Advances in Complex Systems

ACCELERATED INNOVATION AND INCREASED SPATIAL DIVERSITY OF U.S. POPULAR CULTURE

R. ALEXANDER BENTLEY

Anthropology Department, Durham University
Durham, DH1 3LE, U.K.
r.a.bentley@durham.ac.uk

PAUL ORMEROD

Anthropology Department, Durham University
Durham, DH1 3LE, U.K.
Volterra Consulting, 135c Sheen Lane,
London, SW14 8AE, U.K.
pormerod@volterra.co.uk

Received (received date)
Revised (revised date)

At a time of increasing globalization of much popular culture, we observe not a rise but a marked fall in the spatial homogeneity of the popularity of first names across the United States in recent decades. We explain this apparent paradox by calibrating a modified standard model of neutral cultural evolution to the record of first name popularities for the United as a whole since 1900 and across the individual states over the last 50 years. We obtain estimates of both the temporal and spatial diversity of the speed of cultural evolution during the 20^{th} century and early 21^{st} century. We find that the speed of innovation of popular baby names accelerated substantially since the end of the 20^{th} century, increasing the geographic diversity across the United States. This suggests that accelerated innovation can lead to a more heterogeneous cultural landscape, even in the midst of growing globalization.

Keywords: Cultural transmission; evolution; social learning, cultural drift, pattern formation.

1. Introduction

Modern popular culture undergoes competing forces; globalization promotes homogeneity, whereas local innovation promotes diversity. The question of how local differences arise amid regular social interaction could be considered one of the essential pursuits of social science. It is a question which is receiving a great deal of renewed attention in the current century of online communications and global but decentralized popular culture.

Approaches range from the highly descriptive, in-depth socio-cultural research of local communities to, at the other end of the spectrum, quantitative models aimed at general insights into group-level phenomena. Among the latter has been game theory, in which a new frontier has emerged, in modeling ways that spatial heterogeneity of group-level norms can evolve [1, 2]. Related cultural evolution research indicates how such geographic heterogeneity can be attributable to different local adaptations [3, 4], but might equally reflect chance or drift, in the unique evolutionary histories of social learning.

One useful proxy for investigating popular culture is through records of first names, which can reveal aspects of kinship patterns, popular culture trends, and social values. They are found throughout time and space, are easily measured and counted [5–9]. In modern Western culture, the choices of first names reflect three general principles of collective behavior that apply to fashion (popular culture) in general [2, 10]: (a) they involve a number of people carrying out the same or similar actions at a point in time; (b) the behavior exhibited is transient or continually changing; (c) there is some kind of dependency among the actions, individuals are not acting independently [7].

Here we analyze data on first name popularities in the United States to estimate the temporal and spatial evolution of the relative weights of the globalization factors which lead to homogeneity in popular culture, and the speed of local innovation, which promotes diversity.

Despite an increasing globalization of popular culture in general in the United States [7, 11, 13], the naming landscape has diversified markedly in the last fifty years. In 1960 the most popular girls' name was 'Mary' in almost every U.S. state except for 'Susan' in the Northwest and Northeast, and some variety in the Western states (Figure 1a). This homogeneity in 1960 is also reflected in boys' names (Figure 1c), when the five that were locally most popular (David, James, Michael, John, Robert) comprised the top five for most states, and none was lower than 8^{th} place in any state. There was also a clear geography to the boys' names in 1960; 'David' was most popular in almost all states west of the Mississippi River (Figure 1c), 'James' was most popular in the South, and 'John' and 'Michael' dominated the Northeast.

By 2009, however, this homogeneity was broken up considerably. The hegemony of 'Mary' has given way to multiple competing girls' names (Figure 1b), and the geography of boys' names has become even diverse (Figure 1d). Among boys' names in 2009, some continuity remained in the South (where 'William' replaced 'James') and northeast (where 'Michael' gained), but west of the Mississippi the geographic pattern had dissolved into a mosaic (Figure 1d). The differences between states have deepened as well; in 2009, a boy's name such as 'Logan' - the most popular name in Minnesota, Idaho and New Hampshire - was not even among the top 30 in New Jersey or California.

