

The persistence of inferior cultural-institutional conventions[§]

Marianna Belloc* and Samuel Bowles**

Preliminary draft - December 30, 2012

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 are endogenous. We explain why Pareto-dominated cultural-institutional configurations may persist over long periods and how transitions may nonetheless occur. Unlike models in which elites may 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 Codes: C73, D02, D03

[§] For presentation at the AEA Annual Meeting (Session on Culture, Institutions, and Historical Persistence), San Diego, January 2013. Thanks to the Santa Fe Institute Behavioral Sciences program for support of this work and to Sung-Ha Hwang for comments. An online appendix provides derivations and proofs of the main results, and additional citations to the literature (<http://www.santafe.edu/~bowles>).

* 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 (NM), 87501, USA. Email: bowles@santafe.edu

I. 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.” But economists have had little success in providing the mechanisms that would account for this felicitous result. Evidence of socially dysfunctional but enduring cultural practices suggests that invisible hand arguments work no better when applied to cultures than to institutions (Edgerton (1992)).

II. THE PERSISTENCE OF INFERIOR CULTURAL-INSTITUTIONAL CONVENTIONS

Given that culture and institutions are often implicated in explaining enduring poverty among nations (e.g. Clark (2007), Acemoglu and Robinson (2012), Nunn (2012)), a further question has attracted economists: 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 of culture and institutions? “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. Moreover, in light of recent history, persistence cannot be an intrinsic characteristic of either culture or institutions. Examples include the precipitous demise of such heavily defended institutions as Communist Party rule in many parts of the world 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 the distribution 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), Krusell and Rios-Rull (1996)). 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 (Acemoglu, *et al.* (2012)). A key advantage of these elite-based explanations is that they provide both a motive – self-interest – and a mechanism – concentrated but not absolute political power – for the persistence of Pareto-dominated cultural-institutional conventions.

But there is another way that inferior cultural-institutional conventions may persist indefinitely. 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. 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 culture and institutions 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 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 conventions 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 actions on the basis of past play 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)).

We measure the impediments to transitions to a superior convention by the expected waiting time for a transition to take place. We show (unsurprisingly) that the expected waiting time is less the greater is the superiority of the alternative convention, but that for sufficiently rational agents, inferior conventions can persist indefinitely. The expected waiting time is greater if people are more rational in that they more closely approximate expected payoff maximization in their behavior. Transitions to superior conventions will also be less likely if a society is more individualistic in that actions by a given individual have little impact on actions of others unless the relevant material incentives are affected.

III. A BOTTOM-UP MODEL OF CULTURAL-INSTITUTIONAL PERSISTENCE AND INNOVATION

We study two classes, large sub-populations whose members interact economically with each other in ways governed by a set of institutions. These populations 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. We refer to the economic institutions governing relations between the two classes as contracts, which are implemented by the actions that members of the second class in each of the pairs just mentioned (who we call the *Bs*). The actions taken by the first class (the *As*) are the adoption of 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 two distinct contract-norm matches as 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 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; Volkswagens workers would not benefit by adopting a militant oppositional culture; and so on.

For simplicity we study the evolutionary dynamics governing transitions between just two conventions in a 2x2 game in which the z members of each class are randomly paired with members of the other class to interact in a non-cooperative game. Following matching, each has a type revision opportunity. To ensure that the process is acyclic, we let the revision process be asynchronous (as in Binmore, Samuelson and Young, 2003). When revising their contracts and norms, best responding individuals maximize their expected payoffs based on the distribution of the population in the previous period. We index contracts and norms by $j=0,1$ and groups by $i=A,B$, and represent the payoffs to the four possible cultural-institutional matches as $\pi_{11}^i > \pi_{00}^i > \pi_{01}^i = \pi_{10}^i = 0$ for both classes, where, for example, π_{10}^i is the payoff of an individual in class i implementing a 1-contract (adopting a 1-norm) 1 when his partner adopts a 0-norm (implements a 0-contract), and payoffs are normalized so that the two "mismatches" are zero. Expected payoffs are given by $v_j^i = (1 - \phi_{-i})\pi_{j0}^i + \phi_{-i}\pi_{j1}^i$, where ϕ denotes the fraction of individuals who adopted 1-contracts (1-norms) in the previous period and $-i$ refers the other group. Expected payoffs are shown in Figure 1.

