The Scope of Cooperation: norms and incentives*

Guido Tabellini

(IGIER, Bocconi University; CEPR; CES-Ifo; CIAR)

First version: January 2007; this version: February 2007.

Abstract

What explains the range of situations in which individuals cooperate? This paper studies a theoretical model where individuals respond to incentives but are also influenced by norms of good conduct inherited from earlier generations. Parents rationally choose what norms to transmit to their offspring, and this choice is influenced by features of the external environment, such as the quality of external enforcement, or the pattern of likely future economic transactions. The equilibrium displays strategic complementarities between norms and current behavior, which reinforce the effects of changes in the external environment. Norms evolve gradually over time, and if the quality of external enforcement is chosen endogenously under majority rule, the equilibrium displays histeresis: adverse initial conditions may lead to a unique equilibrium path where external enforcement remains weak and inidividual norms discourage cooperation. Evidence from GSS surveys in the US is consistent with some of the model predictions.

*I wish to thank seminar participants in seminars at IGIER and University of Naples for their comments, Giovanna D'Adda, Carlo Prato and Francesco Vecchi for outstanding research assistance. I am grateful to CIAR and Bocconi University for financial support.

1 Introduction

What determines the range of situations in which individuals choose to cooperate with others? This question has been addressed by a large literature in economics, political science and sociology. The traditional approach by economists poses this question in terms of reputation: the scope of cooperation is explained by the strenght of the incentives to preserve one's reputation in repeated interactions, relative to the temptation to cheat.[1]

While the traditional economic approach has yielded important insights, it misses an important dimension. In many social situations individuals behave contrary to their immediate material self interest, not because of an intertemporal calculus of benefits and costs, but because they have internalized a norm of good conduct. Whether we risk our lives fighting in war, or bear the cost of voting in large elections, or refreain from stealing or cheating in an economic transaction, is also determined by our values and beliefs about what is right or wrong.[2] This observation raises several natural questions: what is the origin of specific norms of good conduct? What determines the range of situations over which they are meant to apply? Why do specific norms persist in some environments and not in others? How do values and norms evolve over time? And how do they interact with economic incentives, and with the economic and political environment?

Until recently and with few exceptions, economists have generally refrained from asking these questions and have accepted a division of labor. Other social sciences, primarily sociology, discuss the endogenous evolution of values and preferences. Economics studies the effects of incentives on individual decisions and aggregate outcomes, taking individual preferences as given. Even when social norms have been acknowledged as playing a crucial role, as in the selection of focal points when there are multiple equilibria, economists have studied the implications of these norms, but not their endogenous evolution. A byproduct of this division of labor is that, until recently, the analysis of social norms has generally escaped the discipline of methodological individualism, the paradigm of economics.[3]

[1] Dixit (2004) provides an excellent overview and makes several original contributions taking the economic approach. Axelrod (1984) and Gambetta (1988) are influential contributions in political science and sociology, that overlap with the economic approach.

[2] See for instance the evidence in Fehr, Fischbacher and Gachter (2002), or in Fisman and Miguel (2006).

[3] Besides the pathbreaking work of Gary Becker (see Becker 1993, 1996), recent con-

This paper studies the scope of cooperation combining ideas from economics and sociology. Throughout I neglect the role of reputation, and view cooperation as resulting from a tradeoff between material incentives and individual values. From sociology I borrow the question and the emphasis on norms of good conduct. Namely, I ask how individual values that sustain cooperation evolve endogenously over time. But I address this question with the traditional tool kit of economists, individual optimization and equilibrium analysis, and I focus on how norms interact with economic incentives.

The model is adapted from Dixit (2004). Individuals are randomly matched with others located along a circle, to play a prisoner's dilemma game. They play only once, so there is no role for reputation and cooperation can only be sustained by individual values (a dislike for cheating). The scope of cooperation corresponds to the set of matches over which cooperation can be sustained, and this depends both on economic incentives and individual values.

The model is designed to capture an important idea stressed by sociologists, that rests on the distinction between limited vs generalized morality (eg. Banfield 1958, Platteau 2000). Norms of limited morality are applicable only to a narrow circle of friends or relatives; outside of this narrow circle, cheating is allowed and regularly occurs. Norms of generalized morality instead are meant to apply generally towards everyone. Individuals who have internalized norms of generalized morality are likely to cooperate over a larger range of situations.

To analyze how norms of generalized morality evolve endogenously over time, I build on the work of Bisin and Verdier (2001), Bisin, Topa and Verdier

tributions by economists have started undermining this division of labor. Guiso, Sapienza and Zingales (2006) and Fernadez (2007) review much of this recent (mainly empirical) literature. See also Algan and Cahuc (2006), (2007) and Giuliano (2007). Important theoretical contributions on this line of research include Bisin and Verdier (2001), Akerlof and Kranton (2000), (2006), Benabou and Tirole (2006). Bisin and Verdier (2005) provide a review. Hauk and Saez Marti (2002), Francois and Zabojnik (2005) and Francois (2006) have applied the methodology pioneered by Bisin and Verdier (2001) to closely related issues. Other recent contributions by economists, with a similar approach but a different focus, include Benabou and Tirole (2006), Lyndbeck and Nyberg (2006), Doepke and Zilibotti (2005). The literature by sociologists on these issues is just too large to be mentioned here. But see Nisbet and Cohen (1996) for an interesting example of an attempt to explain the endogenous evolution of individual values. Bowles and Gintis (2000) have also formally studied the evolution of norms facilitating cooperation, but without relying on individual optimization.

(2004). Parents optimally choose what values to pass on to their children, but evaluate their chidren's welfare with their own values. This assumption of "imperfect empathy" implies that the equilibrium is both forward and backward looking. It is forward looking, since parents adapt their educational choices to the future environment of their children. This creates a strategic complementarity between norms and behavior. If more individuals follow a norm of generalized morality, then those who abide by this norm are induced to expand the scope of cooperation (i.e they cooperate over a larger range of matches). And conversely, an expansion in the scope of cooperation facilitates the diffusion of norms of generalized morality. Thus, norms and behavior mutually reinforce each other, and this strenghtens the effects of changes in the external environment. But, the equilibrium is also backward looking, because the parents' values also influence their educational choices. Thus, norms evolve gradually over time and during the adjustment to the steady state the equilibrium reflects historical features of the external environment.

Consider for instance an improvement in the external enforcement of cooperation that is expected to last for ever. This immediately expands the scope of cooperation, as players adapt their behavior to the new environment, for given norms. But the scope of cooperation expands further in subsequent periods, as parents gradually adapt the norms they transmit to their children to the better environment. In the long run, an improvement in external enforcement changes both individual behavior and norms of good conduct. The two changes are self reinforcing and enhance the beneficial effects of better external enforcement. During the adjustment process, the scope of cooperation reflects both current and historical features of the external environment.

The endogeneity of norms has additional implications if, as in Benabou and Tirole (2006), the external environment is also endogenous and reflects political or economic decisions. Better external enforcement of cooperation benefits individuals who abide by norms of generalized morality, and is likely to hurt those who cheat. This gives rise to an additional strategic complementarity. On the one hand, better external enforcement breeds norms that foster cooperation. On the other hand, if a large majority values cooperating with others and dislikes cheating, the political equilibrium supports formal institutions that strongly enforce cooperation. Hence, policies (or formal institutions) and norms of good conduct are mutually self reinforcing.

In the dynamic equilibrium of the model, this strategic complementarity implies histeresis or multiple equilibria, and initial conditions acquire a

special importance. If a norm of generalized morality is initially widespread, then the equilibrium converges to a steady state where a majority retains these positive norms and supports institutions that enforce cooperation. As a result, and for both reasons, the scope of cooperation is large. If instead limited morality initially dominates, then the economy ends up in another steady state, with opposite features: lax external enforcement, poor norms and lack of cooperation. In both cases the equilibrium is unique, although its features are determined by initial conditions. For intermediate values of initial conditions, the model has multiple equilibria and the economy might converge to one or the other steady state, depending on the players expectations.[4]

These results can explain the puzzling persistence of institutions discussed in the recent literature on economic development (eg. Acemoglu et al. 2001, Acemoglu and Robinson 2006, Tabellini 2005, Rajan and Zingales 2006, Rajan 2006, Glaeser et al. 2005). In particular, they can explain why current institutional and organizational failures are often observed in countries and regions that centuries ago were ruled by despotic governments, or where powerful élites exploited uneducated peasants or slaves. In such countries or regions, not only current institutions function poorly and economic outcomes are disappointing, but also individuals typically mistrust others and display values and beliefs that are consistent with norms of limited morality - see the evidence in Tabellini (2005).

This lack of social capital in environments with a history of political abuse and exploitation could be both an independent cause and an effect of the malfunctioning of current institutions. The results of this paper point out that in practice it is bound to be very difficult to identify which specific institutional features are responsible for observed economic outcomes. In the equilibrium of the model, both formal institutions and norms of good conduct are jointly determined, and their evolution is dictated by initial and possibly random historical circumstances.

Nevertheless, micro data strongly support the idea that distant history has an effect on individual values and beliefs. Exploiting what Fernadez (2007) has called the "epidemiological approach", I consider the attitudes towards trusting others displayed by third generations US immigrants, using

[4]Francois (2006) and Hauk and Saez Marti (2002) also study the two-way interaction between endogenous norms and features of the external environment (formal institutions), but they don't focus on political decisions.

data from the General Social Surveys. After controlling for observable individual attributes, trust is higher amongst the US immigrants whose ancestors lived in countries that had better political institutions and more educated population over a century ago, while past economic development of the country of origin does not seem to matter. Overall, this evidence supports the idea that the intergenerational transmission of norms is an important determinant of the scope of cooperation.[5]

The outline of the paper is as follows. Section 2 outlines the model in its simples version with exogenous preferences for cooperation. Section 3 makes preferences endogenous and shaped by the educational choices of optimizing parents. Section 4 adds politics and studies the equilibrium with endogenous preferences and endogenous polcies. Section 5 presents the evidence on the intergenerational transmission of norms. Section 6 concludes.

2 The Scope of Cooperation with Exogenous Values

2.1 Preliminaries

The model is adapted from Dixit (2004), chapter 3. A continuum of one-period lived individuals is distributed along a circle. The density of individuals per unit of arc length is 1, and the size of the circumference is $2S$. Thus the maximum distance between two individuals is S, and S measures the size of the community.

Each individual is randomly matched with another. As in Dixit (2004), the matching technology has a local bias: individuals are more likely to be matched with those located nearby. Specifically, the probability of a match with someone located at distance y decreases exponentially with y and is given by (the denominator is a normalizing factor that insures that the probabilities of all matches between 0 and S sum to 1):

$$\frac{\alpha e^{-\alpha y}}{2\left[1-e^{-\alpha S}\right]} \tag{1}$$

[5] Guiso, Sapienza and Zingales (2006) and Algan and Cahuc (2006), (2007) used a similar approach to study trust in the GSS data, but they did not focus on historical variables in the ancestors' country. See also Dohmen et al. (2006) who use data on German households.

Thus, the parameter α captures the matching technology, as in Dixit (2004). As α increases, the probability of more distant matches drops. A higher α thus corresponds to an environment in which more matches are local.

The two matched individuals observe their respective locations and play a simple prisoner's dilemma game. Each player simultaneously chooses whether to cooperate (play C) or not to cooperate (play NC). The material payoffs from playing the game are illustrated in Table 1:

Table 1. Payoffs from Prisoner's Dilemma

	C	NC
C	c, c	$-l, c+w$
NC	$c+w, -l$	$0, 0$

where $c, l, w > 0$. It is natural to interpret the parameters w and l as reflecting the quality of external enforcement. A better enforcement of private contractual arrangements would reduce the temptation to cheating on a cooperating partner (w), and it would reduce the loss of being cheated (l).