This new geographic heterogeneity in name popularity reflects the process of cultural transmission over the years. It has been said that the choice of a name 'connects us to society in a way that encapsulates the great contradiction in human social life: between the desire to fit in and the desire to be unique' [14]. The question

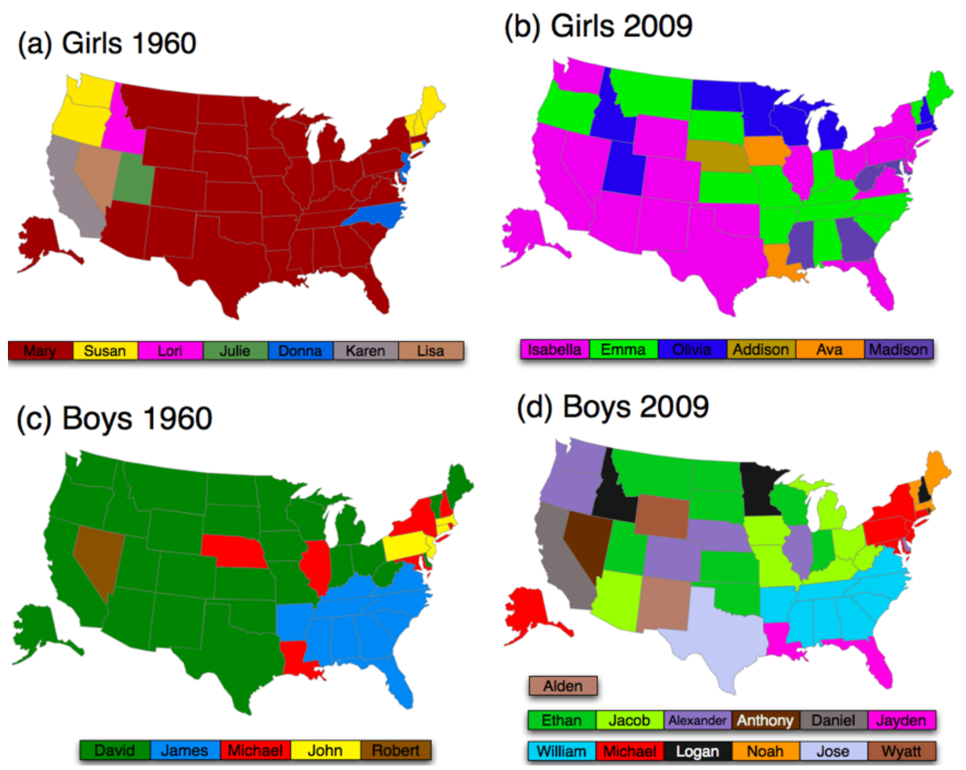

Fig. 1. Spatial heterogeneity in the most popular name for each state. **(a)** Girls, 1960; **(b)** Girls, 2009; **(c)** Boys, 1960; **(d)** Boys, 2009

is, how has this balance changed? Even today, truly unique names are relatively rare; the vast majority of first names given to babies have previously been given to someone else, i.e., most names are copied.

While copying is ubiquitous in human social behavior, original independent invention is nonetheless persistent, if relatively weak compared to social learning. In one Repeated Prisoner's Dilemma Game experiment involving real people, about 10% of players continued inventing new strategies until the last rounds [15]. In real society, most studies indicate that ten percent appears to be around the upper bound of the percentage of truly original behavior in a population [33, 34]. No matter how it originates, the persistent minority of independent action injects new variation into the population, for those in the majority potentially to copy [16].

Since copying with occasional original invention is closely analogous to inheritance with sporadic mutation, simple models of 'unbiased' social transmission have been adapted in studies of cultural evolution from neutral models of population biology [17–19]. In particular, a model of unbiased social learning in a well-mixed, homogeneous population, has been productively applied to cultural evolution [20].

This process was used, for example, to identify social network arrangements favorable to selection versus random drift in the sorting of variation [20], and it was also used to show how highly clustered social networks favor the evolution of cooperation[21].

