The Persistence of Inferior Cultural-Institutional Conventions

Marianna Belloc
Samuel Bowles

SFI WORKING PAPER: 2013-01-002

The Persistence of Inferior Cultural-Institutional Conventions*

Marianna Belloc† and Samuel Bowles‡

January 24, 2013

Abstract

Our theory of cultural-institutional persistence and innovation is based on uncoordinated updating of individual social norms and contracts, so that both culture and institutions co-evolve. We explain why Pareto-dominated cultural-institutional configurations may persist over long periods and how transitions nonetheless occur. Unlike models in which elites impose inferior institutions or cultures as a self-interested distributional strategy, in our model, the exercise of elite power plays no role in either persistence or innovation, and transitions occur endogenously. We show that persistence will be the greater the more inferior is the Pareto-dominated configuration and the more rational and individualistic is the population.

Keywords: endogenous institutions, endogenous social norms, cultural-institutional persistence, evolutionary game theory

JEL Classification: C73, D02, D03

* *American Economic Review: Papers and Proceedings* in press. Acknowledgments: Thanks to the Santa Fe Institute Behavioral Sciences program for support of this work, and to Quamrul Ashraf, Roy Cerqueti, Sung-Ha Hwang, Suresh Naidu, and Peyton Young for comments.

† (Corresponding author) Sapienza University of Rome, via del Castro Laurenziano 9, 00161, Rome (Italy). Email: marianna.belloc@uniroma1.it.

‡ (Corresponding author) Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, 87501, NM (USA). Email: bowles@santafe.edu.

1 Introduction

Economic institutions and cultures (including social norms) are often dynamically complementary, meaning that the persistence of each is facilitated by the presence of the other. The term "feudal society", for example, refers jointly to the economic relationship of lord and serf and to the norms of subordination and reciprocity that both contributed to the smooth functioning of the system and that were its cultural expression. This complementarity provides one mechanism for the long-term persistence of particular configurations of cultures and institutions. Given the institutional relationship of serf to lord (to continue the example), adopting the culturally prescribed norms of subordination and reciprocity was a best response for individuals in the two classes respectively. And given this culture of subordination and reciprocity, conforming to the institutional arrangements defining serf and lord was also a best response. We refer to this pair of mutual best responses as a cultural-institutional convention.

Otherwise similar populations with differing recent histories may exhibit differing conventions: free cities coexisting with feudal manors in Germany and elsewhere during the 13th-15th centuries, for example. Historians and social scientists have long asked why some cultural-institutional conventions appear independently multiple times in human history and persist over long periods – monogamy, markets and primogeniture, for example – while others rarely emerge, and when they do tend to be short lived (Parsons, 1964). In reply, some economists simply extend invisible hand arguments to the selection of institutions. Thus Oliver Williamson (1985: 394) writes that "viable modes of economic organization...ordinarily possess an efficiency advantage." The mechanisms that would account for this felicitous result, however, have remained elusive, and evidence of enduring institutional failure is widespread. Accounts of socially dysfunctional but enduring cultural practices suggest that invisible hand arguments work no better when applied to cultures (Edgerton, 1992).

2 Why do inferior cultures and institutions endure?

Given that culture and institutions are often implicated in explaining enduring poverty (e.g. Clark, 2007; Acemoglu and Robinson, 2012) a pressing question is: what accounts for the persistence of cultural-institutional conventions that are inferior in the sense that almost everyone could be made better off under an alternative set of technically feasible configurations? "Cultural inertia" is sometimes said to be the result of the transmission of learned behaviors from parent to child; but for plausible degrees of transmission this process alone would result not in persistence but in the dissipation of cultural differences between populations in a matter of just a few generations (see appendix). Moreover, in light of recent history, persistence cannot be an intrinsic characteristic of either culture or institutions. Examples include the precipitous demise of such well defended institutions as Communist Party rule in many countries and of apartheid in South Africa and the extraordinarily rapid spread or retreat of cultural practices such as female genital cutting in parts of Africa and the use of honorific pronouns in many European languages.

A more plausible answer – proposed in a variety of models and documented with historical and contemporary examples – is that a concentration of political power allows favored groups to command a larger slice by means of policies that result in a smaller pie (Sokoloff and Engerman, 2000; Acemoglu, 2003). The hypothesis that Pareto-dominated allocations are implemented as part of a strategy of distribution is readily motivated by problems of commitment. For example an elite may resist moving to a Pareto-superior convention because there is no way that non elites can commit to not exploiting the instability of that superior convention in order to introduce a further transition under which the elite would lose.

But there is another way that inferior cultural-institutional conventions may persist indefinitely, one that relies on strategic complementarity rather than cultural transmission across generations or elite power. If individual conformity to the status quo institutions and cultural norms is a mutual best response, and if individuals update their behaviors non-cooperatively, an inferior convention can persist simply because it is evolutionarily stable by dint of its being a mutual best response (Young, 1998). This "bottom-up" mechanism for cultural-institutional persistence is complementary to the "top-down" models just mentioned. But the mechanisms accounting for persistence are diametrically opposite. In the top-down models, institutions persist because elites are organized and powerful enough to implement allocations that limit the claims of others. In our bottom-up model, inferior cultural-institutional conventions persist because nobody is organized in that sense, and the actions of individuals in conforming to or deviating from the status quo institution and cultural norm are entirely decentralized.