Besides obtaining the material payoffs described above, individuals also get additional psychological benefits or losses from playing the game. In particular, each individual incurs a non-economic *cost* $d > \max(l, w)$ whenever it plays NC irrespective of how its opponent played. These non-economic costs decay with distance in the match at exponential rate $\theta > 0$. Thus, playing NC against an opponent located at distance y results in non-economic costs $d\, e^{-\theta y}$. This formulation captures the idea that norms of good conduct apply with particular force with regard to a circle of close friends or relatives, but are weaker in encounters with more distant individuals (whatever the space over which distance is measured). These additional individual consequences from not cooperating might differ across individuals, and later in the paper are determined endogenously. This set up and notation are illustrated in Figure 1.

As will become clear below, when playing the matching game individuals compare their material payoffs with the non-economic cost of not-cooperating. Here the non-economic costs of cheating decrease exponentially at the rate θ with the distance in the match, while the economic payoffs of the prisoner's dilemma game do not depend on distance. An alternative fomulation, suggested by Dixit (2004), would have the material payoffs increase exponentially at rate $\theta > 0$ with distance, capturing the idea that matches between more distant traders are likely to entail bigger gains from trade. This formulation was pursued in a previous version and the results (but not the algebra)

were identical. More generally, the parameter θ can be interpreted as capturing the rate at which non-economic costs decay, relative to the rate at which economic payoffs increase with distance. The general point that the model seeks to capture is that interactions between more distant individuals are likely to entail bigger gains from trade, but also weaker self restraints against purely selfish motivations. We will refer to this parameter θ as the rate at which norms of reciprocity decay with distance.

In section ?? below, we discuss an extension that allows for reciprocity in the non-economic costs of cheating. Namley, the cost $d\ e^{-\theta y}$ is born only if the opponent cooperates, but not if both players cheat. All the results discussed in the paper go through, except that with reciprocity we get additional strategic complementarities and hence additional equilibria.

2.2 Equilibrium with a single representive individual

In this subsection d and θ are fixed at the same value for everyone. Consider the perspective of someone who has to decide whether to play NC or C in a match with a partner at distance y. Throughout, we denote by $\pi(y)$ the probability that his partner will play C. We can express his net expected material gain from playing NC rather than C in a match with y as:

$$T(\pi(y)) = [(w-l)\,\pi(y) + l] > 0 \tag{2}$$

We can think of this expression as the temptation not to cooperate. The right hand side of (2) is strictly positive: it is always better not to cooperate. Nevertheless, the function $T(\pi(y))$ is non-increasing in $\pi(y)$, the probability that the opponent will play C, since $w \leq l$.

This temptation must be balanced against the non-economic costs of not cooperating, $de^{-\theta y}$. An individual is just indifferent between playing C or NC in a match with someone at distance $\tilde{y}$ if:

$$T(\pi(\tilde{y})) = de^{-\theta \tilde{y}} \tag{3}$$

Solving for $\tilde{y}$, we obtain:

$$\tilde{y} = \{\ln[d] - \ln[(w-l)\,\pi(\tilde{y}) + l]\}\,/\theta \tag{4}$$

Note that the cost of not cooperating, $de^{-\theta y}$, is strictly decreasing in y. This follows from the assumption that the norm of good conduct applies

with greater strenght to closer partners. Hence, holding π constant, this individual prefers to play C in a match with someone at distance $y < \tilde{y}$, and he prefers to play NC if $y > \tilde{y}$.

To pin down the equilibrium, we have to solve for $\pi(y)$, the probability that an opponent located at distance y cooperates. This is done in Appendix 1, which proves that the equilibrium outcome depends on the distance y between the two partners. Cooperation is sustained if the distance y falls short of some thresholds, while it fails above those thresholds. Specifically define the distance thresholds Y' and Y:

$$Y' = [\ln d - \ln l] / \theta \tag{5}$$

$$Y = [\ln d - \ln w] / \theta \tag{6}$$

and let y_{Min} and y_{Max} be respectively $y_{Min} = Min\{Y', Y\}$ and $y_{Max} = Max\{Y', Y\}$. Then Appendix 1 proves the following:

Proposition 1 *Let the distance in a match be y. (i) If $y < y_{Min}$ then both partners play C and the equilibrium is unique. (ii) If $y > y_{Max}$, then both partners play NC and the equilibrium is unique. (iii) If $y \in [y_{Min}, y_{Max}]$ and $y_{Min} \neq y_{Max}$, then there are multiple equilibria. Specifically, suppose that $w < l$, so that $y_{Max} = Y > Y' = y_{Min}$. Then for $y \in [y_{Min}, y_{Max}]$ there are two equilibria in pure strategies, one in which both partners play C, and the other in which both partners play NC. Suppose instead that $w > l$, so that $y_{Max} = Y' > Y = y_{Min}$. Then for $y \in [y_{Min}, y_{Max}]$ there are two equilibria in pure strategies, one in which one partner plays C and the other plays NC, and the other equilibrium in which the roles are reversed.*

Throughout the rest of the paper, I restrict attention to the case $w \leq l$, so that the equilibrium is symmetric, and I only consider the more efficient equilibria, to give the best possible chances to cooperation. Hence, if everyone has the same cost parameters d and θ, then the best equilibrium entails reciprocal cooperation in a match of distance $y \leq Y$, and non-coperation if the distance is $y > Y$.

This equilibrium provides a simple theory of the scope of cooperation, and the variable Y defined by (6) summarizes all the relevant information. In particular, individuals cooperate over a larger range of matches (the distance Y increases):

- if the benefit of cheating (w) falls;

- if the non-economic cost of cheating (d) rises.
- if norms of good conduct decay more slowly with distance (if θ falls);

These results are similar to those obtained by Dixit (2004) in his model based on reputation, despite the different reason why here individuals refrain from cheating. In contrast to Dixit (2004), however, here the range of cooperation does not depend on the likelihood of matches with more distant partners, α, nor on the overall size of the economy, S. Note also that, in the Pareto superior equilibrium, the range of cooperation does not depend on the cost of being cheated, l.

2.3 Equilibrium with two types of agents.

In this subsection I continue to assume that the cost of not cooperating is an exogenous parameter, but now I allow for two possible types indexed by $k = 0, 1$. Both types bear the same cost d of cheating. They differ in the rate at which this cost decays with distance, say θ^1 and θ^0, with $\theta^0 > \theta^1 + \ln(\frac{l}{w}) \geq \theta^1$. For shortness, I refer to those with $k = 1$ as trustworthy or "good", since in the Pareto superior equilibrium they cooperate in a larger range of matches, while those with $k = 0$ are called not-trustworthy or "bad". Individuals in a match observe distance, y, but not the trustworthiness of their partner. The fraction of good (θ^1) types in the population is a known parameter n, with $1 > n > 0$.[6]

Repeating the analysis of the previous subsection, it is easy to see that, for both types, there is a distance threshold $\tilde{y}^k$, $k = 0, 1$, that leaves that type indifferent between playing C and NC, given the probability $\pi(\tilde{y}^k)$ that his partner will cooperate. Such threshold $\tilde{y}^k$ is still defined by (4), with θ^k on the right hand side, for $k = 0, 1$. As stated above, we consider only the Pareto superior equilibrium that sustains the maximum possible degree of cooperation (here too there are multiple equilibria similar to those of the previous subsection).

To characterize such an equilibrium, we need to pin down the equilibrium probability of cooperation $\pi(y)$ for all possible values of y. Repeating the

[6]The assumption that $\theta^0 > \theta^1 + \log(\frac{l}{w})$, rather than just $\theta^0 > \theta^1$, simplifies the analysis because it reduces the possible types of equilibria that may exist, but all the results go through (with some additional complications) under the weaker condition that $\theta^0 > \theta^1$. A previous version solved for the case in which different types have the same value of θ, but different non-economic costs of cheating, say $d^1 > d^0$.

steps in the previous subsection, it is useful to define the following thresholds that induce cooperation by the two types:

$$Y^0 = [\ln d - \ln w] / \theta^0 \quad (7)$$
$$Y^1 = [\ln d - \ln [(w - l) n + l]] / \theta^1 \quad (8)$$

By construction, in a match of distance $y \leq Y^0$, all types with $\theta = \theta^0$ find it optimal to cooperate if they expect their partner always to cooperate; if the distance exceeds Y^0, they prefer not to cooperate, irrespective of what their partner does. The term Y^1 corresponds to the distance threshold that sustains cooperation of the good types, given that their expectations are consistent with equilibrium.

Since $n \geq 0$, our maintained assumption that $\theta^0 > \theta^1 + \ln(\frac{l}{w})$ implies that $Y^1 > Y^0$. Hence, the good types cooperate over a strictly larger range of matches. Those with $\theta = \theta^0$ continue to behave as described above: they cooperate if $y \leq Y^0$, and they don't cooperate if $y > Y^0$. This behavior is optimal by definition of Y^0, and given their expectations that everyone cooperates if $y \leq Y^0$ (see the proof of Proposition 1 for more details). Those with $\theta = \theta^1$ find it optimal to cooperate up to distance $y \leq Y^1$, given that they expect cooperation if their partner is good, and no cooperation if he is bad (and given that the type cannot be observed).[7] We summarize this discussion in the following:

Proposition 2 *In the Pareto superior equilibrium of the matching game, individuals of type k, cooperate in a match of distance $y \leq Y^k$ and do not cooperate if $y > Y^k$, for $k = 0, 1$ and with $Y^1 > Y^0$.*

The properties of this equilibrium are the same as those of the equilibrium described in the previous subsection, with a single type, except that now we get two additional implications. The maximum range over which at least some individuals cooperate (the threshold Y^1) increases:

- if the loss from cooperating against a cheating opponent (l) falls;

[7] If $\theta^1 > \theta^0$ but $\theta^0 < \theta^1 + \log(\frac{l}{w})$, then the good and bad types would behave identically if $n > 0$ but small. For n sufficiently large, we would obtain again that $Y^1 > Y^0$ and different types behave differently. Intuitively, the probability of encountering a good type must be sufficiently high to make a difference, or else the difference in preferences between the two types must be sufficiently large.

- if the fraction of good types (n) increases.

The first implication follows from imperfect information: as individuals cannot observe their opponent type, in equilibrium the good players bear the risk of cooperating against a cheating opponent. Clearly, the smaller is the resulting loss, the larger is the range of matches over which cooperation can be sustained. The second implication reflects a strategic complementarity: given $l > w$, individuals are more willing to cooperate the higher is the probability that their partner will also cooperate.

In the introductory section we stressed the distinction between limited vs generalized morality, namely between norms of good conduct that apply in a narrow or in a large set of social interactions. The equilibrium summarized in Proposition 2 provides an analytical foundation to this distinction. Matches within the distance Y^0 can be interpreted as interactions within a small group of friends or relatives. Everyone can be trusted to cooperate and behave well within this narrow group. Matches of distance higher than Y^0 can be interpreted as interactions in the market or in a larger and more anoymous set of individuals. Not everyone can be trusted to behave well in these less frequent interactions, because the temptation to capture the material benefits of cheating might exceed the psychological discomfort of violating an internalized norm of good conduct, at least for some individuals in the population. The scope of maximal sustainable cooperation over these more distant matches is summarized by the variable Y^1. This variable reflects the features of the external environment that determine individual incentives to cooperate outside of the narrow circle corresponding to the distance Y_0.

Finally, note that, in this model with exogenous preferences, as the external environment changes, individuals react immediately by altering their equilibrium behavior. The scope of cooperation is enhanced by better external enforcement (lower w or lower l). But there is no dynamics and what matters is current enforcement, not institutions in the distant past. Hence, this version of the model is unable to explain institutional persistence.

3 Endogenous Values

3.1 The model

This section models the endogenous evolution of the norms that sustain cooperation, as captured by the parameter θ^k. Our goal is to study how par-

ents rationally choose what values to transmit to their children, and how this choice is affected by economic incentives and by features of the external environment. For simplicity θ^k can only take two values, θ^1 and θ^0 with $\theta^0 > \theta^1 + \ln(\frac{l}{w}) \geq \theta^1$ as in the previous section. But we assume that the actual value taken by θ^k for each individual reflects two forces: the exogenous influence of nature or of the external environment, and the deliberate and rational efforts of parents, through education or time spent with their children. The crucial assumption is that parents are altruistic and care about the utility of their offspring, but evaluate their kid's expected welfare with their own preferences. This assumption of "imperfect empathy" (cf., Bisin and Verdier 2001) implies that in some circumstances parents devote effort to try and shape the values of their children to resemble their own.