Here we use a model of unbiased copying based on the same principles, but adding the variable of 'memory' (crucial in the social learning tournament), since social learning can reach back to previous 'generations' of behavior. We consider a homogeneous population of size N, each individual agent of which holds a single variant, representative of some particular choice or behavior. The central parameters are invention fraction μ, and memory m. The effective population size remains constant, but the population is replaced in each generation by N new agents. With probability $1-\mu$, an incoming agent copies its variant from that of an agent within the previous m time steps, or else with probability μ, the agent innovates by choosing a unique new variant at random. In other words, the agent either copies the choice of an existing agent from the last m steps, or chooses a new variant.

Previous versions of this unbiased copying model have proven powerful and versatile, in applications to phenomena where widespread copying is inherent to popularity. The model yields a wide variety of heavy tailed distributions through varying μ and m, which can be closely fit to popularity distribution from phenomena that also undergo continual turnover, or drift, in these distributions, which also is replicated by the model [22, 23, 32].

The unbiased copying model is well suited to the analysis of popular baby names, considered to serve as a proxy for popular cultural exchange. Each name is a discrete entity that is either a copy of an existing name (usually) or else completely novel (rarely). To apply the model to baby names, each baby is considered an agent, and its name as the associated variant.

The United States Social Administration provides a database on baby names that has been studied in a variety of ways [8, 24, 25], owing to its exceptionally deep and chronologically-resolved records. These superb data include the top 100 baby names by US state since 1960, and, for the US as a whole, all of names with at least 5 occurrences in each year since 1879 [44]. With population increase and the issuing of Social Security Cards in 1937, the length of this list of names increases tremendously over the 130 year record - from a few thousand different names for boys and for girls in the early 1900s, to fourteen thousand different boys' names and over twenty thousand girls' names in 2009[a].

More recently data were made available for the top 100 names in individual U.S. states. For each state, these data were compared with the top 100 observations generated by the theoretical model (see methods, below). All data were normalized so that each observation is expressed as its proportion of the total sum of the raw

[a]Because the SSA do not list any names with fewer than five occurrences (in order to safeguard privacy), it is not possible to derive an invention rate directly from the number of unique names in a given year.

observations of the top 100. For consistency with the Top 100 by state, we fit our model results to the Top 100 names in the U.S. as a whole, and derive the U.S. invention fraction μ through time.

2. Methods

We examined 125 combinations of the model parameters, with μ between 0.01 and 0.1 and values of $m \leq 10$ (consistent with the evidence for recent memory being the most valuable in social learning [29] and plausible values for μ being less than 10% [15, 33, 34, 16]. For each parameter combination, we averaged 500 independent results of the model, each one running for 2000 time steps. We tested the null hypothesis that the distributions of the empirical data for any given geographical entity and year and the model generated data for any given parameter combination are the same. We used the Anderson-Darling test, a more powerful test than the widely used Kolmogorov-Smirnov when the data are distinctly non-Gaussian. Only a distinct minority of model parameter combinations were compatible with the empirical data in any given case. Of the 125 combinations of the model parameters examined for each state, about 100 of the resulting distributions could be rejected ($p < 0.05$). In contrast, for many states there were parameter combinations where the null hypothesis is only rejected at $p \geq 0.5$. We found that among the small set of accepted parameter combinations for each state, the effect of the accepted m values was not significant on the resulting distributions. We therefore report results simply for the average value of μ for these cases. For the state-level data, we fit the model to the top 100 baby names from each individual state in the year in question, and then averaged the values of μ among the small minority of solutions where the null hypothesis is only rejected at $p > 0.20$, which implies that the model fits the data extremely well.

3. Results

For the U.S. as a whole, we find that for most of the 20^{th} century the model-derived μ was essentially constant for both boys and girls. However, there was a rapid acceleration around the turn of the 20^{th} century, with estimates of μ more than doubling of μ for girls and tripling of μ for boys by the end of the first decade of the 21^{st} century (Figure 2).