There are other ways in which our approach is distinct. First, we explicitly model the interactions of cultures and institutions and their co-evolution rather than treating institutional or cultural dynamics in isolation. Second, we consider large populations without political differentiation, so that no single actor (for example an "elite") has any appreciable influence on outcomes. Third, both the persistence of cultural-institutional conventions and transitions between them are captured in the same model, without the intervention of exogenous changes. Finally, in contrast to many of the classical game theoretic treatments, our agents, while strategic, have limited cognitive capacities, updating their culture and institutions on the basis of past distribution of institutional and cultural traits in the population rather than foresight.

Whether these *differentia specifica* of our model are taken to be features rather than bugs will of course depend on the questions at hand. But we think that the resulting model provides insights complementary to the top-down approach in the understanding of such durable institutions as land tenure, inheritance systems and property rights more generally, as well as employment contracts and marital practices. It also provides important insights into bottom-up "cultural-institutional tipping events" resulting in rapid transitions. An example is the end of apartheid in South Africa: individual firms and trade unions were privately working out the terms of a non-racial order years before the ruling National Party freed Nelson Mandela and conceded non-racial elections (Wood, 2000).

3 A bottom-up model of cultural-institutional persistence and innovation

We study the evolutionary dynamics of culture and institutions in an economy with two classes. These are large sub-populations whose z members are paired randomly to interact in a non-cooperative game governed by a set of institutions. The classes may be employees and employers, share cropping farmers and landlords, slaves and slave owners, independent farmers and government officials, and so on. As these examples suggest, these class relations are asymmetric. The alternative economic institutions governing relations between the classes are one of two contracts, which are implemented by the members of the second class (who we call the Bs) in each of the pairs just mentioned. The first class (the As) may adopt two alternative social norms. To represent the complementarity between cultures and institutions and the possibility of the persistence of inferior cultural-institutional conventions, we assume that some contract-norms matches are more productive than others and can be Pareto-ranked.

For concreteness, think of a somewhat idealized rendition of the institutional structure and culture of two firms, Volkswagen and Fiat (Jurgens, 2002; Nuti, 2001). In the former, an institutional structure based on a works-council and co-management matches with a workforce with norms of cooperation with management, resulting in high levels of productivity and, as a result, mutual gains. In the latter, a top-down management structure is matched with an oppositional workers' culture resulting in reduced productivity. What is important for our model is not only that the idealized Volkswagen match Pareto-dominates the Fiat match, but that the two matches are best responses for owners and workers alike. Given the oppositional culture of Fiat's workers, the owners would do even worse if they implemented a co-management structure; a militant oppositional culture would not benefit Volkswagens workers; and so on.

We index contracts and norms by j=0,1 and classes by i=A,B, and represent the payoffs to the four possible cultural-institutional matches as $\pi^i_{11} > \pi^i_{00} > \pi^i_{01} > \pi^i_{10} = 0$ for both classes, where, for example, π^i_{10} is the payoff of an individual in class i implementing contract 1 (adopting norm 1) when his partner from the other class adopts norm 0 (implements contract 0), and payoffs are normalized so that the two "mismatches" are zero. Expected payoffs for members of class i implementing contract j (or adopting norm j) are given by $\nu^i_j = (1-\phi_{-i})\pi^i_{j0}+\phi_{-i}\pi^i_{j1}$, where ϕ_{-i} denotes the fraction of individuals in the other class $(-i)$ who implemented contract 1 (adopted norm 1) in the previous period. Expected payoffs lines are shown in Fig. 1.

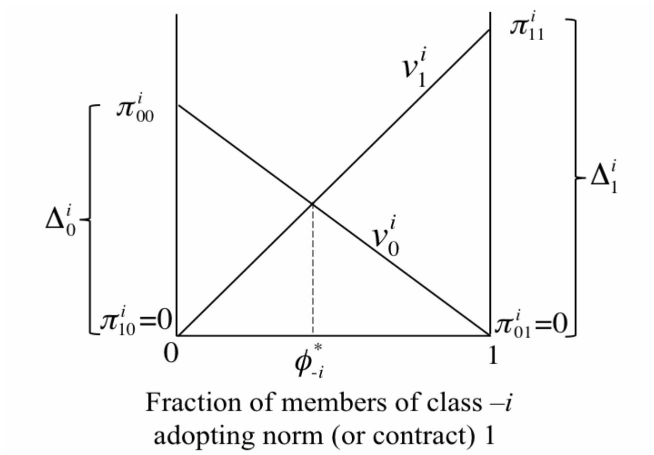

Figure 1. *Expected payoffs of members of class i*

The state space for this process (shown in Fig. 2) is all possible combinations of the number of individuals in the two classes respectively adopting norm 1 and implementing contract 1, $(z\phi^*_A, z\phi^*_B)$. The two matches with non-zero payoffs are Pareto-ranked Nash equilibria, and in the Markov process that represents this model the states (denoted {0,0} and $\{z,z\}$) in which all members of both classes adopt one or the other of these two profiles are absorbing when all individuals best respond.