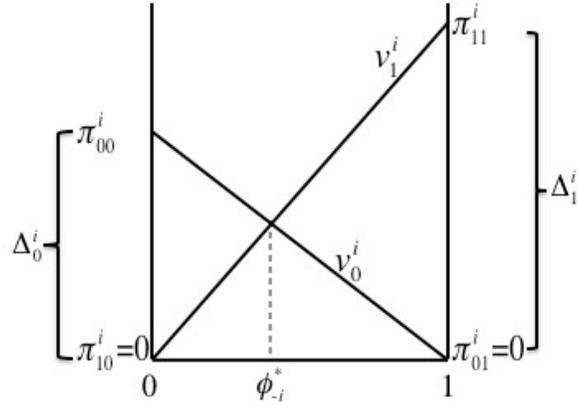


Figure 1: Expected payoffs.

The state space for this process (shown in Figure 2) is all possible combinations of the number of 1-contract and 1-norm individuals in the two classes, $(z\phi_A, z\phi_B)$. 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 play one or the other of these two profiles are absorbing when idiosyncratic (non-best response) play is absent. 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), we have the following expression for the probability of deviating from the best response when the population is at $\{0,0\}$:

$$\sigma(\Delta_0^i, \beta) = \frac{1}{1 + e^{\beta\Delta_0^i}}, \text{ with } i = A, B, \quad (1)$$

where $\Delta_0^i = \pi_{00}^i - \pi_{10}^i$ is the (positive and finite) cost of deviation from the status quo culture or institution. We interpret β as a measure of rationality because the larger is β , the smaller the probability that she or he will deviate from the best response. When $\beta = 0$ the agent chooses randomly between the two options, and as β goes to infinity, the individual never deviates. Of

course, individuals may have non-economic reasons to deviate from the status quo even in the absence of cognitive failures; so, strictly speaking, β measures the degree to which agents maximize their expected payoffs.

Under this dynamic $\{z,z\}$ is the stochastically stable state (because it is risk dominant). But 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 waiting time for a realization of sufficient non-best response play to tip the process from the neighborhood of one convention to the basin of attraction of the other will be very prolonged.

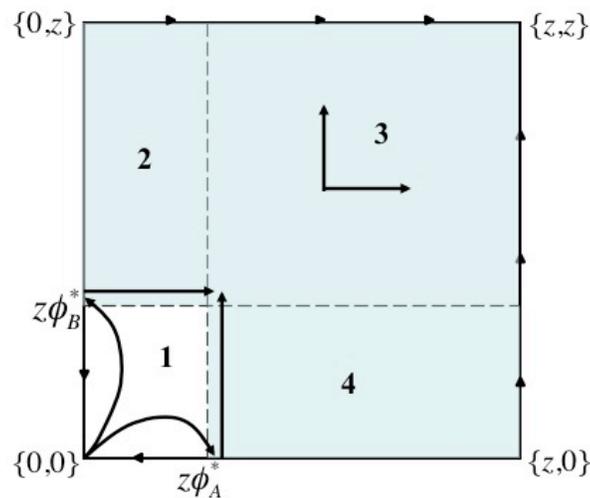


Figure 2: Transitions from $\{0,0\}$ to $\{z,z\}$.

IV. IMPEDIMENTS TO PARETO-IMPROVING CULTURAL AND INSTITUTIONAL CHANGE

To study transitions from the inferior $\{0,0\}$ to the superior $\{z,z\}$ convention, 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, the state following idiosyncratic play will almost always lie within the basin of attraction of the status quo convention and, because agents have only a one-period memory, the excursion from the convention will have no lasting effects.

But suppose that, from the initial state $\{0,0\}$, $z\phi_B^*$ of the B s play idiosyncratically and offer 1-contracts instead of best responding with a 0-contract, where $z\phi_B^*$ is the smallest number such that the A s' best response is to switch to the 1-norm and $\phi_B^* = \Delta_0^A / (\Delta_0^A + \Delta_1^A)$. (Since z is large we may avoid notational clutter by abstracting from integer considerations.) In response, each A will adopt 1-norms with probability $1-\sigma$. But as β goes to infinity σ goes to zero, so there exists some finite $\underline{\beta}$ such that for $\beta > \underline{\beta}$, as a result of the A s' updating, with virtual certainty we will have $z\phi_A \geq z\phi_A^*$, where $z\phi_A^*$ is the smallest number such that the B s best response is to switch to the 1-contract (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 1-contracting B s sufficient to escape from the inferior equilibrium is $z\phi_B^*$. Analogous reasoning applies to the minimum number of deviant A -players, $z\phi_A^*$, sufficient to induce a transition. It follows that from the initial state $\{0,0\}$ the basin of attraction of the superior equilibrium is composed of quadrants 2, 3 and 4 in Figure 2.

