law exponent (between about  $-1.6$  and  $-1.8$ ) when variants are counted cumulatively (e.g. [16]). Notably, the exponent of best fit varies with the choice of mutation rate, and in fact as we move away from  $N\mu = 1$  ( $N$  is population,  $\mu$  is mutation), a power law no longer fits [17]; increasing  $N\mu$  tends to reduce highest frequencies and push the distribution towards an exponential distribution, whereas decreasing  $N\mu$  increases the highest frequency, ultimately to a ‘winner-take-all’ distribution at  $N\mu = 0$  [3,8,18]. In addition, the added memory parameter  $m$  enables the modified neutral model to generate a much larger family of right-skew frequency distributions [8].

Secondly, Reali & Griffiths [1] report an inverse power law relating word frequency to replacement rate, comparable to real languages [19]. This has also already been shown in the population version of the model [14], albeit in a slightly different way. Consider a ranked list of the  $V$  most popularly used variants in a particular variant pool, such as the  $V$  most popularly used words in a language. Regarding this list of variants, ranked in order of popularity from 1 (most popular) to  $V$  (least popular), the turnover on the entire list (word replacement rate among the top  $V$  words) is simply proportional to the size,  $V$ , of the list—as has been shown through simulation [14,20] and analytically [21]. The turnover just at rank  $V$  is equivalent to the turnover on the entire list from rank 1 to rank  $V$ , because each new entry must, at some instant, displace the one at the bottom or  $V$ th position (this turnover decreases moving up the rankings, because reaching rank  $V$  does not guarantee reaching rank  $V - 1$ ). Zipf’s law implies that the frequency of a word at rank  $V$  has usage frequency proportional to  $V^{-a}$ , meaning the probability distribution function (PDF) has an exponent of  $-(1 + 1/a)$ . Plotting turnover at position  $V$  on the  $y$ -axis versus the word frequency  $V^{-a}$  on the  $x$ -axis therefore yields an inverse power law with negative slope  $1/a$ , which is just the slope of the PDF  $(1 + 1/a)$  minus 1. Reali & Griffiths ([1], fig. 2) show word frequency PDF with exponent  $-1.74$ , so we expect a plot of turnover versus word frequency to have a power law slope of  $-0.74$ , which is quite close to their demonstrated slope of  $-0.8$ .

Thirdly, Reali & Griffiths [1] demonstrate ‘S’ curves of replacement that follow naturally from their rational-individual version of the model, following Wright–Fisher. This is an important realization, and by allowing for a population view, we can (i) add the memory parameter to gain flexibility in the resulting turnover, and (ii) apply to multiple variants in such contemporary representations as the ranked popularity list. To see why, consider an original pool of  $V$  different variants at time zero, and let  $x(t)$  be the number of new variants replacing this original pool. These  $V$  variants could represent the different words in a

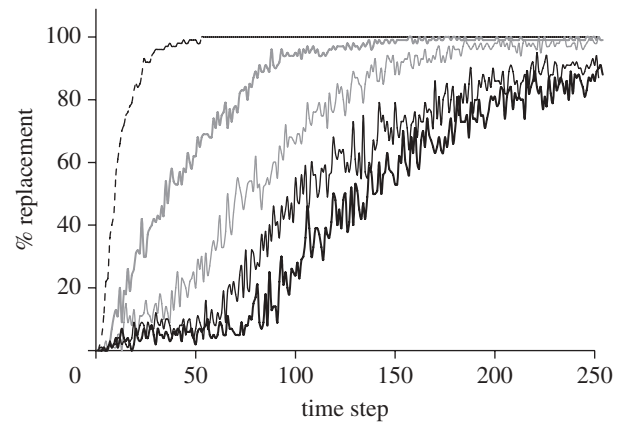


Figure 1. The effect of the simple memory parameter on replacement rate under the simple neutral model, with  $N = 1000$ , and  $\mu = 10\%$ . Here we consider an original vocabulary of 100 popular words, and how those words are replaced over time under the model (note that an ‘S’ curve does not follow from the simpler version of the model, where  $m = 1$ ). Dashed line,  $m = 1$ ; thick grey line,  $m = 5$ ; thin grey line,  $m = 10$ ; thin black line,  $m = 15$ ; thick black line,  $m = 20$ .

vocabulary, for example, or perhaps different possible grammatical rules in use. This replacement rate is proportional to the constant turnover rate,  $z$ , times the diminishing proportion  $(1 - x/V)$  of original variants remaining in the pool:

$$\frac{dx}{dt} = z\left(1 - \frac{x}{V}\right), \tag{1}$$

which is a separable differential equation, with a simple solution

$$x(t) = V\left(1 - e^{-zt/V}\right). \tag{2}$$

This is not an S curve, because even though it asymptotically approaches  $V$  over time, it begins quickly (when  $x = 0$ ,  $dx/dt = z$ ), rather than beginning slowly and speeding up. Adding the memory parameter, however, for certain combinations, such as  $m = 10$  time steps and  $\mu = 10\%$ , enables the modified neutral model to generate an S curve of replacement (figure 1).