Following matching, members of each class have the opportunity to update their contracts and norms. To ensure that the process is acyclic, we let the revision process be asynchronous (as in Binmore, et al., 2003) with all members of one class updating first, followed by updating of the other class. When revising their contracts and norms, best responding individuals maximize their expected payoffs based on the distribution of respectively norms and contracts in the sub-population with which they are matched in the previous period.

But individuals are boundedly rational and with probability $\sigma > 0$ they adopt the norms or institutions that are not the best response, with σ strictly decreasing in both the cost of deviating from the best response and the agents' degree of rationality (defined below).

Following Blume (2003), the probability of deviating from the best response when the population is at {0,0} is:

$$\sigma_i(\Delta^i_0, \beta) = \frac{1}{1 + e^{\beta \Delta^i_0}}, \text{ with } i = A, B, \tag{1}$$

where $\Delta^i_0 = \pi^i_{00} - \pi^i_{10}$ is the cost of deviation from the status quo culture or institution at {0,0}. We interpret β as a measure of rationality because the larger is β, the smaller the probability that the individual will deviate from the best response. When $\beta = 0$ the agent randomizes between the two alternatives, and as β goes to infinity, the individual never deviates. Of course, individuals may have non-economic reasons to deviate from the status quo

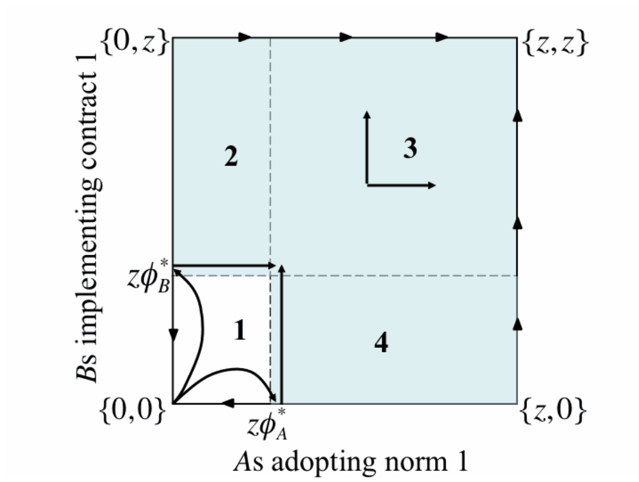

Figure 2. *State space and transitions*

even in the absence of cognitive failures; so, strictly speaking, β measures the degree to which agents maximize their expected payoffs.

For sufficiently rational agents, once a population is in the neighborhood of either of the two absorbing states, the associated convention may persist over very long periods. The reason is that, for sufficiently large populations and sufficiently rational individuals, the expected waiting time for a realization of sufficient non-best responses to tip the process from the neighborhood of one convention to the basin of attraction of the other will be very prolonged. Cultural-institutional conventions are perpetuated in every period; inertia is not involved as individuals have just a single period memory.

4 Impediments to Pareto-improving cultural and institutional change

To study transitions from the inferior {0,0} to the superior convention $\{z,z\}$, we first determine the minimum numbers of deviant members of each class, such that with sufficiently rational agents, the population will enter the basin of attraction of the superior convention. The basin of attraction of a state is the set of states from which, for the above dynamics and sufficiently rational agents, the revision process we have just described leads to that state. For sufficiently rational individuals, beginning near the convention {0,0} the state following updating will almost always lie within quadrant 1 in Fig. 2 and, because agents have just a one-period memory, even an unlikely substantial excursion away from the convention will have no lasting effect.

But suppose that, from the initial state {0,0}, $z\phi_B^*$ of the Bs deviate from the status quo institution and offer contract 1 instead of best responding with contract 0, where $z\phi_B^*$ is the smallest number such that the As' best response is to switch to norm 1 and $\phi_B^* = \Delta_0^A/(\Delta_0^A + \Delta_1^A)$. (See Fig. 1. Since z is large we may avoid notational clutter by abstracting from integer considerations.) In response, each A will adopt norms 1 with probability $1 - \sigma$. But as β goes to infinity σ goes to zero, so there exists some finite $\hat{\beta}$ such that for $\beta > \hat{\beta}$, as a result of the As' updating, with virtual certainty we will have

$z\phi_A > z\phi_A^*$, where $z\phi_A^*$ is the smallest number such that the Bs' best response is to switch to contract 1 (with $\phi_A^* = \Delta_0^B/(\Delta_0^B + \Delta_1^B)$). When this occurs, the population will be in the set of states for which both classes' best responses will lead to $\{z,z\}$. Thus the minimum number of Bs implementing contracts 1 sufficient to escape from the inferior convention is $z\phi_B^*$. Analogous reasoning applies to the minimum number of innovating A-members, $z\phi_A^*$, sufficient to induce a transition to $\{z,z\}$. It follows that from the initial state $\{0,0\}$ the basin of attraction of the superior convention is composed of quadrants 2, 3 and 4 in Fig. 2.