Because deviations from the best response contract or social norm are independent, waiting times for a transition from one stationary state to the other are approximated by the inverse of the probability that in a given period the number of deviants will be sufficient to enter the basin of attraction of the other equilibrium, which, for large populations and sufficient rationality, is approximated by the probability, μ_i , that exactly the minimum number of deviants in class i ($z\phi_i^*$, with $i=A,B$) will occur (Binmore, *et al.* (2003); our results to follow would not be affected by taking account of the probability that larger than minimal numbers deviate.):

$$\mu_i = \left(\begin{array}{c} z \\ z\phi_i^* \end{array} \right) (\sigma_i)^{z\phi_i^*} (1 - \sigma_i)^{z - z\phi_i^*}, \text{ with } i = A, B. \quad (2)$$

V. DISCUSSION

Equation (2) gives us four results. First, “cultural and 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 the critical fractions required for a transition from $\{0,0\}$ to $\{1,1\}$ – 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 ($1/\mu_i$) is decreasing in the superiority of the Pareto-dominant convention.

Second, because deviations from the status quo are less likely the greater 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.

Third, the greater is the cost of deviating from the inferior convention (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 (which we demonstrate in Belloc and Bowles (2012)): because the gains from trade increase the payoffs for the contract-norm match at both conventions, 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.

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 (that is $z_A > z_B$), most of the transitions will be induced by the idiosyncratic play of members of the elite. But this unsurprising result – history tends to be driven by elite actions – 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.

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 actions of others 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 waiting time for a transition is correspondingly reduced.

A degree of collectivism (in this sense) is suggested by the high levels of spatially and temporally correlated actions observed, for example in series of oppositional acts during the French Revolution, the fall of Communism, the U.S. civil rights movement, or the end of apartheid (Markoff (1996), Lohmann (1994), McAdam (1983), Bowles, *et al.* (2012).) To provide a simple illustration, suppose in our model that employees 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.

Accounting for the possibility of collective action in this model, of course, requires both a more explicit representation of the political process and more complex network structures than the bipartite graph implicit in this employer-worker case, perhaps along the lines suggested in Acemoglu and Jackson (2011), Bowles (2004), Kets, *et al.* (2011) and Young (2003). Because technical change may affect equilibrium network structures, it may also alter effective population sizes and the transition process. For example, an economy made up of a few large

industrial firms (e.g. St. Petersburg in 1917) may provide an environment more favorable to transitions induced by non-elites than an economy of sharecroppers.

VI. REFERENCES

- Acemoglu, Daron. 2003. "Why not a political Coase theorem? Social conflict, commitment and politics." *Journal of Comparative Economics*, 31, pp. 620-52.
- Acemoglu, Daron, Gregory Egorov, and Konstantin Sonin. 2012. "Dynamics and Stability of Constitutions, Coalitions, and Clubs." *American Economic Review*, 102, pp. 1446-76.
- Acemoglu, Daron and Matthew Jackson. 2011. "History, expectations and leadership in the evolution of social norms." *NBER Working Paper No. 17066*.
- Acemoglu, Daron and James Robinson. 2012. *Why nations fail: The origins of power, prosperity and poverty*: Crown.
- Bello, Marianna and Samuel Bowles. 2012. "International Trade and the Persistence of Cultural-Institutional Diversity." *Santa Fe Institute Working Paper*.
- Binmore, Ken, Larry Samuelson, and H. Peyton Young. 2003. "Equilibrium selection in bargaining models." *Games and Economic Behavior*, 45, pp. 296-328.
- Blume, Lawrence. 2003. "How noise matters." *Games and Economic Behavior*, 44, pp. 251-71.
- Bowles, Samuel. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press.
- Bowles, Samuel, Suresh Naidu, and Sung Ha Hwang. 2012. "Persistent institutions." *Santa Fe Institute*.
- Clark, Gregory. 2007. *A Farewell to Alms: A Brief Economic History of the World*. Princeton: Princeton University Press.