In summary, the basic versions of the two models are mathematically the same, but radically different in interpretation: one is aimed at populations of simple imitators, whereas the other is focused on rationally based individual learning through iterated sampling. We argue that there are two reasons not to abandon the population representation. Firstly, it keeps things simple, and modifications and advances from this model—e.g. adding new extra parameters such as copying bias [6] or memory [8]—are more easy to benefit from and to apply generally. Secondly, in many situations of collective behaviour—which includes language, by its very definition—a representation of very simple social learning in a population may be more realistic than a model of individual rational actors. In any case, as vast sets of new word frequency data become available [22], the simplest, most flexible model has the best chance of generating insights across the range of social evolutionary phenomena where the same patterns occur.

R. Alexander Bentley<sup>1,\*</sup>, Paul Ormerod<sup>1,2</sup> and Stephen Shennan<sup>3</sup>

<sup>1</sup>*Department of Anthropology, Durham University, South Road, Durham DH1 3LE, UK*

<sup>2</sup>*Vòlterra Consulting, 135c Sheen Lane, London SW14 8AE, UK*

<sup>3</sup>*Institute of Archaeology, University College London, 31-34 Gordon Square, London WC1H 0PY, UK*

\*Author for correspondence (r.a.bentley@durham.ac.uk).

## REFERENCES

- 1 Reali, F. & Griffiths, T. L. 2010 Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proc. R. Soc. B* **277**, 429–436. (doi:10.1098/rspb.2009.1513)
- 2 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. (doi:10.1006/jaar.1997.0314)
- 3 Neiman, F. D. 1995 Stylistic variation in evolutionary perspective. *Am. Antiquity* **60**, 7–36. (doi:10.2307/282074)
- 4 Shennan, S. J. & Wilkinson, J. R. 2001 Ceramic style change and neutral evolution: a case study from Neolithic Europe. *Am. Antiquity* **66**, 577–594. (doi:10.2307/2694174)
- 5 Lieberman, E., Hauert, C. & Nowak, M. A. 2005 Evolutionary dynamics on graphs. *Nature* **433**, 312–316. (doi:10.1038/nature03204)
- 6 Mesoudi, A. & Lycett, S. J. 2009 Random copying, frequency-dependent copying and culture change. *Evol. Hum. Behav.* **30**, 41–44. (doi:10.1016/j.evolhumbehav.2008.07.005)
- 7 Bentley, R. A. 2008 Random drift versus selection in academic vocabulary. *PLoS ONE* **3**, e3057. (doi:10.1371/journal.pone.0003057)
- 8 Bentley, R. A., Ormerod, P. & Batty, M. In press. Evolving social influence in large populations. *Behav. Ecol. Sociobiol.* (doi:10.1007/s00265-010-1102-1)
- 9 Laland, K. N. 2004 Social learning strategies. *Learn. Behav.* **32**, 4–14.
- 10 Shennan, S. J. 2002 *Genes, memes and human history*. London, UK: Thames and Hudson.
- 11 Farmer, J. D., Patelli, P. & Zovko, I. I. 2005 The predictive power of zero intelligence in financial markets. *Proc. Natl Acad. Sci. USA* **102**, 2254–2259. (doi:10.1073/pnas.0409157102)
- 12 Kahneman, D. 2003 Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* **93**, 1449–1475. (doi:10.1257/00028280322655392)
- 13 Simon, H. A. 1955 On a class of skew distribution functions. *Biometrika* **42**, 425–440.
- 14 Bentley, R. A., Lipo, C. P., Herzog, H. A. & Hahn, M. W. 2007 Regular rates of popular culture change reflect random copying. *Evol. Hum. Behav.* **28**, 151–158. (doi:10.1016/j.evolhumbehav.2006.10.002)
- 15 Newman, M. E. J. 2005 Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351. (doi:10.1080/00107510500052444)
- 16 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), S120–S123. (doi:10.1098/rsbl.2003.0045)
- 17 Evans, T. S. 2007 Exact solutions for network rewiring models. *Eur. Phys. J. B* **56**, 65–69. (doi:10.1140/epjb/e2007-00084-8)
- 18 Bentley, R., Hahn, M. W. & Shennan, S. J. 2004 Random drift and culture change. *Proc. R. Soc. Lond. B* **271**, 1443–1450. (doi:10.1098/rspb.2004.2746)
- 19 Pagel, M., Atkinson, Q. D. & Meade, A. 2007 Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717–720. (doi:10.1038/nature06176)
- 20 Bentley, R. A., Ormerod, P. & Madsen, M. E. 2009 Physical space and long-tail markets. *Physica A* **388**, 691–696. (doi:10.1016/j.physa.2008.11.009)
- 21 Eriksson, K., Jansson, F. & Sjöstrand, J. 2010 Bentley's conjecture on popularity toplist turnover under random copying. *Ramanujan J.* **23**, 371–396. (doi:10.1007/s11139-010-9280-z)
- 22 Michel, J.-P. *et al.* & The Google Books Team. 2011 Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182. (doi:10.1126/science.1199644)