Because deviations from the best response contract or social norm are independent, expected waiting times for a transition from one absorbing state to the other induced by each of the classes respectively are approximated by the inverse of the probability that in a given period the number of deviants of that class will be sufficient to enter the basin of attraction of the other convention. For large populations and sufficient rationality, this probability is approximated by P_i, the likelihood that exactly the minimum number of innovators in class i (ϕ_i^*) will occur (Binmore, et al., 2003. Our results are not affected by taking account of the probability that larger than minimal numbers deviate.):

$$P_i = \binom{z}{z\phi_i^*}(\sigma_i)^{z\phi_i^*}(1-\sigma_i)^{z-z\phi_i^*}, \ i = A, B. \tag{2}$$

Hence the expected waiting time for a transition is the inverse of the probability that the number of innovators in at least one class will be sufficient to tip the population to the basin of attraction of $\{z,z\}$, that is

$$E[W] = (P_A + P_B - P_A \times P_B)^{-1}. \tag{3}$$

5 Discussion

Equations (2) and (3) give us four results applicable equally to either of the two classes (proofs in the appendix). First, "culture- or institution-biased" technical change may accelerate transitions by making an alternative convention more productive relative to the status quo. We find that for sufficiently rational agents, because both of the critical fractions required for a transition from $\{0,0\}$ to $\{z,z\}$ – ϕ_A^* and ϕ_B^* – are decreasing in the productivity advantage of the superior convention (Δ_1^i), the expected waiting time for a transition ($E[W]$ given by (3)) is decreasing in the superiority of the Pareto-dominant convention. Our evolutionary dynamic thus favors superior cultural-institutional configurations.

But, second, because deviations from the status quo are less likely the greater is the degree of individual rationality, the expected waiting time for a transition is increasing in β. Then, for sufficiently rational agents, a cultural-institutional convention can last virtually forever even if there exists an alternative Pareto-superior convention. There is no invisible hand for cultural-institutional configurations, at least not on historically relevant time scales.

Third, the greater is the cost of deviating from the inferior culture or institution (Δ_0^i), the longer will be the expected waiting time for a transition to the superior convention. This unsurprising result has a somewhat unexpected implication (Belloc and Bowles, 2013): because the gains from trade increase the payoffs for the appropriate contract-norm match at both cultural-institutional nexus, a shift from autarchy to free trade will increase the costs of deviating and hence will delay a transition away from the Pareto-inferior convention. Trade liberalization, thus, does not favor cultural-institutional convergence to superior configurations.

Fourth, because transitions require extreme realizations of the sum of deviations relative to population size, for sufficiently rational individuals, the expected waiting time for a transition is increasing in the group size (z). Extending the model to allow the "upper" B class to be less numerous ($z_A > z_B$), most of the transitions will be induced by the innovations by members of the elite. But this unsurprising result – history tends to be driven by the elite – is unrelated to the fact that smaller groups may more readily coordinate their actions in producing the public good represented by a transition in which their members do better. The result occurs because the extreme realization of the number of innovators required to induce a transition are more likely the smaller is the population size.

From the above result it follows that, by relaxing the ultra-individualism of the model and allowing for collective action, the expected waiting time for a transition will vary with the degree to which a society is "individualist" or "collectivist" – in the terms of Avner Greif (1994). We use these terms to mean that in an individualist society the action of one person does not affect the other individuals' actions unless the action alters the incentives facing the others. For example, if one member of a family deviates from the status quo, this has no effect on other family members' actions unless it changes their expected payoffs. By contrast, in a collectivist society individuals sometimes act in groups, such that if one brother deviates all of the siblings will also deviate. In a collectivist society the effective population size is less than the census size and the expected waiting time for a transition is correspondingly reduced. To provide a simple illustration, suppose in our model that employees (As) work in firms of size n_A and that all employees in given firm either conform to the status quo (best respond) or they jointly deviate. Effective population size in this case is not z_A but instead z_A/n_A. Worker-induced transitions will be correspondingly accelerated.

References

Acemoglu, Daron. 2003. Why Not a Political Coase Theorem? Social Conflict, Commitment and Politics. *Journal of Comparative Economics*, 31, 620-52.

Acemoglu, Daron, and James Robinson. 2012. *Why nations fail: The origins of power, prosperity and poverty*: Crown.

Belloc, Marianna, and Samuel Bowles. 2013. Cultural-Institutional Persistence under Autarchy, International Trade, and Factor Mobility. *Santa Fe Institute Working Paper*, 13-01-003.

Binmore, Ken, Larry Samuelson, and Peyton Young. 2003. Equilibrium Selection in Bargaining Models. *Games and Economic Behavior*, 45, 296-328.

Blume, Lawrence. 2003. How Noise Matters. *Games and Economic Behavior*, 44, 251-71.

Clark, Gregory. 2007. *A farewell to alms.* Princeton: Princeton University Press.

Edgerton, Robert B. 1992. *Sick societies.* New York: The Free Press.

Greif, Avner. 1994. Cultural Beliefs and the Organization of Society: An Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy*, 102(5), 912-50.