Specifically, consider an ongoing economy that lasts for ever. Individuals live two periods. In the first period of their life they are educated by their parents and, once education is completed, they are active players in the game described above. In the second period, each individual is the parent of a single kid and his only activity is to devote effort to educate him. Parental education increases the probability that the kid becomes good (i.e. that $\theta^k = \theta^1$), but it is costly for the parent. To obtain a closed form solution we assume a quadratic cost function: $-\frac{1}{2\varphi}f^{\,2}$, where $f \geq 0$ denotes parental effort to educate his kid, and $\varphi > 0$ is a parameter that captures the marginal cost of effort (higher φ corresponds to a lower marginal cost). Parental effort is chosen by each parent before observing his kid's value. Conditional upon parental effort, the probability of having a good kid does not depend on the the value parameter of the parent. Specifically, if a parent exerts no effort to educate his kid, then with probability $1 > \delta > 0$ the kid is born good ($\theta^k = \theta^1$), and with probability $1-\delta$ the kid is born bad ($\theta^k = \theta^0$). If instead the parent exerts effort f to educate his kid, then the probability of having a good kid is $\delta + f$, and the probability of a bad kid is $1 - \delta - f$. [8]

[8]Note the asymmetry. We let parents exert effort to increase the expected trustworthiness of their kid, but we assume that they cannot exert effort to reduce it. With a slight change in notation, this asymmetry can be interpreted almost literally as saying that inculcating trustworthiness in one's kid is costly, while inculcating dishonesty or non-trustworthyness does not cost any effort to the parent. A previous version of this paper removed the asymmetry, and assumed that it was equally costly for a parent to increase or decrease the trustworthyness of one's kid, relative to the choice made by nature. The qualitiative results were similar, although the derivation was more complicated and additional conditions on parameter values had to be imposed to obtain some of the comparative

Once parents have completed the education, each young player observes his own type and plays the matching game described in the previous section. Thus, the economy in any given period t behaves exactly like in the matching game of the previous section with two exogenous types of agents, except that here we have to keep track of time, because the composition of types is endogenous and varies with time. As already noted, the matching game has multiple equilibria if $w \neq l$. Throughout, I maintain the assumption that $l > w$ and I restrict attention to the Pareto superior equilibrium of the matching game described in the previous section. Let n_t denote the proportion of good (θ^1) individuals in the population at the end of period t (i.e., after parents have exerted effort into educating their kids during period t). Then, by Proposition 2, players of type k cooperate in a match of distance $y \leq Y_t^k$, and do not cooperate if $y > Y_t^k$, where the distance threshold that triggers cooperation, Y_t^k, is still given by (7) and (8), except that it is indexed by t since Y_t^k might depend on time through n_t.

Consider a parent of type p who gives birth to a kid of type k in period t, for $k, p = 0, 1$. Let V_t^{pk} denote the parent's evaluation of his kid's expected utility in the Pareto superior equilibrium of the matching game described in subsection 2.3. By the assumption of imperfect empathy, we can write V_t^{pk} as:

$$V_t^{pk} = U_t^k - \frac{\alpha d}{2\left[1 - e^{-\alpha S}\right]} \int_{Y_t^k}^{S} e^{-(\alpha + \theta^p)z} dz \tag{9}$$

where U_t^k denotes the expected equilibrium material payoffs of a kid of type k, while the second term on the RHS of (9) is the parent's evaluation of his kid's expected non-economic cost of not cooperating in matches of distance greater than Y_t^k. Note that this evaluation is done with the parent's value parameter, θ^p, rather than with the kid's value. Thus, if the kid is born with the same value of his parent (if $\theta^p = \theta^k$), then parent and kid evaluate the outcome of the kid's matching game identically. But if the kid and the parent have different values, then V_t^{pk} differs from the kid's own evaluation: the value parameter in the last term on the right hand side of (9), θ^p, is that of the parent, while the relevant distance thresholds according to which the

statics results mentioned below. Unlike in Bisin and Verdier (2001), and given the different focus of our analysis, we neglect the possibility that the kids'values or the effect of parental effort also depend on the current distribution of types in the population. This implies that to obtain dynamic stability we need to impose additional conditions on parameter values.

game is played, Y_t^k, are those of the kid.

Exploiting Proposition 2 in the previous section, the kid's expected material payoffs in the matching game are:

$$U^k = \frac{\alpha}{2\left[1-e^{-\alpha S}\right]}\left[\int_0^{Y_t^k} e^{-\alpha z}[c\pi_t(z) - l(1-\pi_t(z))]dz + \int_{Y_t^k}^{S} e^{-\alpha z}(c+w)\pi_t(z)dz\right] \tag{10}$$

where $\pi_t(z)$ denotes the probability that a partner at distance z will cooperate in period t in the Pareto superior equilibrium - $\pi_t(z)$ is indexed by time because it might depend on n_t. The first term on the right hand side is the expected utility when cooperating, given that the partner cooperates with probability $\pi_t(z)$. The second term on the right hand side is the expected utility of not cooperating, again given the probability that the partner cooperates (recall by Table 1 that if both partners do not cooperate then their payoffs are normalized to 0). Subsection 2 of the Appendix writes down the expressions for U_t^k in the Pareto superior equilibrium considered in Proposition 2, replacing $\pi_t(z)$ with the corresponding equilibrium expressions.

The following Lemma, proved in subsection 3 of the appendix, verifies that a parent always prefers to have a kid with his own values, and this is a strict preference if different values induce different behavior (i.e. if $Y^1 > Y^0$):

Lemma 3 *If $k \neq p$, then $V_t^{pp} \geq V_t^{pk}$, with strict inequality if $Y^1 > Y^0$.*

This intuitive result reflects two assumptions. First, individual types are not observable, and hence there is no incentive for strategic delegation (i.e there is no strategic gain in distorting the kid's preferences when he plays the subsequent game).[9] Second, imperfect empathy implies that the only reason for changing one's kid value θ^k is to induce him to change his behavior: the disutility from non-cooperation is evaluated by the parent with his own value,

[9]The assumption that a kid's value is not observable is not necessarily always appropriate. Levy and Razin (2006), for instance, formulate a theory of religion based on the assumption that one's religion is obervable (maybe only within a subset of the population). This creates a strategic incentive to join a religious organization, to signal one's type. But in Levy and Razin (2006), individual values are stable and exogenous, and individuals choose their own religion (i.e. there is no role for parents to shape their kid's values or beliefs).

θ^p, and hence the parent does not directly benefit from a lower cost of non-cooperative behavior by his kid, except through the induced effects on the kid's behavior.

Given that effort to educate one's kid costs the parent some disutility according to the quadratic function summarized above, and given that parental effort is chosen before observing the kid's type, Lemma 3 immediately implies:

Corollary 4 *A "good" parent* $(p = 1)$ *exerts strictly positive effort. A "bad" parent* $(p = 0)$ *exerts no effort.*

Intuitively, by Lemma 3, a bad parent would like to have a bad kid. Hence, he will never exert any effort to increase his kid's expected trustworthiness. Conversely, a good parent would like to have a good kid. Hence at the maring he is prepared to exert at least some effort to increase the probability of this happening.

Given this result, the proportion of good individuals playing the matching game in period t, n_t, evolves endogenously over time according to the following law of motion:

$$n_t = n_{t-1}(\delta + f_t) + (1 - n_{t-1})\delta = \delta + n_{t-1}f_t \tag{11}$$

where from here onwards, with a slight abuse of notation, f_t denotes effort by the good type parents only. Intuitively, if parents exerted no effort, then the average fraction of good kids in the population would just equal δ. But the good parents (of which there is a fraction n_{t-1} in period t) exert effort f_t in period t, and this increase the fraction of good kids in the population on averge by $n_{t-1}f_t$.

3.2 The parent's optimization problem

This subsection describes how the good parents choose effort, f_t. Each parent takes as given the effort choices of all the other parents, and takes into account the equilibrium implications of his kid's value for his own welfare, according to (10) and (9). At an interior optimum, the first order condition for an optimum equates the marginal cost and the expected net marginal benefit of effort, and by (10) and (9) it can be written as:

$$f_t/\varphi = (U_t^1 - U_t^0) + \frac{\alpha d}{2\left[1 - e^{-\alpha S}\right]} \int_{Y^0}^{Y_t^1} e^{-(\alpha+\theta^1)z} dz \tag{12}$$

Consider the right hand side of (12), that captures the net marginal benefit of effort. The first term is the change in the kid's expected material payoffs, if his value switches from θ^0 to θ^1. This term is always negative, since for any probability that the partner in a match will cooperate, the kid's expected material payoffs are always higher if the kid plays NC (see (33) in subsection 2 of the appendix for a proof). The second term is the expected benefit of extending the scope of the kid's cooperative behavior to a larger range of matches, evaluated with the parent's values, $\theta^p = \theta^1$ (note that Y^0 is time invariant by (5)). This term is always positive, since extending the scope of the kid's cooperative behavior decreases the direct non-economic cost born by the parent. Hence, the parent perceives a tradeoff. Increasing his kid's trustworthiness hurts the kid's expected material payoffs, and this cost is internalized by the parent. But a good kid also provides expected direct non-economic benefits to the parent. By Corollary 4, we know that the benefits exceed the costs, and hence $f_t > 0$.

Exploiting the equilibrium expression for $U_t^1 - U_t^0$ as given by (33) in subsection 3 of the Appendix, we can rewrite the parents' optimality conditions, (12) as:

$$f_t = \frac{\varphi\alpha d}{2\left[1-e^{-\alpha S}\right]}\left[-e^{-\theta^1 Y_t^1}\int_{Y^0}^{Y_t^1} e^{-\alpha z}dz + \int_{Y^0}^{Y_t^1} e^{-(\alpha+\theta^1)z}dz\right] \equiv F(Y_t^1) > 0 \quad (13)$$

Note that f_t denotes a probability. Thus, implicit in (12) is a restriction on parameter values (and in particular on φ) guaranteeing that $1 \geq f_t$. As we shall see below, dynamic stability of the equilibrium requires $1 > f_t$, which we assume throughout. Equation (13) defines f_t as a known function of Y_t^1, $f_t = F(Y_t^1)$ - note that all other terms on the right hand side of (13) are fixed parameters, including Y^0. Subsection 4 of the appendix proves:

Lemma 5 *The function $F(Y_t^1)$ is strictly increasing in Y_t^1.*

Intuitively, if the difference in behavior between good and bad types increases (as captured by the variable Y_t^1), then good parents are induced to put more effort to increase the probability of having a good kid. That is, parental effort increases as the behavioral implications of their kids values become more relevant.

This property is important, because it gives rise to a second strategic complementarity. If parents expect others to put more effort into education,

they anticipate that the fraction of good types will increase. They realize that this will expand the scope of cooperation, Y_t^1, and as a result they exert more effort. In fact, it is easy to verify that the educational game described in this section is supermodular (cf. Amir 2003).

3.3 The equilibrium

Replacing f_t with $F(Y_t^1)$ in (11) and simplifying, the equilibrium is thus given by the vector (Y_t^{1*}, n_t^*) that solves the following two equations:

$$Y_t^1 = [\ln d - \ln [(w-l)\, n_t + l]] / \theta^1 \equiv Y(n_t) \quad (14)$$

$$n_t = \delta + n_{t-1} F(Y_t^1) \equiv N(Y_t^1, n_{t-1}) \quad (15)$$

The first equation defines the maximum distance Y_t^1 that sustains cooperation by the good types, as a function of the proportion of other good types in the population, $Y_t^1 = Y(n_t)$. Since we assumed strategic complementarity in the matching game ($l > w$), cooperation is easier to sustain if there are many good types around. Hence, Y_t^1 is an increasing (and convex) function of n_t, as depicted by the curve $Y_t^1 = Y(n_t)$ in Figure 2.