As a check on this, we obtained an independent estimate of μ through estimated proportion of unique names, which we can extrapolate approximately from the most complete U.S. data offered [b]. Interestingly, the correlation with this independent

[b] We obtained an independent estimate of μ using the most complete US baby name data available, listing down to names with only 5 occurrences in a year. In every year, the number of babies with each name are reasonably approximated by a power law distribution through the first several orders of magnitude, with scaling exponents ranging from 1.56 to 1.72 for boys, and 1.62 to 1.76 for girls, using maximum likelihood estimation via Aaron Clauset's MatLab code [12]. We used

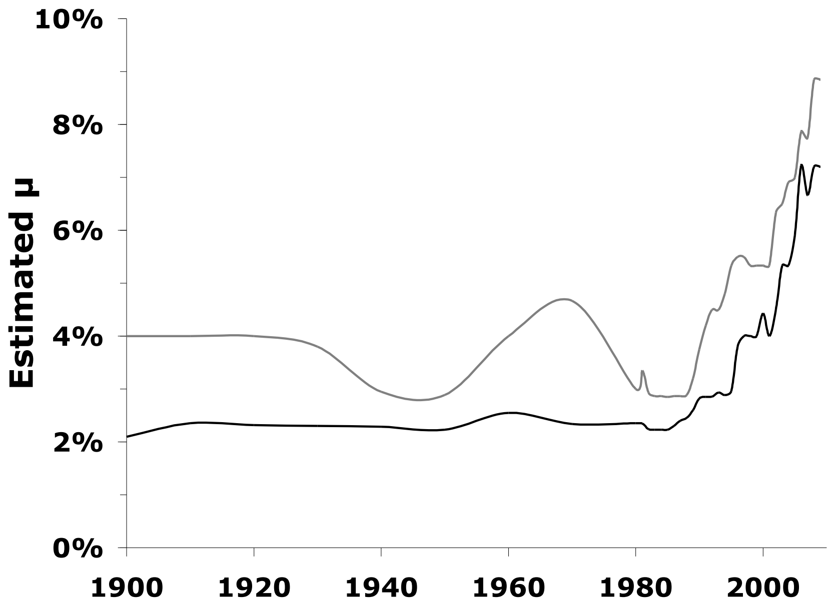

Fig. 2. Mean 'invention' score μ for all U.S. names (grey line, girls; black line, boys), estimated using the unbiased copying model from the top 100 most popular names in the U.S. as a whole.

estimate of μ is best when we allow a twenty-year lag ($r^2 > 0.89$ for both girls' and boys' names); that is, if we compare the modeled estimate of μ, derived from the top 100, to the fraction of unique names twenty years before. The implication is understandable; that changes in invention rate take about a generation (≈ 20 years) to affect the top 100.

As inventiveness μ increased nationwide over the last 50 years, there arose significant differences between individual states across the U.S.. The maps in Figure 3 compare the derived values of μ for individual states in 1960 (the first year of state data) versus 2009. In 1960 there was a noticeable gradient across the country, for both boys' and girls' names, in that the μ values are significantly lower in states of the North Eastern U.S., and higher in the Southern and Western states (Figure 3).

This geographic pattern in 1960 correlated with a geographic cultural landscape familiar to most Americans. The U.S. Bureau of Economic Analysis (BEA)[26], for example, divides the U.S. into eight geographic regions, numbered sequentially from

the parameters for each distribution to extrapolate back to the number of baby names with 2 occurrences. To estimate the number of unique names in a given year, allowing for a possible abundance of one-off names, we subtracted the total with > 2 occurrences from the total births for the year (from the SSA).