Jurgens, Ulrich. 2002. Corporate Governance, Innovation and Economic Performance. *ZBW*, Berlin.

Nuti, D. Mario. 2001. Industrial Relations at FIAT: Dr. Marchionne's Class War. *Transfer: European Review of Labour and Research*, 17, 251-54.

Parsons, Talcott. 1964. Evolutionary Universals in Society. *American Sociological Review*, 29(3), 339-57.

Sokoloff, Kenneth, and Stanley Engerman. 2000. Institutions, Factor Endowments, and Paths of Development in the New World. *Journal of Economic Perspectives*, 14(3), 217-32.

Williamson, Oliver 1985. *The economic institutions of capitalism.* New York: Free Press.

Wood, Elisabeth. 2000. *Forging democracy from below.* Cambridge: Cambridge University Press.

Young, Peyton. 1998. *Individual strategy and social structure: an evolutionary theory of institutions.* Princeton: Princeton University Press.

Appendix

A. Why intergenerational transmission of cultural traits is insufficient to explain cultural persistence

Intergenerational transmission of cultural traits cannot explain the historical evidence for long term persistence of cultures (surveyed in Spolaore and Wacziang, 2013) for two reasons: (A.1) the empirically measured degree of transmission for most traits is quite modest; (A.2) even where transmission is substantial, cultural differences are dissipated over just a few generations if there are no other mechanisms favoring persistence.

A.1. Vertical transmission from parents, either genetic or cultural, appears to be substantial for some traits – those relating to political values and religion in particular – but measures of parent offspring similarity for most traits relevant to the study of culture seem to be quite limited. The surprising weakness of vertical transmission from parents may, of course, reflect the difficulty in measuring preferences or other cultural traits rather than the absence of underlying effects; but the available data does not support the inference of strong transmission.

A meta-study of parent-child transmission of the so-called big five personality traits yields a mean correlation of 0.13 (Loehlin, 2005). Even for cognitive traits in which genetic transmission plays a major role, the parent offspring correlations are quite modest (for example 0.38 for IQ (Black, et al., 2009)). Feldman, et al. (1982) found parent offspring correlations of 0.69 for religion and 0.48 for a measure of political values, with average correlations including these and other cultural traits (concerning beliefs, tastes in entertainment, etc.) of 0.35. Nowak's sample from Poland in the early 1970s (Nowak, 1981) found no significant correlation between the values of parents and those of their (grown) children excepting religion. Kohn (1983: 3) concludes "that the relationships between parents and children's values are probably only modest in magnitude." Surveying a number of studies, he writes "studies of values have consistently found rather modest levels of agreement – correlations of roughly 0.15 to 0.25 – between parents and children." Feldman, Cavalli-Sforza, Dornbusch, et al. (1982) find, in a Taiwanese sample, strong vertical transmission (parental) for religion and political beliefs but not for other traits (preferences for films, recreation). Rozin (1991) likewise finds evidence of the direct transmission of values (about homosexuality, abortion, religiosity, and other values): mid-parent offspring correlations averaged 0.54.

Food tastes differ considerably between cultural groups, a major predictor of food likes and dislikes being one's nation or ethnic group of origin; yet parental tastes are poor predictors of the tastes of offspring, even when parental tastes are themselves congruent. The mean correlation for mid-parent-child tastes for particular foods (black coffee, lima beans, hot sauce, and so on) in the study by Rozin (1991) is 0.17. He calls this the "family paradox" concerning "the sources of variance in preferences. Genetic factors and family influence account for a very small part" (Rozin, 1991: 101). Because cultural differences in tastes are maintained over long periods of time, it seems clear that a significant part of the transmission process is taking place at the societal level, with important roles played by either horizontal transmission (from age peers) or oblique transmission (from non parents in the previous generation).

A.2. Consider two populations. In each it is widely thought that some animal is sacred; but the cultures differ in which the sacred animal is: in one society it is forbidden to eat rabbits, in the other eating deer is taboo. Parents pass on their ideas about sacredness of animals to their offspring, but the process is imperfect, so that with probability r the child will adopt the norm of their parents while with probability $1-r$ the other norm is adopted. For plausible values of the transmission coefficient, r, the cultures of these two populations will become indistinguishable in just a few centuries, or less.

Suppose the cultural aversion to eating rabbits or deer is very strongly transmitted directly from parents to offspring and that initially virtually all of the members of the two societies have adopted their respective norms, and to bias the example to favor persistence suppose further that marital assortment is complete, so that parents always share the same trait. Let p_t^k be the fraction of each population ($k = 1, 2$) sharing the respective norm (either rabbit taboo or deer taboo) at time t. This fraction in the next generation, $t + 1$, becomes $p_{t+1}^k = rp_t^k + (1-r)(1-p_t^k)$. Assume that at time $t = 0$ in each of the two populations $p_0^k = 0.9$ and $1 - p_0^k = 0.1$. Then transmission coefficients $r = 0.7$ and $r = 0.85$ generate a parent-offspring correlation of about a quarter and a half, respectively (see, e.g., Cavalli-Sforza and Feldman, 1981).