The second equation defines the law of motion of the proportion of good types, as a function $n_t = N(Y_t^1, n_{t-1})$. As Y_t^1 increases, good parents are induced to put more effort into changing their kid's value (by Lemma 5, the function $F(Y_t^1)$ is strictly increasing in Y_t^1). Hence, the function $n_t = N(Y_t^1, n_{t-1})$ is also increasing in Y_t^1.

Together, equations (14) and (15) implicitly define the equilibrium vector (Y_t^{1*}, n_t^*) as a function of n_{t-1} :

$$Y_t^{1*} = G^Y(n_{t-1}) \quad (16)$$

$$n_t^* = G^n(n_{t-1}) \quad (17)$$

Setting $n_t = n_{t-1} = n_s$, we obtain the steady state equilibrium:

$$Y_s^{1*} = Y(n_s^*) \quad (18)$$

$$n_s^* = \frac{\delta}{1 - f_t} \quad (19)$$

where $f_s = f_t(Y_s^{1*})$ is the steady state value of educational effort by the good parents.

As both curves in Figure 2 are increasing, multiple equilibria are possible. That is, the same fraction of "good" parents n_{t-1} might imply more than one equilibrium pair for parental effort and scope of cooperation, (Y_t^{1*}, n_t^*). The reason for the possible multiplicity is the already mentioned strategic complementarity between norms and behavior.

The equilibrium is unique if the curve $n_t = N(Y_t^1, n_{t-1})$ always intersects the curve $Y_t^1 = Y(n_t)$ from left to right, as drawn in Figure 2. Subsection 4 of the appendix proves that a sufficient condition for this to happen is:

$$\frac{1}{l-w} > \varphi \left[(\frac{w}{d})^{\lambda^0} - (\frac{w}{d})^{\lambda^1} \right] \tag{A1}$$

Note that this condition is certainly satisfied if φ or $(l-w)$ are sufficiently small - that is if the marginal cost of effort for the parents is sufficiently high, or if the strategic complementarity in the prisoner's dilemma game is sufficiently small. Condition (A1) thus guarantees that the equilibrium (Y_t^{1*}, n_t^*) is unique. In the remainder of the paper we assume that this condition holds.[10]

Subsection 5 of the appendix proves that, under condition (A1), the functions $G^Y(n_{t-1})$ and $G^n(n_{t-1})$ are strictly increasing in n_{t-1}. Subsection 6 of the appendix also proves that there is a $\bar{\varphi} > 0$ such that, if $\bar{\varphi} > \varphi > 0$, then $dG^n(n_{t-1})/dn_{t-1} < 1$. We summarize the implications of this discussion in the following:

Proposition 6 *If condition (A1) holds, then the equilibrium (Y_t^{1*}, n_t^*) is unique. For $\varphi > 0$ but small enough, the equilibrium asymptotically reaches the steady state (Y_s^{1*}, n_s^*) defined by (18)-(19). If (A1) holds, then the path towards the steady state is monotonic and during the adjustment to the steady state (Y_t^{1*}, n_t^*) move in the same direction.*

3.4 Discussion

As already noted, the variable Y_t^1 can be interpreted as the scope of cooperation induced by a norm of generalized morality. As the external environment changes, individuals immediately adjust their behavior responding to incentives, and Y_t^1 reacts accordingly. But this is not the end of the story. The

[10] Of course, the matching game described in section 2 and played in each period by the kids has multiple equilibria. But here we are restricting attention to the Pareto superior equilibrium of the matching game.

diffusion of a norm of good conduct, as captured by the fraction of good individuals, n_t, is also part of the equilibrium. The variable n_t evolves slowly over time, as it reflects both the current features of the environment, as well as the culture of previous generations.

Cultural forces and economic incentives interact through strategic complementarities and have self-reinforcing effects on individual behavior. On the one hand, having more good individuals around expands the scope of cooperation, because it reduces the risk of exploitation by a more shrewd partner (Y_t^1 is an increasing function of n_t). On the other hand, as the scope of cooperation increases, parents put more effort into education, because they realize that the cultural trait that they value in their children will have more pronounced behavioral implications (f_t and hence n_t are increasing in Y_t^1).

We now discuss how the equilibrium is affected by changes in the underlying parameters. Throughout we assume that condition (A1) holds and that φ is sufficienty small that equilibrium is dynamically stable. We also assume that the economy is originally in the steady state, (n_s^*, Y_s^{1*}).

Better external enforcement Suppose that at the beginning of period $t = 0$, before parents choose their educational effort, the payoffs to the matching game change. Specifically, consider a reduction in the loss l, the cost of cooperating against a deviating partner. This change can be interpreted as an improvement in the external enforcement of cooperation.[11] As l is reduced, the curve $Y_t^1 = Y(n_t)$ in Figure 2 shifts to the right - cf (14). Intuitively, for a given n_t, the good types now cooperate over a larger range of matches. Moroever, the threshold Y^0 is not affected by this change. As a result, the curve $n_t = N(Y_t^1, n_{t-1})$ remains unaffected in period 0, since its position does not directly depend on the parameter l if Y^0 remains unchanged - cf. (13). Thus, the scope of cooperation expands.

This improvement in the external environment in turn induces parents to increase their educational effort - the curve $N(Y_t^1, n_{t-1})$ is increasing in Y_t^1, as drawn in Figure 2. Hence, this initial change results in a larger fraction of good types (n_0 rises), which further increases the scope cooperation

[11] A change in the temptation to cheat, w has ambiguous effects on the equilibrium, since it affects both Y^0 and Y_t^1. The next section discusses the consequences of external enforcement more at length, also considering the effect of changing w over some distance ranges. A larger gain from cooperation, c, holding the parameters w and l fixed, has no effects on the equilibrium, since it does not affect any of the margins that are relevant for the kids or the parents decisions.

sustainable in period 0.

But this is not the end of the process, because in period 1 the curve $N(Y_t^1, n_{t-1})$ shifts upwards. Since more parents are good (n_0 has risen), more of them put effort into educating their children. Hence in period 1 the proportion of good kids is even higher than in period 0 ($n_1 > n_0$) and this brings about an even larger range of cooperative matches in period 1, $Y_1^1 > Y_0^1$. The adjustment continues smoothly over time, and for φ small enough a new steady state is reached. This new steady state has both a larger fraction of good types and cooperation is sustained over a longer range of matches. Thus, a permanent change in the external environment continues to have effects for many generations after it has occurred, through the educational choices of rational parents.

A larger economy Next, consider the effect of an increase in the size of the circle, S. As emphasized by Dixit (2004), this corresponds to the economy growing in size, for instance because technological improvements have increased the scope of mutually beneficial economic exchange. As the population remains fixed in size, the circle becomes uniformly less dense, and the probability of a match within any distance $y < S$ is reduced. This means that the probability of a match in the interval $[Y^0, Y_t^1]$ also shrinks. As a result, f_t goes down as parents reduce their educational effort - cf. the right hand side of (13). Intuitively, parental education only matters for matches in the interval $[Y^0, Y_t^1]$, because outside of this interval the kid beahves in the same way irrespective of the education received. This shifts down the $N(.)$ curve, while the $Y(.)$ curve on impact remains unaffected. In the new steady state both n_s and Y_s^1 shrink: an expansion in the size of the economy reduces the fraction of trustworthy types and shrinks the range of cooperative matches. To put it more bluntly, globalization (the equivalent of an increase in S) reduces the scope of cooperation because it destroys the values that induce individuals to cooperate. This is the same qualitative effect found in Dixit (2004), although here the forces at work are very different.

A more localised economy Suppose that α increases. This corresponds to increasing the rate at which the probability of a match decays with distance. In other words, the economy becomes more localised, in the sense that distant matches are less likely. This has no effect on the curve $Y_t^1 = Y(n_t)$, but it shifts the curve $N(Y_t^1, n_{t-1})$, and hence changes educational effort,

since parents realize that thier kid is more likely to interact with closer partners. How exactly effort changes with α depends on the scope of the norms of limited and generalized morality. Specifically, Subsection 7 of the appendix proves that equilibrium effort, f_t^*, increase with α if $Y_t^1 < \bar{Y}$, and decreases with α if $Y_0 > \bar{Y}$, where:

$$\bar{Y} = \frac{1 - e^{-\alpha S}(1 + \alpha S)}{\alpha(1 - e^{-\alpha S})} \tag{20}$$

Note that $\bar{Y} < S$ (see subsection 7 of the Appendix). Thus, quite intuitively, if the scope of norms of generalized morality is small relative to the size of the economy, (Y_t^1 is low), then a more localised economy means that matches in the region $[Y_0, Y_t^1]$ become more likely, and this induces parents to increase effort. Conversely, if the scope of limited morality is large relative to the size of the economy (Y^0 is high), then kids are more likely to interact in distance ranges below Y^0, and in this case a more localised economy reduces parental effort. For intermediate values the effect of a more localized economy can go either way depending on interactions with other parameters of the model. In other words, a more localised economy discourages cooperation if the scope of limited morality is not too narrow, and has the opposite (i.e. positive) effect if the scope of generalized morality is not too wide, relative to the size of the economy. The dynamic adjustment is then as discussed above.

More generally, the effects of changing α and S illustrate a general and intuitive property of the model. Effort to inculcate trustworthiness in one's kid depends on the likely patterns of interaction in the economy. Whatever increases the likelihood of interactions in the region above Y^0 and below Y_t^1, where the distinction between limited and generalized morality matters, also increases the parents incentive to educate their kids. Very local interactions (below Y^0) or very distant interactions (above Y_t^1) have the opposite effect, because the distinction between limited and generalized morality has no behavioral implication in those regions.

3.5 Extensions: Reciprocity

The model assumes that the non-economic cost d is born irrespective of whether the partner cooperates or not. An alternative and perhaps more plausible formulation has the player bearing the cost d only if he cheats against a cooperating partner in a match. This alternative formulation can easily be incorporated in the model, and it would result in two main changes.

First, the Pareto superior equilibrium of the matching game would entail an additional strategic complementarity. Specifically, while the definition of Y^0 is not affected, the upper threshold of cooperation, Y_t^1, now becomes:

$$Y_t^1 = \left[\ln d + \ln n_t - \ln\left[(w-l)\, n_t + l\right]\right]/\theta^1 \equiv Y(n_t) \tag{21}$$

Comparing (21) with the previous expression in (14), we have added the term $\ln n_t$ that was missing in (14). Intuitively, if more good types are around, then the expected cost of cheating rises (since it is more likely to occur againts a cooperating opponent). Hence, a rise in the fraction of good types (a higher n_t) induces a further expansion in the scope of cooperation corresponding to the norm of generalized morality. Note that this strategic complementarity arises even if $l = w$, in which case the matching game without the norm of reciprocity has a unique equilibrium (cf. Proposition 1).[12]

Second, parents too bear the non-economic cost d only if their kid cheats against a cooperating opponent. This means that the optimality condition for effort also changes, and the variable n_t pre-multiplies the second term on the right hand side of both (12) and (13). This has two effects. First, it dampens parental effort (because having a bad kid is now less costly). Second, it introduces a further strategic complementarity also in the educational decision of parents. If parents expect others to increase effort, they realize that their kid is more likely to be matched with a good partner (since n_t is higher). This raises the cost of having a bad kid (since his cheating is more likely to be against a cooperating opponent). Hence, they are induced to exert more effort.

Besides these two effects, the remaining analysis is unaffected (of course, some of the specific conditions discussed above to characterize the equilibrium would also change). But these additional strategic complementarities imply that multiple equilibria are more likely to exist. More generally, reciprocity increases the strategic complementarities between behavior (as captured by the the scope of cooperation, Y_t^1), and norms of generalized morality (as captured by the fraction of good types, n_t). For this reason, a norm of reciprocity also reinforces the effects of changes in the external environment on equilibrium outcomes.