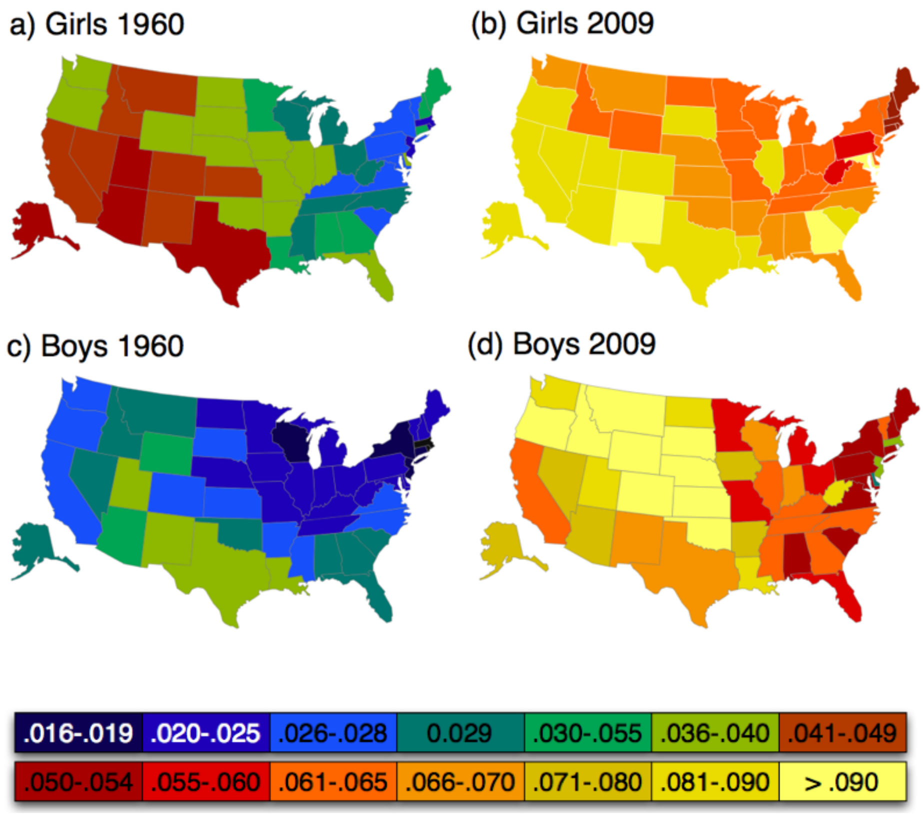

Fig. 3. Temporal increase and geographic pattern in name innovation. Colors depict the mean invention score μ, derived from the model and the top 100 names, for each state. **(a)** Girls, 1960; **(b)** Girls, 2009; **(c)** Boys, 1960; **(d)** Boys, 2009. In every case the null hypothesis that the distributions of model solution with this value of μ and the empirical data are only rejected at a $p > 0.20$, far higher than the standard value of rejection of $p0.05$, implying a close correspondence of the distributions. (the only exception is MD where $p > 0.15$ is used.

New England to the Far West. As Figure 4 shows, the average model-derived values of μ for 1960 correlate well with these BEA region numbers ($r^2 = 0.59$ among boys' names, $r^2 = 0.49$ for girls' names). By 2009, however, these geographic correlations had diminished ($r^2 = 0.40$ and 0.45 for boys' and girls' names, respectively), with more outliers (Figure 4). This was coupled with an almost uniform rise in innovation fractions across regions (Figure 4).

4. Discussion

Overall, the diversification of naming practices across the US appears to be driven by an increase in innovation that drives further drift. In this arena, the diversifying effects of novel invention would seem to have the edge over the homogenizing

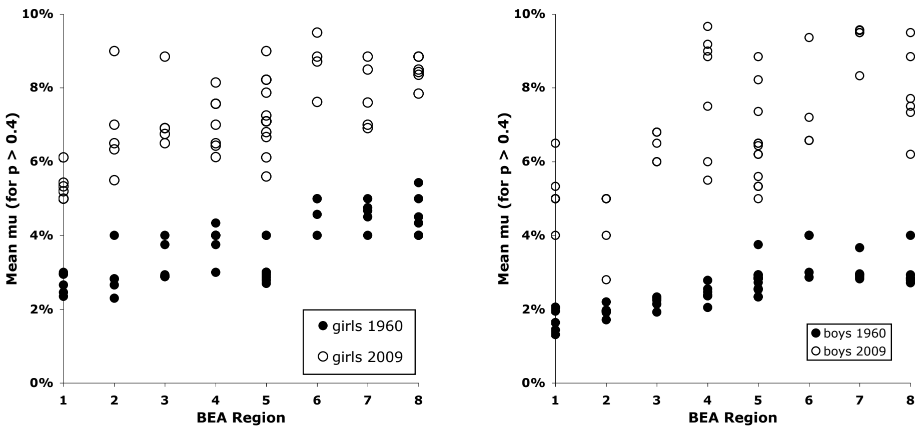

Fig. 4. Model-derived μ versus economic region of the U.S., for girls' names (left) and boys' names (right). The BEA region numbers are: (1) New England, (2) Mideast, (3) Great Lakes, (4) Plains, (5) Southeast, (6) Southwest, (7) Rocky Mountain, and (8) Far West.