Figure A1 illustrates that the two populations would be barely distinguishable after respectively 10 generations and 5 generations.

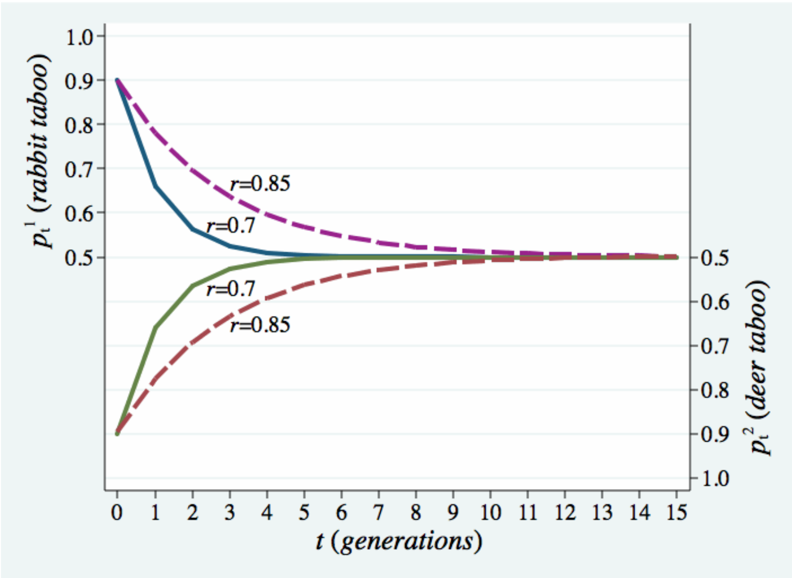

Figure A1. *Cultural convergence with empirically plausible levels of intergenerational transmission*

A.3. References

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2009. Like Father Like Son? A Note on the Intergenerational Transmission of IQ Scores. *Economic Letters*, 105(1), 138-140.

Cavalli-Sforza, Luigi L., and Marcus W. Feldman. 1981. *Cultural transmission and evolution: a quantitative approach*. Princeton: Princeton University Press, chapter 2.

Feldman, Marcus W., Luigi L. Cavalli-Sforza, Stanford M. Dornbusch, and Kuang-Ho Chen. 1982. Theory and Observation in Cultural Transmission. *Science*, 218, 19-27.

Kohn, Melvin L. 1983. On the Transmission of Values in the Family: A Preliminary Formulation. *Research in Sociology of Education and Socialization*, 4, 3-12.

Loehlin, John. 2005. Resemblance in Personality and Attitudes Between Parents and their Children: Genetic and Environmental Contributions, in *Unequal chances: family background and economic success*, eds. Samuel Bowles, Herbert Gintis and Melissa Osborne Groves. Princeton: Princeton University Press, 192-207.

Nowak, Stefan. 1981. Values and Attitudes of the Polish People. *Scientific American*, 245(1), 45-53.

Rozin, Paul. 1991. Family Resemblance in Food and Other Domains: The Family Paradox and the Role of Parental Congruence. *Appetite*, 16, 93-102.

Spolaore, Enrico, and Romain Wacziarg. 2013. How Deep are the Roots of Economic Development? *Journal of Economic Literature*, forthcoming

B. Mathematical appendix

B.1. First result: *For sufficiently rational agents (large β), the expected waiting time for a transition ($E[W]$) is decreasing in the superiority of the Pareto-dominant convention (Δ_1^i).*

Using equation (2) in the text, we have (with $i = A, B$):

$$\lim_{\beta\to\infty} \frac{P_i'(\Delta_1^{i\prime},\beta)}{P_i(\Delta_1^i,\beta)} = \frac{z!/[(z\phi_i^{*\prime})!(z-z\phi_i^{*\prime})!]}{z!/[(z\phi_i^*)!(z-z\phi_i^*)!]} \lim_{\beta\to\infty} \frac{[\sigma_i(\beta)]^{z\phi_i^{*\prime}}}{[\sigma_i(\beta)]^{z\phi_i^*}} \frac{[1-\sigma_i(\beta)]^{z-z\phi_i^{*\prime}}}{[1-\sigma_i(\beta)]^{z-z\phi_i^*}}$$

where $\phi_i^* = \Delta_0^i/(\Delta_0^i + \Delta_1^i) > \phi_i^{*\prime} = \Delta_0^i/(\Delta_0^i + \Delta_1^{i\prime})$ with $\Delta_1^{i\prime} > \Delta_1^i$ and $\sigma_i(\beta) = 1/(1+e^{\beta\Delta_0^i})$. Omitting the constant term and using (1) in the text, it follows:

$$\lim_{\beta\to\infty} \frac{[1/(1+e^{\beta\Delta_0^i})]^{z\phi_i^{*\prime}}}{[1/(1+e^{\beta\Delta_0^i})]^{z\phi_i^*}} \frac{[1-1/(1+e^{\beta\Delta_0^i})]^{z-z\phi_i^{*\prime}}}{[1-1/(1+e^{\beta\Delta_0^i})]^{z-z\phi_i^*}} = \lim_{\beta\to\infty} \frac{(1+e^{\beta\Delta_0^i})^{z\phi_i^*}}{(1+e^{\beta\Delta_0^i})^{z\phi_i^{*\prime}}}$$

where we have used the fact that $\lim_{\beta\to\infty}[1-1/(1+e^{\beta\Delta_0^i})] = 1$. After defining $y \equiv e^{\beta\Delta_0^i}$ and for finite z, we obtain

$$\lim_{\beta\to\infty} \frac{P_i'(\Delta_1^{i\prime},\beta)}{P_i(\Delta_1^i,\beta)} = \lim_{y\to\infty} \left[\frac{y^{\phi_i^*}}{y^{\phi_i^{*\prime}}}\right]^z = \infty.$$