[12] This norm of reciprocity would also add a continuum of other equilibria to those in Proposition 1. In particualr, there would always exist an equilibrium where everyone cheats in any match (or in a subset of matches) just because it expects everyone else to do the same (and hence to bear no cost from cheating). As stated in the text, here we confine attention to the Pareto superior equilibrium.

4 Endogenous Government Policies

If the payoffs of the matching game result from policy choices, different player types might disagree on government policies. When public policies are chosen under majority rule, this creates an additional strategic complementarity. In particular, good types generally prefer better enforcement of cooperation, compared to the bad types. Hence, if the good types are a majority, the government enacts better external enforcement. But the anticipation of better enforcement induces parents to exert more effort into teaching generalized morality to their children. As a result, the fraction of good types in the population increases and might become a majority, just because it is expected to be a majority. Conversely, if parents expect the government to refrain from enforcing cooperation, their incentive to spread generalized morality is diminished, and this expectation might become a political reality.

A similar point is illustrated in an interesting recent paper by Benabou and Tirole (2006) with respect to social insurance policies, in a model where parents conceal information to their kids to overcome a time-inconsistency problem. In this section we illustrate that the same forces are at work in the enforcement of cooperation, even without time inconsistent preferences. Since we have an explicitly dynamic economy, here the interaction between endogenous policies and endogenous preferences has additional implications. Not only we get multiple equilibria, we also obtain histeresis: initial conditions matter, because they might lead the economy to a different steady state. Thus, the interaction of culture and government policies is a source of persistence, which could explain why some economies that started off in political or economic backwardness might remain trapped for ever with poor institutions and adverse cultural traits.

4.1 Two enforcement regimes

To simplify the algebra, throughout this section we assume that $l = w$, so that the matching game does not exhibit strategic complementarity and has a unique equilibrium. We also retain the model as laid out above, without the extension to reciprocity. As discussed in section 2, this implies that the maximum distance that sustains cooperation does not depend on n_t even for the good types, and is given by:

$$Y^k = \left[\ln d - \ln w\right] / \theta^k, \qquad k = 0, 1 \tag{22}$$

In terms of Figure 2, the $Y(n_t)$ curve is vertical. Under this assumption, any strategic complementarity can only arise from the endogeneity of government policy, since for a given policy the equilibrium is unique.

The parameter w (or l) can be interpreted as reflecting external enforcement by the government. A lower value of w corresponds to a smaller temptation not to cooperate and a smaller loss from cooperating against a cheating opponent, and hence better external enforcement. For simplicity, we restrict attention to a binary policy choice, high and low enforcement. We model better external enforcement as a reduced temptation to cheat in matches outside of the safe range $y \leq Y^0$. Specifically, in the regime of low enforcement, the temptation not to cooperate retains the same value w for all matches. In the regime of high enforcement, instead, the temptation not to cooperate drops to $w' = w/\eta < w$ for matches of distance $y \geq \hat{Y}$, with $Y^1 > \hat{Y} > Y^0$. We assume

$$\hat{Y} > [\ln(d) - \ln(w) + \ln(\eta)]/\theta^0 \tag{23}$$

which implies that the bad players (with $k = 0$) continue to find it optimal to cooperate only up to the distance Y_0, even in the high enforement regime. The high enforcement regime, instead, induces the good players to cooperate over a larger range of matches, $Y'^1 > Y^1$, where Y'^1 is still defined by (22) with w' replacing w.

Consider the expected utility of a player who knows his own type and who is about to play the matching game. It is easy to verify that the good types are always better off in the high enforcement regime, since they bear a smaller loss from being cheated (see subsection 8 of the Appendix). This in turn induces the parents to exert more effort to educate their kid in the high than in the low enforcement regime. Let $f\,'$ and f denote educational effort in the high and low enforcement regimes respectively, with $f = F(Y^1)$ as given by (13) above with $w = l$ in it. Note that parental effort no longer depends on time, in either regime, since with $w = l$ effort no longer depends on n_t. Subsection 7 of the appendix proves:

Lemma 7 *$f\,' = f + \Delta$, with $\Delta > 0$*

The bad types, instead, might be better off in either regime, depending on parameter values. We assume that they are better off in the low enforcement regime and subsection 8 of the appendix provides a sufficient condition for this to happend.

The timing of events in each period t is as follows. First, parents choose their educational effort. Then, the kids' types become known and the kids vote over the enforcement regime. Finally, the kids play the matching game. Note that, under this timing, the kids only consider their utility in the current period. When the vote is taken, the fraction of good types is already determined. Thus, current enforcement only affects current payoffs, and it has no effect on future equilibrium outcomes.

4.2 Political-economic equilibrium

Under these assumptions, the political equilibrium is straightfoward. If $n_t > 1/2$, then the high enforcement regime prevails; if $n_t < 1/2$, then the low enforcement regime prevails; and for $n_t = 1/2$ a coin is tossed. The expectation of different enforcement regimes, in turn, induces different educational efforts by the parents. If parents expect $n_t > 1/2$, then they anticipate better enforcement and, by Lemma 6, they exert more effort to inculcate trustworthiness in their kid. And viceversa, if they expect $n_t < 1/2$, they reduce effort.

This set up induces a strategic complementarity in the education decision of the parents, and for some parameter values it can give rise to multiple steady states. Specifically, suppose that parents expect the low enforcement regime to prevail. Then the steady state fraction of good types is given by (19) in the previous section, reproduced here for convenience (with f_s replaced by f):

$$n_s^* = \frac{\delta}{1-f} \tag{24}$$

If $n_s^* < 1/2$, this steady state reproduces itself in a political equilibrium. Suppose instead that parents expect the high enforcement regime. Then, by Lemma 6, the steady state fraction of good types is:

$$n_s'^* = \frac{\delta}{1-f-\Delta} \tag{25}$$

If $n_s'^* > 1/2$, this steady state too reproduces itself in a political equilibrium. Thus, both steady states are possible in equilibrium if $n_s'^* > 1/2 > n_s^*$, or, by (25) and (24), if:

$$\Delta > 1 - 2\delta - f > 0 \tag{A3}$$

which of course requires $\delta < 1/2$.

As already noted, if $w = l$, then the curve $Y(n_t)$ is vertical (i.e., the thresholds of maximal cooperation, Y^1 and Y'^1, do not depend on n_t). Thus, both steady states are always dynamically stable as long as $1 > f + \Delta$, which is implied by (A3). Since $f > 0$, the adjustments to both steady states are also monotonic. Which steady state is reached in equilibrium depends on the initial conditions as well as on the parents' expectations, as we now discuss.

The high enforcement regime is an equilibrium in period t if, given that it is expected, we have $n_t > 1/2$. By (11), this condition can be stated as:

$$n_t = \delta + n_{t-1}(f + \Delta) > 1/2 \tag{26}$$

Similarly, the low enforcement regime is an equilibrium in period t if, given that it is expected, $n_t < 1/2$, namely if :

$$n_t = \delta + n_{t-1} f < 1/2 \tag{27}$$

Combining (26) and (27), we obtain two thresholds, that define which equilibria exist in period t, depending on the fraction of good types in period $t-1$. Specifically, let:

$$\underline{n} = \frac{1 - 2\delta}{2(f + \Delta)} \tag{28}$$

$$\bar{n} = \frac{1 - 2\delta}{2f} \tag{29}$$

with $\bar{n} > \underline{n}$. Then we have:

Lemma 8 *If $n_{t-1} < \underline{n}$, then in period t the unique equilibrium has low enforcement. If $n_{t-1} > \bar{n}$ then in period t the unique equilibrium has high enforcement. If $\bar{n} \geq n_{t-1} \geq \underline{n}$, then both the low and the high enforcement regimes exist as equilibria in period t.*

The proof is straightforward. If n_{t-1} is so low that it falls below the threshold $\underline{n}$, then even if parents expect high enforcement, we have $n_t < 1/2$. Hence high enforcement cannot be a political equilibrium. Conversely, if n_{t-1} is so high that it exceeds the threshold $\bar{n}$, then even if parents expect low enforcement we have $n_t > 1/2$, which rules out low enforcement as an equilibrium. For values of n_{t-1} in between the two tresholds, either regime could win a majority depending on the parents' expectations.

Suppose that condition (A3) is satisfied, so that we have two steady states. Suppose further that both steady states fall outside of the interval $[\underline{n}, \bar{n}]$. Manipulating (28)-(29) and (24)-(25), a sufficient condition for this to happend is:

$$\frac{1-f}{2\delta} - 1 > \Delta > 1 - \frac{f}{1-2\delta} \tag{A4}$$

which in turn requires $\delta \leq 1/4$ (and which also implies (A3)). Since the adjustment towards the steady state is monotonic, then the thresholds $\underline{n}$ *and* $\bar{n}$ define three regions with different dynamics. If the economy starts from an initial condition $n_0 < \underline{n}$, then the equilibrium is unique. The economy remains for ever in the low enforcement equilibrium and it converges to the low enforcement steady state. Conversely, if the economy starts from an initial condition $n_0 > \bar{n}$, then the equilibrium is again unique. The economy remains for ever in the high enforcement equilibrium and it converges to the high enforcement steady state. If the initial condition is in between these two tresholds, $n_0 \in [\underline{n}, \bar{n}]$, then both paths are feasible equilibria, and the economy eventually ends up in one or the other steady state depending on expectations.

If condition (A4) is violated, then one of the steady states (or both) are inside the region where multiple equilibria are possible. In this case eventually the economy might end up in the region of multiple equilibria, and one or the other steady state will be reached depending on expectations (if both inequalities in (A4) are violated then both steady states are inside the region of multiple equilibria and the economy certainly reaches this region in finite time for any initial conditions).

We summarize the foregoing discussion in the following.

Proposition 9 *If condition (A3) holds, then the economy has two steady states, one with high external enforcement and where the good players are a majority; and one with low external enforcement and where the good players are a minority. Both steady states are dynamically stable. If condition (A4) also holds, and if the initial fraction of good players, n_0, is outside of the interval $[\underline{n}, \bar{n}]$, then the equilibrium is unique. For $n_0 < \underline{n}$ (for $n_0 > \bar{n}$), the economy remains always under the low (high) enforcement regime and eventually reaches the low (high) enforcement steady state. If condition (A4) is violated, then multiple equilibria are possible during the adjustment path towards one or the other steady states.*

4.3 Discussion

This Proposition can explain why distant historical circumstances have such long lasting effects, and why some countries or societies may remain trapped in cultural, institutional and economic backwardness. Despotic leaders that abuse of their citizens or don't enforce the rule of law are likely to disseminate adverse cultural traits in the community. Such traits then influence the political choices of citizens once the autocrat is replaced by democratic institutions. Even if the country becomes a democracy, it retains weak institutions because adverse cultural traits induce a preference for policies that stifle cooperation. Better institutions are available, and nothing prevents the country from adopting them, but this does not happend in a political equilibrium. The majority prefers to retain worse institutions because it benefits from them, since it can take advantage of those few cooperating citizens who strongly dislike cheating. Whether cheating corresponds to tax evasion, or to free riding on public transportations, or on the public welfare system, the institutions that allow weak law enforcement are chosen and preferred by a majority of citizens themselves. This cultural explanation of institutional persistence is quite different from others suggested in the literature, that emphasize the power of the élites against the will of the citizens at large (eg. Acemoglu and Robinson 2006).

Note that the presence of at least some citizens who strongly value cooperation and who are occasionally exploited by other more shrewed players is not necessary for this result. Even if almost everyone ends up with a low value for cooperation, better external enforcement would still be opposed if it costs resources. The reason is that the benefits of better enforcement would be negligible in a society where trust and cooperation are so low that many mutually advantageous trade opportunities are foregone anyway.