- Edgerton, Robert B. 1992. *Sick Societies: Challenging the Myth of Primitive Harmony*. 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, pp. 912-50.
- Jurgens, Ulrich. 2002. "Corporate governance, innovation and economic performance: a case study on Volkswagen." Open access publication server of the ZBW: Berlin.
- Kets, Willemien, Garud Iyengar, Rajiv Sethi, and Samuel Bowles. 2011. "Inequality and Network Structure." *Games and Economic Behavior*, 73, pp. 215-26.
- Krusell, Per and Jose-Victor Rios-Rull. 1996. "Vested interests in a theory of stagnation and growth." *Review of Economic Studies*, 63:301-330.
- Lohmann, Susanne. 1994. "Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig." *World Politics*, 47:1, pp. 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, pp. 735-54.
- Nuti, D. Mario. 2001. "Industrial relations at FIAT: Dr. Marchionne's class war." *Transfer: European Review of Labour and Research*, 17, pp. 251-54.
- Parsons, Talcott. 1964. "Evolutionary Universals in Society." *American Sociological Review*, 29:3, pp. 339-57.
- Sokoloff, K. and S. Engerman. 2000. "Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic Perspectives*, 14:3, pp. 217-32.
- Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

Wood, Elisabeth Jean. 2000. *Forging Democracy From Below: Insurgent Transitions in South Africa and El Salvador*. Cambridge: Cambridge University Press.

Young, H. Peyton. 2003. "The diffusion of innovations in social networks," in *The Economy as a Complex Evolving System*. Lawrence Blume and Steven Durlauf eds. Oxford: Oxford University Press.

MATHEMATICAL APPENDIX

A.1. First result: for sufficiently rational agents, the expected waiting time for a transition ($1/\mu$) is decreasing in the superiority of the Pareto-dominant convention (Δ_1^i)

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

$$\lim_{\beta \rightarrow \infty} \frac{\mu_i(\Delta_1^{i'}, \beta)}{\mu_i(\Delta_1^i, \beta)} = \frac{\frac{z!}{(z\phi_i^*)!(z-z\phi_i^*)!}}{\frac{z!}{(z\phi_i^i)!(z-z\phi_i^i)!}} \times \lim_{\beta \rightarrow \infty} \frac{\sigma(\Delta_1^{i'}, \beta)^{z\phi_i^{*'}} [1 - \sigma(\Delta_1^i, \beta)]^{z-z\phi_i^{*'}}}{\sigma(\Delta_1^i, \beta)^{z\phi_i^i} [1 - \sigma(\Delta_1^i, \beta)]^{z-z\phi_i^i}}$$

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

$$\lim_{\beta \rightarrow \infty} \frac{\left(\frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z\phi_i^{*'}} \left(1 - \frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z-z\phi_i^{*'}}}{\left(\frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z\phi_i^i} \left(1 - \frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z-z\phi_i^i}} = \lim_{\beta \rightarrow \infty} \frac{(1 + e^{\beta\Delta_0^i})^{z\phi_i^i}}{(1 + e^{\beta\Delta_0^i})^{z\phi_i^{*'}}$$

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

$$\lim_{\beta \rightarrow \infty} \frac{\mu_i(\Delta_1^{i'}, \beta)}{\mu_i(\Delta_1^i, \beta)} = \lim_{y \rightarrow \infty} \frac{y^{\phi_i^i}}{y^{\phi_i^{*'}} = \infty.$$

Hence, there exists $\bar{\beta}$ such that for $\beta > \bar{\beta}$, it must be $\mu_i(\Delta_1^{i'}) > \mu_i(\Delta_1^i)$ with $\Delta_1^{i'} > \Delta_1^i$.

A.2. Second result: the expected waiting time for a transition ($1/\mu$) is increasing in the degree of individual rationality (β)

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

$$\frac{d\mu_i(\sigma)}{d\sigma} = z\phi_i^* \sigma^{z\phi_i^*-1} (1 - \sigma)^{z\phi_i^*} - (z - z\phi_i^*) (1 - \sigma)^{z-z\phi_i^*-1} \sigma^{z\phi_i^*}$$

that can be written as

$$\begin{aligned}\frac{d\mu_i(\sigma)}{d\sigma} &= z\sigma(\beta)^{z\phi_i^*} (1-\sigma)^{z\phi_i^*} [\phi_i^* \sigma(\beta)^{-1} - (1-\phi_i^*)(1-\sigma)^{-1}] \\ &= \frac{\phi_i^*(1-\sigma) - (1-\phi_i^*)\sigma}{\sigma(1-\sigma)} = \frac{\phi_i^* - \sigma(\beta)}{\sigma(\beta)[1-\sigma(\beta)]}\end{aligned}$$

that is positive iff $\phi_i^* - \sigma > 0$ which is always the case for small σ .¹ Hence, since $\sigma(\beta) = 1/(1 + e^{\beta\Delta_0^i})$ is decreasing in β , it follows that $\mu_i(\beta') < \mu_i(\beta)$ with $\beta' > \beta$.