Hence, there exists $\bar{\beta}$ such that for $\beta > \bar{\beta}$ it must be that $P_i'(\Delta_1^{i\prime},\beta) > P_i(\Delta_1^i,\beta)$ with $\Delta_1^{i\prime} > \Delta_1^i$ and $i = A, B$. From equation (3) in the text it thus follows that $E[W'] = (P_A' + P_B' - P_A' \times P_B')^{-1} < E[W] = (P_A + P_B - P_A \times P_B)^{-1}$.

B.2. Second result: *The expected waiting time for a transition ($E[W]$) is increasing in the degree of individual rationality (β).*

Using equations (1) and (2) in the text, we obtain (with $i = A, B$):

$$\frac{dP_i[\sigma_i(\beta)]}{d\beta} = \frac{dP_i[\sigma_i(\beta)]}{d\sigma_i(\beta)} \frac{d\sigma_i(\beta)}{d\beta} =$$

$$= \binom{z}{z\phi_i^*} {\sigma_i}^{z\phi_i^*}(1-\sigma_i)^{z-z\phi_i^*}[z\phi_i^*\sigma_i^{-1} - (z-z\phi_i^*)(1-\sigma_i)^{-1}][-\Delta_0^i e^{\beta\Delta_0^i}/(1+e^{\beta\Delta_0^i})^2]$$

which is negative if and only if $\phi_i^* - \sigma_i > 0$, which is always the case for sufficiently large β. Hence from (3) in the text it follows that $E[W] = (P_A + P_B - P_A \times P_B)^{-1}$ is increasing in β.

B.3. Third result: *For sufficiently rational agents, the expected waiting time for a transition ($E[W]$) is increasing in the cost of deviation from the inferior convention (Δ_0^i).*

Using equation (2) in the text, we have (with $i = A, B$)

$$\lim_{\beta\to\infty} \frac{P_i'(\Delta_0^{i\prime},\beta)}{P_i(\Delta_0^i,\beta)} = \frac{z!/[(z\phi_i^{*\prime})!(z-z\phi_i^{*\prime})!]}{z!/[(z\phi_i^*)!(z-z\phi_i^*)!]} \lim_{\beta\to\infty} \frac{[\sigma_i'(\beta)]^{z\phi_i^{*\prime}}}{[\sigma_i(\beta)]^{z\phi_i^*}} \frac{[1-\sigma_i'(\beta)]^{z-z\phi_i^{*\prime}}}{[1-\sigma_i(\beta)]^{z-z\phi_i^*}}$$

where $\phi_i^{*\prime} = \Delta_0^{i\prime}/(\Delta_0^{i\prime} + \Delta_1^i) > \phi_i^* = \Delta_0^i/(\Delta_0^i + \Delta_1^i)$ and $\sigma_i(\beta) = 1/(1+e^{\beta\Delta_0^i}) > \sigma_i'(\beta) = 1/(1+e^{\beta\Delta_0^{i\prime}})$ with $\Delta_0^{i\prime} > \Delta_0^i$. Omitting the constant term, we can write:

$$\lim_{\beta\to\infty} \frac{[1/(1+e^{\beta\Delta_0^{i\prime}})]^{z\phi_i^{*\prime}}}{[1/(1+e^{\beta\Delta_0^i})]^{z\phi_i^*}} \frac{[1-1/(1+e^{\beta\Delta_0^{i\prime}})]^{z-z\phi_i^{*\prime}}}{[1-1/(1+e^{\beta\Delta_0^i})]^{z-z\phi_i^*}} = \lim_{\beta\to\infty} \frac{(1+e^{\beta\Delta_0^i})^{z\phi_i^*}}{(1+e^{\beta\Delta_0^{i\prime}})^{z\phi_i^{*\prime}}}.$$

After defining $y \equiv e^{\beta}$ and for finite z, we obtain

$$\lim_{\beta\to\infty} \frac{P_i'(\Delta_0^{i\prime},\beta)}{P_i(\Delta_0^i,\beta)} = \lim_{y\to\infty} \left[\frac{y^{\Delta_0^i \phi_i^*}}{y^{\Delta_0^{i\prime} \phi_i^{*\prime}}}\right]^z = 0,$$

because $\Delta_0^i \phi_i^* < \Delta_0^{i\prime} \phi_i^{*\prime}$. Hence, there exists $\bar{\beta}$ such that for $\beta > \bar{\beta}$ it must be that $P_i'(\Delta_0^{i\prime},\beta) < P_i(\Delta_0^i,\beta)$ with $\Delta_0^{i\prime} > \Delta_0^i$ and $i = A, B$. From equation (3) in the text it thus follows that $E[W'] = (P_A' + P_B' - P_A' \times P_B')^{-1} > E[W] = (P_A + P_B - P_A \times P_B)^{-1}$.