The importance of initial conditions also implies that temporary shocks might have permanent effects, if they move the economy from one equilibrium trajectory to another. Suppose for instance that the economy is on the path that leads to the better steady state, with n_t above the region of multiple equilibria. But suppose that, because of a war, or of an inept political leader, external enforcement temporarily deteriorates. If the deterioration of the external environment lasts long enough to have an impact on individual cultural traits, the fraction of good types might move outside the region of multiple equilibria, and the economy and the institutions never recover, even when the external circumstances are changed.

Finally, this example has assumed that individuals vote or act politically according to their self interest. Here cultural traits influence political preferences only because they have implications for how individuals are affected by the quality of external enforcement. If instead individual values also have a direct impact on political acts, as seems plausible, then this would open up an additional channel through which the external environment interacts with individual values, and this could reinforce the results presented above.[13] We have also assumed that the parents generation does not vote and does not particpate in the matching game. Relaxing these assumptions would increase the sources of persistence described in this and in the previous sections.

5 Some evidence

As already anticipated, the theoretical results of the previous sections are consistent with the evidence in Tabellini (2005), that a history of political backwardness is associated with current lack of social capital and institutional malfunctioning across and within countries. This section presents additional and more direct evidence that distant history shapes individual values and beliefs, as they are passed from one generation to the next.

One of the central results of the model is that the values of currently alive individuals reflect the external environment in which earlier generations were living. The so called "epidemiological" approach used in several recent papers is one way to test this result (Fernandez 2007 reviews the relevant empirical literature, and Guiso, Sapienza and Zingales 2006 and Algan and Cahuc 2006, 2007 have applied the same approach to closely related issues). This exploits data on the opinions of immigrants, to test whether they reflect features of the country of origin of their ancestors.

To measure the attitude of individuals towards cooperation, I use the measure of trust that many other empirical papers have studied before, namely the answer to the question of how much the respondent thinks he can trust others. This is not really a measure of trustworthiness, but of trusting others. Nevertheless, it is strongly correlated with more precise indicators of trustworthiness in opinion polls where both measures are available, and it

[13] Alesina and Angeletos (2006) consider a model where individuals vote according to their self interest and also according to a notion of what is fair and unfair. In their model, however, individual values are exogenously given and do not interact with the economic environment.

has the advantage that it is widely available in many surveys.

I use the US General Social Surevey (GSS), that contains data on trust, as well as other relevant information on the respondent. My sample consists of about 4300 third generation immigrants to the US, namely individuals born in the US who have at least two grand-parents born outside of the US. To classify where these immigrants come from, I rely on a question that asks what is the ancestors' country of origin.[14] I mainly focus on countries that originated at least 25 individuals in the sample, but the results are robust to including a larger or smaller set of countries of origin. Table 2 lists the relevant coutries and how many respondents in the sample originated from each country. With the exception of Russia and Mexico, the set of countries is fairly homogenous in terms of current level of development, though not in terms of political and economic history in the distant past.

Do the current attitudes of these third generation immigrants reflect historical or current features of their ancestors' country of origin? To answer this question, I estimate a probit model, where the dependent variable is *trust*, defined as one if the respondent thinks that most people can be trusted, and zero otherwise (don't know or no answers are omitted). Throughout I control for several features of the respondent, such as his gender, income, education, employment status, age, religion, whether he is married or has children, the education of his parents, whether he lives in a large metropolitan area, the number of grand parents born abroad, and a set of dummy variables for his area of residence and for the decade in which the survey was carried out (the surveys span the period 1977-2004, since before this period the question on the birthplace of the respondent was not included in the questionnaire) - see the notes to Table 3 for a complete list. Several but not all of these variables are statistically significant, and in particular income, gender, employment status, age and the mother's education. Controlling for this long list of individual attributes should make it likely that, if we find an effect on individual trust of variables measuring features of the country of origin of ancestors, it reflects transmitted cultural traits, rather than economic variables such as income or human capital.

For this type of probit regression, Table 3 shows the estimated coefficients of variables that measure alternative features of the ancestors' country

[14] Unfortunately the GSS survey does not ask about the country of origin of grandparents, but of the more vaguely defined ancestors. Since this is probably interpreted by the respondent as reflecting the ancestors who had more influence on his family history, it need not introduce much measurement error.

of origin (in other words, the variables reported in Table 3 only vary across different countries of origin of the ancestors). Standard errors are clustered by ancestors'country, to allow for arbitrary patterns of correlation of the residuals by ancestors' country. Column 1 of Table 3 reproduces a finding already discussed in recent papers by Guiso, Zinagales and Sapienza (2006) and by Algan and Cahuc (2006), (2007) on a slightly different sample of GSS respondents. Namely, trust attitudes of US immigrants are strongly correlated with average trust in the ancestors' country, as measured by the latest World Value Surveys conducted around the year 2000 (the variable *Trust from 2000 WWS*). This is already important evidence of the transmission of cultural traits across generations. Third generation US immigrants have had time to adapt to their new environment, that certainly differs from that of their ancestors.

Columns 2 and 3 replace contemporaneous trust in the ancestor's country with a measure of historical political institutions in the country origin. The variables *Constr. Exec up to 1900* and *Polity2 up to 1900* are the first principal component of the variables Constraints on the Executive and Polity2 in the Polity IV data set, measured in the years 1850, 1875 and 1900. Higher values correspond to stronger constraints on the executive or more democratic political institutions. Both variables are highly statistically significant and show that trust is higher in third generation US immigrants that come from countries that over a century ago had better political institutions. Similar results are obtained if political institutions are sampled in different years or are aggregate in different ways.

Could this result be due to the fact that immigrants from countries with better political institutions were richer, and this in turn increases the stock of wealth of currently alive respondents ? Since we are already controlling for individual income and education, and for parental education, this does not seem very likely. But to allow for this possibility, columns 4 and 5 add as a regressor per capita income in the ancestors' country in 1870 and in 1930 respectively (the source is Maddison 2001).[15] Political history remains significant and its coefficient does not vary, while per capita income in the country of origin has a positive estimated coefficient which however is not statistically significant.

Column 6 adds a historical measure of education in the country of origin,

[15] Adding this variable implies that we loose immigrants from Russia and the former socialist countries in Eastern Europe, for which the Maddsion data are not available.

namely primary school enrollment in 1910 (*Primary school enr. 1910*), taken from Benavot and Riddle (1988). Since education is likely to foster trust, immigrants that were coming from countries with higher school enrolment on average were likely to be better educated. We expect that this cultural feature is transmitted to subsequent generations and thus we expect this variable to have a positive effect on trust of current generations. This is what we find: the estimated coeffiecient of this variable is statistically significant, and the estimated coefficient on constraints on the executive remains positive and significant.

Finally, the last column of Table 3 reports the full specification, with variables measuring political history, past education, past income and contemporaneous trust in the country of origin. All variables except per capita income are statistically significant and with the expected sign. Thus, the historical variables of the ancestors' country of origin contribute to explain the attitides of third generation immigrants, but so does current average trust in the country. A plausible interpretation of this finding is that national culture is determined by more than the sparce historical variables included in the regression (hence current average trust retains its staistical significance); but at the same time a country's history has additional explanatory power besides current trust, because the latter also reflects more recent events that could not influence the cultural traits of earlier generations and hence of currently alive US respondents.

These national variables explain a signficant fraction of the of current trust of US respondents, averaged by country of origin. Replacing the national variables reported in Table 3 with dummy variables for the ancestors' origins, we can estimate the average effects of different ancestors origin. If we regress the estimated coefficients of these dummy variables on the three historical variables reported in column 6 (weighting observations by the number of third generation US immigrants from each country), we reproduce similar results to those reported in Table 3, and these three variables explain up to 63% of the variance in the estimated coefficients. Adding current average trust from the WWS surveys, as in column 7, the fraction of variance explained goes up to over 77%. Figure 3, that plots the estimated coefficients of the dummy variables against all the country of origin regressors included in column 7 illustrates that the estimates are not due to any outlier observations.

Of course, we have to be aware of the data limitations. First, the number of ancestors' country is small (from 21 to 17, depending on the specifica-

tion). Second, immigrants are certainly not a representative sample of the underlying population. Third, although we can explain most of the variation in country averages, the individual responses display a lot of unexplained variation. Nevertheless, none of these problems suggests that the estimated coefficients are likely to be biased upwards. On the contrary, to the extent that many of the national variables meant to capture national cultural traits are measured with error, their estimated coefficients would be biased towards zero.

6 Concluding remarks

Economic backwardness is typically associated with a large range of institutional, organizational and government failures, simultaneously along many dimensions. In many stagnating countries or regions, politicians are ineffective and corrupt, public goods are under-provided, law enforcement is inadequate, corruption and moral hazard is widespread inside public and private organizations. There is not just one institutional failure. Typically, the countries or regions that fail in one dimension also fail in many other aspects of collective behavior. And conversely, when institutions or organizations function well, they do so along many dimensions. Moreover, institutional failures or successes do not just reflect the design of public policies; they also result from the behavior of public officials, or of private individuals inside private or public organizations.

Economists, unlike other social scientists, have until recently refrained from relying on a cultural explanation of these patterns, arguing that culture is endogenous and hard to pin down. This paper has studied a model in which cultural traits are transmitted by forward looking and rational parents, who adapt their educational choices to the environment in which their children will live. If the external environment gurantees better enforcement of cooperation, both norms and individual behavior adapt with mutually reinforcing effects through strategic complementarities. But parents evaluate their children's welfare with their own values, and this creates cultural persistence. In equilibrium, the range of situations in which individuals cooperate also depends on the external environment that prevailed several generations ago. A history of respect for the rule of law and effective law enforcement encourages current cooperation because it facilitates the diffusion of norms of generalized (as opposed to limited) morality. If the current law enforce-

ment regime is also endogenous, the forward looking behavior of parents gives rise to an additional strategic complementarity between current educational choices and future political decisions, and the equilibrium displays histeresis. Parents refrain from encouraging their children to cooperate with others if they anticipate that such cultural trait will remain minoritarian. Hence, cultural persistence induces persistence of formal institutions, and a country that starts out with inefficient institutions and adverse cultural traits retains them for ever. The aggregate evidence discussed in the literature, and the microeconomic evidence presented in this paper, are consistent with this phenomenon.

Much remains to be done, to pin down more precisely the channels of cultural transmission both inside and outside of the family, to understand the role of learning and formal education, and to study empirically the relevance of specific cultural traits. But the general issue of how individual norms of conduct evolve and interact with the economic and political environment is a very exciting area of research, that can be fruitfully studied with the standard tools of economic analysis.

7 Appendix

7.1 Proof of Proposition 1

Consider first the simplest case in which $w = l$. In this case the net material gain of playing NC does not depend on the strategy played by the opponent. Then the threshold $\tilde{y}$ that leaves the player indifferent between cooperating or not simplifies to $\tilde{y} = Y' = Y$. The proposition then immediately follows by the definition of $\tilde{y}$ and the fact that the cost of cheating, $de^{-\theta y}$, is strictly decreasing in y, while the temptation $T(\pi(y))$ equals w for all $\pi(y)$. Here each player has a simple dominant strategy and the equilibrium is unique.

If $w \neq l$, then the optimal choice of each player depends on his beliefs about what his opponent will do, and some matches entail multiple equilibria. In particular, consider the threshold of indifference, $\tilde{y}$, in (4). Replacing $\pi(y)$ with 0 and 1 respectively, equation (4) yields $\tilde{y} = Y'$ and $\tilde{y} = Y$, as defined in (5) and (6). Parts (i) and (ii) of the proposition follow again immediately from the definition of $\tilde{y}$, and the fact that the cost of cheating, $de^{-\theta y}$, is strictly decreasing in y, while the temptation $T(\pi(y))$ does not depend on y, holding π constant. But if $y \in [y_{Min}, y_{Max}]$, then multiple equilibria are

possible, depending on the value of $\pi(y)$.