A.3. Third result: for sufficiently rational agents, the expected waiting time for a transition ($1/\mu$) is increasing in the cost of deviating from the inferior convention (Δ_0^i)

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

$$\lim_{\beta \rightarrow \infty} \frac{\mu_i(\Delta_0^{i'}, \beta)}{\mu_i(\Delta_0^i, \beta)} = \frac{\frac{z!}{(z\phi_i^{*'})!(z-z\phi_i^{*'})!}}{\frac{z!}{(z\phi_i^*)!(z-z\phi_i^*)!}} \times \lim_{\beta \rightarrow \infty} \frac{[\sigma(\Delta_0^{i'}, \beta)]^{z\phi_i^{*'}} [1-\sigma(\Delta_0^{i'}, \beta)]^{z-z\phi_i^{*'}}}{[\sigma(\Delta_0^i, \beta)]^{z\phi_i^*} [1-\sigma(\Delta_0^i, \beta)]^{z-z\phi_i^*}},$$

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

$$\lim_{\beta \rightarrow \infty} \frac{\left(\frac{1}{1+e^{\beta\Delta_0^{i'}}}\right)^{z\phi_i^{*' }} \left(1 - \frac{1}{1+e^{\beta\Delta_0^{i'}}}\right)^{z-z\phi_i^{*' }}}{\left(\frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z\phi_i^*} \left(1 - \frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z-z\phi_i^*}} = \lim_{\beta \rightarrow \infty} \frac{(1 + e^{\beta\Delta_0^i})^{z\phi_i^*}}{(1 + e^{\beta\Delta_0^{i'}})^{z\phi_i^{*' }}}$$

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

$$\lim_{\beta \rightarrow \infty} \frac{\mu_i(\Delta_0^{i'}, \beta)}{\mu_i(\Delta_0^i, \beta)} = \lim_{y \rightarrow \infty} \frac{[y^{\Delta_0^i}]^{z\phi_i^*}}{[y^{\Delta_0^{i'}}]^{z\phi_i^{*' }}} = 0$$

because $\Delta_0^i \phi_i^* < \Delta_0^{i'} \phi_i^{*' }$. Hence, there exists $\bar{\beta}$ such that for $\beta > \bar{\beta}$, it must be $\mu_i(\Delta_0^{i'}) < \mu_i(\Delta_0^i)$ with $\Delta_0^{i'} > \Delta_0^i$.

¹ We assume throughout that the rate of idiosyncratic play is sufficiently small that the equilibrium conventions described in the text are defined as having an expected duration of more than one period (i.e. $\phi^* > \sigma$).

A.4. Forth result: for sufficiently rational agents, the expected waiting time for a transition ($1/\mu$) is increasing in the group size (z)

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

$$\lim_{\beta \rightarrow \infty} \frac{\mu_i(z')}{\mu_i(z)} = \frac{\frac{z'!}{(z'\phi_i^*)!(z'-z'\phi_i^*)!}}{\frac{z!}{(z\phi_i^*)!(z-z\phi_i^*)!}} \times \lim_{\beta \rightarrow \infty} \frac{\sigma(\beta)^{z'\phi_i^*} [1-\sigma(\beta)]^{z'-z'\phi_i^*}}{\sigma(\beta)^{z\phi_i^*} [1-\sigma(\beta)]^{z-z\phi_i^*}}$$

with $z' > z$. Omitting the constant term, we can write:

$$\lim_{\beta \rightarrow \infty} \frac{\left(\frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z'\phi_i^*} \left(1 - \frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z'-z'\phi_i^*}}{\left(\frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z\phi_i^*} \left(1 - \frac{1}{1+e^{\beta\Delta_0^i}}\right)^{z-z\phi_i^*}} = \lim_{\beta \rightarrow \infty} \frac{(1+e^{\beta\Delta_0^i})^{z\phi_i^*}}{(1+e^{\beta\Delta_0^i})^{z'\phi_i^*}}.$$

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

$$\lim_{\beta \rightarrow \infty} \frac{\mu_i(z')}{\mu_i(z)} = \lim_{y \rightarrow \infty} \frac{y^{z\phi_i^*}}{y^{z'\phi_i^*}} = 0.$$

Hence, there exists $\bar{\beta}$ such that for $\beta > \bar{\beta}$, it must be $\mu_i(z') > \mu_i(z)$ with $z' > z$.