B.4. Fourth result: *For sufficiently rational agents, the expected waiting time for a transition ($E[W]$) is increasing in the group size (z).*

Using equation (2) in the text, we have (with $i = A, B$):

$$\lim_{\beta\to\infty} \frac{P_i'(z',\beta)}{P_i(z,\beta)} = \frac{z'!/[(z'\phi_i^*)!(z'-z'\phi_i^*)!]}{z!/[(z\phi_i^*)!(z-z\phi_i^*)!]} \lim_{\beta\to\infty} \frac{[\sigma_i(\beta)]^{z'\phi_i^{*\prime}}}{[\sigma_i(\beta)]^{z\phi_i^*}} \frac{[1-\sigma_i(\beta)]^{z'-z'\phi_i^{*\prime}}}{[1-\sigma_i(\beta)]^{z-z\phi_i^*}}$$

with $z' > z$. Omitting the constant term and using equation (1) in the text, we can write:

$$\lim_{\beta\to\infty} \frac{[1/(1+e^{\beta\Delta_0^i})]^{z'\phi_i^*}}{[1/(1+e^{\beta\Delta_0^i})]^{z\phi_i^*}} \frac{[1-1/(1+e^{\beta\Delta_0^i})]^{z'-z'\phi_i^*}}{[1-1/(1+e^{\beta\Delta_0^i})]^{z-z\phi_i^*}} = \lim_{\beta\to\infty} \frac{(1+e^{\beta\Delta_0^i})^{z\phi_i^*}}{(1+e^{\beta\Delta_0^i})^{z'\phi_i^*}}.$$

After defining $y \equiv e^{\beta\Delta_0^i\phi_i^*}$ and for finite z, we obtain

$$\lim_{\beta\to\infty} \frac{P_i'(z',\beta)}{P_i(z,\beta)} = \lim_{y\to\infty} \frac{y^z}{y^{z'}} = 0.$$

Hence, there exists $\bar{\beta}$ such that for $\beta > \bar{\beta}$ it must be that $P_i'(z',\beta) < P_i(\Delta_0^i,\beta)$ with $z' > z$ and $i = A, B$. From equation (3) in the text it thus follows that $E[W] = (P_A' + P_B' - P_A' \times P_B')^{-1} > E[W'] = (P_A + P_B - P_A \times P_B)^{-1}$.

C. Additional sources

A key next step in the approach outlined here is to model idiosyncratic play as an intentional innovation that may be coordinated by leadership and network structure:

Acemoglu, Daron, and Matthew Jackson. 2011. History, Expectations and Leadership in the Evolution of Social Norms. *NBER Working Paper*, 17066.

Bowles, Samuel. 2004. *Microeconomics: behavior, institutions, and evolution.* Princeton: Princeton University Press, chapter 12.

Hwang, Sung Ha, Suresh Naidu, and Samuel Bowles. 2013. Persistent Institutions. *Santa Fe Institute.*

Kets, Willemien, Garud Iyengar, Rajiv Sethi, and Samuel Bowles. 2011. Inequality and Network Structure. *Games and Economic Behavior*, 73, 215-26.

Young, H. Peyton. 2003. The Diffusion of Innovations in Social Networks, in *The economy as a complex evolving system*, eds. Lawrence Blume and Steven Durlauf. Oxford: Oxford University Press.

Our representation of the joint dynamics of cultures and institutions has borrowed from the working group on the Co-evolution of Behaviors and Institutions at the Santa Fe Institute and on the literature on cultural evolution initiated by:

Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the evolutionary process.* Chicago: University of Chicago Press.

Cavalli-Sforza, Luigi L., and Marcus W. Feldman. 1981. *Cultural transmission and evolution: a quantitative approach.* Princeton: Princeton University Press.

On why inefficient institutions persist:

Bowles, Samuel. 2004. *Microeconomics: behavior, institutions, and evolution.* Princeton: Princeton University Press, chapters 1 and 2.

Krusell, Per, and Jose-Victor Rios-Rull. 1996. Vested Interests in a Theory of Stagnation and Growth. *Review of Economic Studies*, 63, 301-330.

Rosenthal, Jean-Laurent. 1998. The Political Economy of Absolutism Reconsidered, in *Analytic narratives*, eds. Robert H. Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. Princeton: Princeton University Press, 64-108.

Historical evidence on collective actions deviating from status quo cultural-institutional conventions:

Lohmann, Susanne. 1994. Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig. *World Politics*, 47(1), 42-101.

Markoff, John. 1996. *The abolition of feudalism: peasants, lords and legislators in the French revolution.* University Park: Pennsylvania State University Press.

McAdam, Doug. 1983. Tactical Innovation and the Pace of Insurgency. *American Sociological Review*, 48(6), 735-54.