Specifically, consider first the case $w < l$, so that $Y > Y'$. Suppose that $y \in [Y', Y]$. If the opponent is expected to cooperate ($\pi(y) = 1$), then $\tilde{y} = Y > y$, so reciprocal cooperation is a best response. While if the opponent is expected not to cooperate ($\pi(y) = 0$), then $\tilde{y} = Y' < y$, so reciprocal non-cooperation becomes a best response.

Next, consider the case $w > l$, so that $Y < Y'$. Suppose that $y \in [Y, Y']$. If the opponent is expected to cooperate ($\pi(y) = 1$), then $\tilde{y} = Y < y$, so non-cooperation is a best response. While if the opponent is expected not to cooperate ($\pi(y) = 0$), then $\tilde{y} = Y' > y$, so now cooperation becomes a best response. Hence, here the two matched players find it optimal to play opposite strategies. In this case, there is also a symmetric equilibrium where both players play the same mixed strategy. QED

7.2 Expected Utility in the Equilibrium of Proposition 2

Here we write down the players' expected utility in the Pareto superior equilibrium summarized in Proposition 2, letting n_t be indexed by time. For those with $\theta = \theta^0$, (10) simplifies to:

$$U_t^0 = \frac{\alpha}{2\left[1 - e^{-\alpha S}\right]} \left[c \int_0^{Y^0} e^{-\alpha z} dz + (c + w) n_t \int_{Y^0}^{Y_t^1} e^{-\alpha z} dz \right] \tag{30}$$

The first term on the right hand side of (30) corresponds to the expected material benefit in a match within the safe distance where both partners always cooperate; the second term is the expected outcome in the intermediate area where only the good cooperate, while the bad types play non-cooperatively. The expected utility of those with $\theta = \theta^1$ instead is:

$$U_t^1 = \frac{\alpha}{2\left[1 - e^{-\alpha S}\right]} \left[c \int_0^{Y^0} e^{-\alpha z} dz + \left[c n_t - l(1 - n_t)\right] \int_{Y^0}^{Y_t^1} e^{-\alpha z} dz \right] \tag{31}$$

where the first term on the right hand side of (31) continues to have the same interpretation, while the second term is the expected outcome, given that only the good types cooperate in the intermediate distance range.

Note that, (8) implies:

$$de^{-\theta^1 Y_t^1} = [l + (w - l)n_t)] \tag{32}$$

Hence, (30) and (31) imply:

$$\begin{aligned} U_t^1 - U_t^0 &= \frac{-\alpha[l + (w - l)n_t)]}{2\left[1 - e^{-\alpha S}\right]} \int_{Y^0}^{Y_t^1} e^{-\alpha z} dz = \\ &= \frac{-\alpha d e^{-\theta^1 Y_t^1}}{2\left[1 - e^{-\alpha S}\right]} \int_{Y^0}^{Y_t^1} e^{-\alpha z} dz < 0 \end{aligned} \tag{33}$$

where the last equality follows from (32).

7.3 Proof of Lemma 3

Here we omit the time indexes since they are redundant. Consider the solution to the problem of maximizing V^{pk}, as defined in (9), by choice of θ^k. As discussed in the text, θ^k enters the expression for V^{pk} only through the distance threshold Y^k that triggers non-cooperation by the kid. Hence, by (10) and (9), differentiating V^{pk} with respect to θ^k and rearranging, we have:

$$\frac{\partial V^{pk}}{\partial \theta^k} = \frac{\alpha e^{-\alpha Y^k}}{2\left[1 - e^{-\alpha S}\right]} \frac{\partial Y^k}{\partial \theta^k} \left\{ d e^{-\theta^p Y^k} - \left[(w - l)\pi(Y^k) + l\right] \right\} \tag{34}$$

By (7) and (8), $\frac{\partial Y^k}{\partial \theta^k} < 0$. Hence, the optimal value of d^k is such that Y^k solves the expression

$$de^{-\theta^p Y^k} = [(w - l)\pi(Y^k) + l] \tag{35}$$

for $\pi(Y^k)$ corresponding to the equilibrium probability of cooperation by a partner located at distance Y^k. But by (4), this implies $\theta^k = \theta^p$. Hence the parent strictly prefers to have a kid with his own value parameter. QED

7.4 Proof of Lemma 5

Differentiating the RHS of (13) with respect to Y_t^1 and simplifying, we have:

$$F_{Y_t^1} = \frac{\varphi \alpha d}{2\left[1 - e^{-\alpha S}\right]} \left[\theta^1 e^{-\theta^1 Y_t^1} \int_{Y^0}^{Y_t^1} e^{-\alpha z} dz \right] > 0 \tag{36}$$

QED

7.5 Slope of the functions $n_t = N(Y_t^1, n_{t-1})$ and $Y_t^1 = Y(n_t)$

Equation (15) implies that $N_{Y_t^1} = n_{t-1} F_{Y_t^1} > 0$

Differentiating the RHS of (14) with respect to n_t, we also have:

$$Y_{n_t} = \frac{1}{\theta^1} \frac{l-w}{x_t} > 0 \tag{37}$$

where $x_t = l + (w-l)n_t \geq 0$ (since $n_t \leq 1$). The sign Y_{n_t} follows from $l > w$.

The function $N(Y_t^1)$ intersects the function $Y(n_t)$ from left to right, as drawn in Figure 2, if $N_{Y_t^1} < 1/Y_{n_t}$, or, by (35), (36) and (37), if:

$$\frac{1}{l-w} > \frac{\varphi \alpha n_{t-1}}{2(1-e^{-\alpha S})} \left[\int_{Y^0}^{Y_t^1} e^{-\alpha z} dz \right] = \frac{\varphi n_{t-1}}{2(1-e^{-\alpha S})} [(\frac{w}{d})^{\lambda^0} - (\frac{x_t}{d})^{\lambda^1}] \tag{38}$$

where $\lambda^k = \alpha/\theta^k$ and where the last equality follows from (7) and (8). Note that $\frac{n_{t-1}}{2(1-e^{-\alpha S})} < 1$ and $x_t \geq w$. Thus, a sufficient condition for (38) to hold is:

$$\frac{1}{l-w} > \varphi \left[(\frac{w}{d})^{\lambda^0} - (\frac{w}{d})^{\lambda^1} \right] \tag{A1}$$

which is certainly satisifed if $l - w$ or φ are sufficiently small - recall that $d > w$ and that $\lambda^1 > \lambda^0$, so that the right hand side of (**??**) is strictly positive.

7.6 Dynamic stability of the steady state

Applying the implicit function theorem to (16) and (17), we have:

$$\begin{aligned} \frac{dn_t^*}{dn_{t-1}} &= \frac{N_{n_{t-1}}}{1 - Y_{n_t}.N_{Y_t^1}} \\ \frac{dY_t^{1*}}{dn_{t-1}} &= \frac{dn_t^*}{dn_{t-1}} Y_{n_t} \end{aligned}$$

Under (A1), $1 - Y_{n_t}.N_{Y_t^1} > 0$ (see the previous subsection of the appendix). Moreover:

$$N_{n_{t-1}} = f_t > 0$$

where the inequality follows from Corollary 4. Hence, (A1) implies $\frac{dn_t^*}{dn_{t-1}} > 0$ and (since $Y_{n_t} > 0$) $\frac{dY_t^{1*}}{dn_{t-1}} > 0$.

To prove that $\frac{dn_t^*}{dn_{t-1}} < 1$, we need to prove that $N_{n_{t-1}} < 1 - Y_{n_t}.N_{Y_t^1}$, or equivalently, that

$$f_t + Y_{n_t}.N_{Y_t^1} < 1 \tag{39}$$

By (36) and (37), the term $Y_{n_t}.N_{Y_t^1}$ is proportional to φ. By (13), the term f_t is also proportional to φ. Define $\bar{\varphi}_t$ as the value of φ such that $f_t + Y_{n_t}.N_{Y_t^1} = 1$. Note that $\bar{\varphi}_t$ depends on t through the terms x_t and n_{t-1} (cf. (37), (36) and (13)). Define $\bar{\varphi} = \arg\min(\bar{\varphi}_t)$, where the minimization is taken over all feasible values of n_{t-1} and x_t. Since $N_{n_{t-1}} > 0$, $Y_{n_t} > 0$ and $N_{Y_t^1} > 0$, then $\bar{\varphi} > 0$. Then, for any $0 < \varphi < \bar{\varphi}$, (39) also holds. QED

7.7 Effect of a more localised economy (higher α)

Taking the derivative of f_t in (13) with respect to α and rearranging, we have:

$$\frac{\partial f_t}{\partial \alpha} = \frac{\varphi d}{2\left[1 - e^{-\alpha S}\right]} \left[\int_{Y^0}^{Y_t^1} e^{-\alpha z}(e^{-\theta^1 z} - e^{-\alpha Y_t^1})\Psi(z)dz \right] \tag{40}$$

where the function $\Psi(z)$ is:

$$\Psi(z) = 1 - e^{-\alpha S}(1 + \alpha S) - \alpha z(1 - e^{-\alpha S}) \tag{41}$$

Let $\bar{Y}$ be such that $\Psi(\bar{Y}) = 0$. By (41), we obtain that $\bar{Y}$ is defined as in (20). Since $\Psi_z(z) < 0$, then $Y_t^1 < \bar{Y}$ implies $\Psi(Y_t^1) > 0$, which in turn implies $\Psi(z) > 0$ for any $z \in [Y^0, Y_t^1]$. By equation (40), then, $Y_t^1 < \bar{Y}$ implies $\frac{\partial f_t}{\partial \alpha} > 0$. Conversely, $Y^0 > \bar{Y}$ implies $\Psi(Y^0) < 0$, which in turn implies $\Psi(z) < 0$ for any $z \in [Y^0, Y_t^1]$. By equation (40), then, $Y^0 < \bar{Y}$ implies $\frac{\partial f_t}{\partial \alpha} < 0$.

Finally, consider $\bar{Y} - S$. By (41), we have $\bar{Y} - S < 0$ if $1 - e^{-\alpha S} < \alpha S$, which is true for any $\alpha S > 0$ (as can be seen by differentiating both sides with respect to αS). Hence, $\bar{Y} < S$.

7.8 Proof of Lemma 6

Let a $'$ denote the high enforcement regime. In this regime, the difference in the expected material payoffs of a good and a bad kid can be written as (cf.

subsection 2 of the appendix):

$$U_t'^1 - U_t'^0 = \frac{-\alpha}{2\left[1-e^{-\alpha S}\right]}\left[w\int_{Y^0}^{\hat{Y}} e^{-\alpha z}dz + w'\int_{\hat{Y}}^{Y'^1} e^{-\alpha z}dz\right] =$$

$$= \frac{-\alpha}{2\left[1-e^{-\alpha S}\right]}\left[w\int_{Y^0}^{Y^1} e^{-\alpha z}dz + w'\int_{Y^1}^{Y'^1} e^{-\alpha z}dz - (w-w')\int_{\hat{Y}}^{Y^1} e^{-\alpha z}dz\right] \quad (42)$$

Repeating the previous analysis, parental effort is given by the following first order condition, adapted from (12):

$$f\,'/\varphi = (U'^1 - U'^0) + \frac{\alpha d}{2\left[1-e^{-\alpha S}\right]}\int_{Y^0}^{Y'^1} e^{-(\alpha+\theta^1)z}dz \quad (43)$$

Combining (42) and (43), and exploiting (13) for $w = l$, we have:

$$f_t' = f_t + \Delta$$

where

$$\Delta = \frac{\varphi\alpha}{2\left[1-e^{-\alpha S}\right]}\left\{(w-w')\int_{\hat{Y}}^{Y^1} e^{-\alpha z}dz - w'\int_{Y^1}^{Y'^1} e^{-\alpha z}dz + d\int_{Y^1}^{Y'^1} e^{-(\alpha+\theta^1)z}dz\right\} \quad (44)$$

We also have:

$$\frac{\partial\Delta}{\partial w'} = \frac{\varphi\alpha}{2\left[1-e^{-\alpha S}\right]}\left\{-\int_{\hat{Y}}^{Y'^1} e^{-\alpha z}dz - w'e^{-\alpha Y'^1}\frac{\partial Y'^1}{\partial w'} + de^{-(\alpha+\theta^1)Y'^1}\frac{\partial Y'^1}{\partial w'}\right\} =$$

$$= \frac{-\varphi\alpha}{2\left[1-e^{-\alpha S}\right]}\int_{\hat{Y}}^{Y'^1} e^{-\alpha z}dz < 0$$

where the second equality follows from $d = w'e^{\theta^1 Y'^1}$. Note that when $w' = w$, we have $Y'^1 = Y^1$ so that $\Delta = 0$, by (44). Thus, $w' < w$ implies $\Delta > 0$. QED

For later use, it is convenient to simplify the expression on the right hand side of (44), to obtain:

$$\Delta = \frac{\varphi\alpha d^{-\lambda}\hat{w}^{1+\lambda}}{2\left[1-e^{-\alpha S}\right](\alpha+\theta^1)}\left\{\theta^1(w^{1+\lambda}-w'^{1+\lambda})/\hat{w}^{1+\lambda}-(1+\theta^1)(w-w')/\hat{w}\right\} \tag{45}$$

where $\lambda = \alpha/\theta^1$ and $\hat{w} = de^{-\theta^1\hat{Y}}$.