effects of globalization. It exemplifies how social learning can yield spatial diversity in cultural choices as different portions of the population drift apart stochastically [1]. As an efficient strategy of social learning, copying other individuals is often successful, and is seen across a range of species, from fish to primates and humans right from infancy [27–30]. The principle is reflected in a recent computer tournament, where entrants submitted strategies specifying how to use both social learning and alternatives such as trial-and-error to acquire adaptive behavior in a complex multi-player environment. The most successful strategy in the tournament was not (as many expected) a predominance of independent learning supplemented by a relatively minor degree of social learning. In fact, the winning strategy in the tournament relied almost exclusively on copying the successful strategies of others, biased towards copying recent success by 'discounting' older information [29].

Our hypothesis is that this diversification since 1960 is due to random drift via social learning, but an alternative hypothesis is immigration. Although 'Jose' is currently the most popular boys name in Texas, and of course different ethnic groups have distinctive sets of names [7], migration and ethnic diversity appear to be minor factors in the geographic diversification of US name popularity. The model-derived average μ values correlate only weakly with the number of foreign-born residents by state (for 2008, $r^2 = 0.08$ for girls, $r^2 = 0.14$ for boys) [31].

Another factor is surely the recent decline of monolithic mass media, first to a multiplicity of media sources, and finally the online media that much of the public now participates in. If the cultural landscape is more conducive now to individuals copying each other, rather than people responding to centralized media, we could expect the kind of diversification through local interactions that Thomas Schelling demonstrated in his classic segregation model [10]. Names are but one example of course, but this phenomenon suggests a way in which the interconnectedness of 'globalization' - a word that tends to connote homogeneity - may instead promote cultural diversity by allowing random drift to occur more pervasively

References

[1] Helbing D., Yu W., The outbreak of cooperation among success-driven individuals under noisy conditions. *PNAS* **106** (2009) 3680-3685.

[2] Henrich J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., Camerer, C., McElreath, R., Gurven, M., Hill, K., Barr, A., Ensminger, J., Tracer, D., Marlowe, F. W., Patton, J. Q., Alvard, M., Gil-White, F. J., and Henrich, N. S., "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* **28** (2005) 795-855.

[3] Laland K. N., Odling-Smee J., and Myles, S., How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews Genetics* **11** (2010) 137-148.

[4] Richerson P. J, Boyd R., and Henrich, J., Gene-culture coevolution in the age of genomics. *PNAS* **107** (2010) 8985-8992.

[5] Zelinsky, W., Cultural variation in personal name patterns in the Eastern United States. *Annals of the Association of American Geographers* **60** (1970) 743-769.

[6] Lieberson, S., *A Matter of Taste: How Names, Fashions and Cultures Change* (Yale University Press, New Haven, CT, 2000).

[7] Fryer R. G., and Levitt, S. D., The causes and consequences of distinctively black names. *Quarterly Journal of Economics* **119** (2004) 767-805.

[8] Berger, J., and Le Mens, G., How adoption speed affects the abandonment of cultural tastes. *PNAS* **106** (2009) 8146-8150.

[9] Jones, D., The universal psychology of kinship: Evidence from language. *Trends in Cognitive Sciences* **8** (2004) 210-215.

[10] Schelling, T. C., Dynamic models of segregation. Journal of Mathematical Sociology **1** (1971) 143-186.

[11] Amin, A., Spatialities of globalisation, *Environment and Planning A* **34** (2002) 385-399.

[12] Clauset, A. Shalizi, C. R. and Newman, M. E. J., Power-law distributions in empirical data. *SIAM Review* **51** (2009) 661-703.

[13] Crothers, L., *Globalization and American Popular Culture* (Rowman and Littlefield, Boston, 2009).

[14] Pinker, S., *The Stuff of Thought: Language as a Window into Human Nature*, (Penguin, New York, 2007).

[15] Traulsen A., Semmann, D., Sommerfeld, R. D., Krambeck, H.-J., and Milinski, M., Human strategy updating in evolutionary games. *PNAS* **107** (2010) 2962-2966.

[16] Kirby, S., Cornish, H., and Smith, K., Cumulative cultural evolution in the laboratory. *PNAS* **105** (2008) 10681-10686.

[17] Hubbell, S. P., *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, 2001).

[18] Iizuka, M., Tachida, H., and Matsuda, H., A neutral model with fluctuating population size and its effective size. *Genetics* **161** (2002) 381-388.

[19] Lachlan, R. F., and Slater P. J. B., Song learning by chaffinches: how accurately and from where? A simulation analysis of patterns of geographical variation. *Animal Behaviour* **65** (2003) 957-969.

[20] Lieberman, E., Hauert, C., and Nowak M. A., Evolutionary dynamics on graphs. *Nature* **433** (2005) 312-316.

[21] Ohtsuki, H., Hauert, C., Lieberman, E., and Nowak, M. A., A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441** (2006) 502-505.

[22] Bentley, R. A, Ormerod, P., and Batty, M., Evolving social influence in large populations. *Behavioral Ecology and Sociobiology* (2011) in press.

[23] Eriksson, K., Jansson, F., and Sjostrand, J., Bentley's conjecture on popularity toplist turnover under random copying. *The Ramanujan Journal* (2011) in press.

[24] Hahn, M. W., and Bentley R. A., Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society B* **270** (2003) S120-S123.

[25] Bentley R. A., and Ormerod, P., Traditional models already explain adoption-abandonment pattern. *PNAS* **106** (2009) E109.

[26] http://www.bea.gov/regional/docs/regions.cfm

[27] Laland K. N., Social learning strategies, *Learning and Behavior* **32** (2004)4-14.

[28] Kendal, J. R., Rendell, L., Pike, T. W., and Laland K. N., Nine-spined sticklebacks deploy a hill-climbing social learning strategy, *Behavioral Ecology* **20** (2009) 238-244.

[29] Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Fledman, M. W., Fogarty, L., Ghirlanda, S., Lillicrap T., and Laland, K. N., Why copy others? insights from the social learning strategies tournament. *Science* **328** (2010) 208-213.

[30] Gergely, G., Bekkering, H., and Kiraly, I., Rational imitation in preverbal infants, *Nature* **415** (2002) 755.

[31] http://www.migrationinformation.org/datahub/acscensus.cfm

[32] Bentley, R. A., Lipo, C. P., Herzog, H. A., and Hahn M. W., Regular rates of popular culture change reflect random copying. *Evolution and Human Behaviour* **28** (2007) 151-158.

[33] Eerkens, J. W. Practice makes within 5% of perfect. *Current Anthropology* **41** (2000) 663-668.

[34] Rogers, E. M., *Diffusion of Innovations* (Free Press, New York, 1962)

[35] Simon, H. A., On a class of skew distribution functions. *Biometrica* **42** (1955) 425-440.

[36] Gallegati, M., Keen, S., Lux, T., and Ormerod, P., Worrying trends in econophysics. *Physica A* **370** (2006) 1-6.

[37] Salganik, M. J., Dodds, P. S., and Watts, D. J. Experimental study of inequality and unpredictability in an articial cultural market. *Science* **311** (2006) 854-856.

[38] Kahneman, D., Maps of bounded rationality: psychology for behavioral economics. *Am. Economic Rev.* **93** (2003) 1449-1475.

[39] Smith, V. L., Constructivist and ecological rationality in economics. *Am. Economic Rev.* **93** (2003) 465-508.

[40] Akerlof, G. A., The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly J. Economics* **84** (1970) 488-500.

[41] Stiglitz, J. E., Information and the change in the paradigm in economics. *Am. Economic Rev.* **92** (2002) 460-501.

[42] Stigler G. J., The economics of information. *J. Political Economy* **69** (1961) 213-225.

[43] Dunbar R. I. M., and Shultz, S., Evolution in the social brain. *Science* **317** (2007) 1344-1347.

[44] Data: www.ssa.gov/OACT/babynames