7.9 Policy preferences

Here we show that the good players always prefer the high enforcement regime, and we provide a sufficient condition that guarantees that the bad players prefer the low enforcement regime.

Let $V_t'^1$ denote the expected utility of the good types under the high enforcement regime. Exploiting the envelope theorem and adapting (31) and (9) to the high enforcement regime, we have:

$$\frac{\partial V_t'^1}{\partial w'} = -\frac{\alpha}{2\left[1-e^{-\alpha S}\right]}(1-n_t)\int_{\hat{Y}}^{Y'^1} e^{-\alpha z}dz < 0$$

where Y'^1 is their maximum threshold for cooperation under this regime. Since this expression holds for any w', the good types are always in favor of the high enforcement regime, the more so the smaller is w' relative to w.

Next, consider the bad types. Since the threshold Y^0 is not affected by the regime, the difference in their overall expected utility under the two regimes is:

$$U_t'^0 - U_t^0 = \frac{\alpha n_t}{2\left[1-e^{-\alpha S}\right]}\left[(c+w')\int_{Y^1}^{Y'^1} e^{-\alpha z}dz + (w'-w)\int_{\hat{Y}}^{Y^1} e^{-\alpha z}dz\right] \tag{46}$$

Exploiting the fact that $e^{-\alpha Y^1} = (\frac{w}{d})^\lambda$ and that $e^{-\alpha Y'^1} = (\frac{w'}{d})^\lambda$, where $\lambda = \alpha/\theta^1$ and $\hat{w} = de^{-\theta^1\hat{Y}}$, and after some algebra, it is possible to show that the expression on the right hand side of (46) is strictly positive if:

$$c\frac{w^\lambda - (w')^\lambda}{w-w'} + \frac{w^{1+\lambda}-(w')^{1+\lambda}}{w-w'} > \hat{w}^\lambda \tag{A2}$$

which is certainly satisfied if c is large enough, or if $\lambda \geq 1$ and $w - w'$ is large enough.

References

[1] Acemoglu, D., S. Johnson and J. A. Robinson (2001), "The colonial origins of Comparative Development : An Empirical Investigation", *American Economic Review*, 91:1369-1401.

[2] Acemoglu, D. and J. A. Robinson (2006), "Persistence of Power, Elites and Institutions", mimeo, MIT

[3] Akerlof, G. A. and Kranton, R. E., "Economics and Identity," *Quarterly Journal of Economics* CXV(3), August 2000, pp. 715-733.

[4] Akerlof, G. A. and Kranton, R. E., (2006), "*Economics and Identity*", Mimeo, Berkeley.

[5] Alesina, A. and Angeletos, M. (2005), "Fairness and Redistribution", *American Economic Review*, September 95:4.

[6] Algan, Y. and Cahuc, P. (2006), "Why Is The Minimum Wage So High In Low-Trust Countries?", Mimeo, University of Paris 1.

[7] Algan, Y. and Cahuc, P. (2007), "Social Attitudes and Macroeconomic Performance: an Epidemiological Approach", Mimeo, University of Paris 1

[8] Amir, R. (2003), "Supermodularity and Complementarity in Economics: An Elementary Survey", mimeo, CORE.

[9] Axelrod, R. (1984), "*The Evolution of Cooperation*", New York: Basic Books

[10] Banfield, E. C. (1958), "*The Moral Basis of a Backward Society*", New York : The Free Press.

[11] Becker, Gary S. (1993) "Nobel Lecture: The Economic Way of Looking at Behavior", *The Journal of Political Economy*, Vol.101, No.3. (Jun.1993), pp.385-409.

[12] Becker, Gary S. (1996), "*Accounting for Tastes*", Harvard U. P.

[13] Benabou, R. and Tirole, J. (2006), "Belief in a Just World and Redistributive Politics", *Quarterly Journal of Economics*, 121(2), May 2006, 699-746.

[14] Benabou, R. and Tirole, J. (2006), "Incentives and Prosocial Behavior", *American Economic Review*, December 2006.

[15] Benavot, A. and P. Riddle(1988) "The Expansion of Primary Education: 1870-1940", *Sociology of Education* 61 (3): 191-210

[16] Bisin, A., G. Topa and T. Verdier. (2004), "Cooperation as a transmitted cultural trait", *Rationality and Society*, 16, 477-507

[17] Bisin, A. and Verdier, T. (2001), "The Economics of Cultural Transmission and the Dynamics of Preferences", *Journal of Economic Theory*, 97, 289-319

[18] Bisin, A. and Verdier, T. (2005), "Cultural Transmission", *The New Palgrave Dictionary of Economics*

[19] Bowles, S. and Gintis, H. (2000), "Social Capital and Community Governance", *Economic Journal* 112, 2002, 419-436.

[20] Dixit, A. (2004), "*Lawlessness and Economics - Alternative Models of Governance ",* Princeton University Press, Princeton

[21] Doepke, Matthias and Zilibotti, Fabrizio (2005), "Patience Capital and the Demise of the Aristocracy", mimeo, April 2005, UCLA.

[22] Dohmen, T., A. Falk, D. Huffman and U. Sunde (2006) "The Intergenerational Transmission of Risk and Trust Attitudes", mimeo, IZA

[23] Fehr, E., Fischbacher, U. and Gachter S. (2002), "Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms", *Human Nature*, 13 (2002):1-25.

[24] Fernadez, R. (2007), "Culture, Women, and Work", *Journal of the European Economic Association*, forthcoming

[25] Fisman, Ray and Miguel, Edward (2006), "Cultures of Corruption: Evidence from Diplomatic Parking Tickets", mimeo, Columbia University and NBER.

[26] Francois P. (2006), "Norms and the Dynamics of Institution Formation", mimeo, University of British Columbia

[27] Francois P. and J. Zabojnik (2005) "Trust, Social Capital & Economic Development", *Journal of the European Economic Association, 3(1) 51-94*

[28] Gambetta, Diego (ed.) (1988), "*Trust: Making and Breaking Cooperative Relations*", Oxford: Basil Blackwell

[29] Giuliano, P. (2007), "On the Determinants of Living Arrangements in Western Europe: Does Cultural Origin Matter?", *Journal of the European Economic Association*, forthcoming

[30] Glaeser, E., R. La Porta, and F. Lopez-de-Silanes and A. Schleifer (2004), "Do Institutions cause Growth", NBER W. P. 10568.

[31] Guiso L., Sapienza P. and Zingales L. (2006), "Does Culture Affect Economic Outcomes?," *Journal of Economic Perspectives,* Spring

[32] Hauk E. and M. Saez-Mart´ı (2002), "On the Cultural Transmission of Corruption", *Journal of Economic Theory* 107(2), 311-335

[33] Levy G. and Razin, R. (2006), "A Theory of Religion: Linking Individual Beliefs, Rituals, and Social Cohesion", Mimeo, London School of Economics.

[34] Lindbeck, A. and S. Nyberg (2006), "Raising Children to Work Hard: Altruism, Work Norms and Social Insurance", *Quarterly Journal of Economics*, November

[35] Maddison, A. (2001), "*The World Economy: Historical Statistics*", OECD.

[36] Nisbet, Richard E. and Cohen, Dov (1996), "*Culture of Honor - The Psychology of Violence in the South*", Westview Press.

[37] Platteau, J.P. (2000), "*Institutions, Social Norms, and Economic Development*",Harwood: Academic Publishers & Routledge.

[38] Rajan, R. (2006), " The Persistence of Underdevelopment: Constituencies and Competitive Rent Preservation", mimeo, University of Chicago

[39] Rajan, R. and L. Zingales (2006), "The Persistence of Underdevelopment: Institutions, Human Capital or Constituencies?", NBER working paper 12093

[40] Tabellini, G. (2005), "Culture and Institutions: Economic Development in the Regions of Europe", Mimeo, Bocconi University.

Table 2 – Country of origin of Ancestors

Country of origin	N. of observations	Frequency (%)
Austria	69	1.60
Canada	200	4.63
Denmark	51	1.18
Finland	46	1.06
France	63	1.46
Germany	834	19.31
Greece	38	0.88
Hungary	75	1.74
Ireland	485	11.23
Italy	668	15.46
Lithuania	40	0.93
Mexico	252	5.83
Netherlands	94	2.18
Poland	376	8.70
Portugal	30	0.69
Russia	162	3.75
Slovak Republic	149	3.45
Slovenia	45	1.04
Spain	40	0.93
Sweden	153	3.54
United Kingdom	450	10.42
Total	4320	100

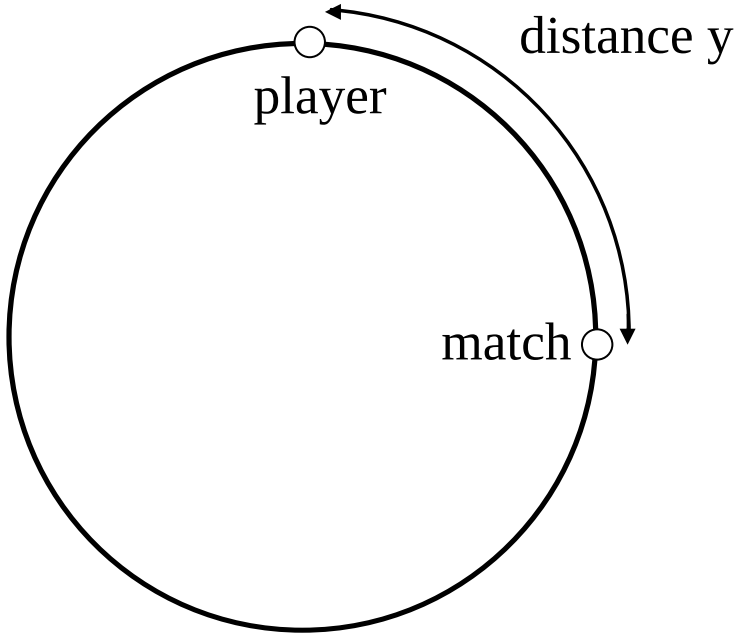
Table 3. Trust and ancestors' countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable				***trust***			
Trust from 2000 WWS	0.63						0.75
	(0.21)***						(0.19)***
Constr. Exec. up to 1900		0.06		0.05	0.06	0.05	0.06
		(0.02)***		(0.02)**	(0.03)**	(0.02)***	(0.01)***
Polity2 up to 1900			0.05				
			(0.02)***				
Per capita income 1870				0.09			
				(0.09)			
Per capita income 1930					0.08	-0.04	-0.15
					(0.10)	(0.12)	(0.08)*
Primary school enr. 1910						0.50	0.50
						(0.22)**	(0.15)***
Observations	4320	4320	4320	3969	3969	3593	3593
Pseudo R-squared	0.03	0.03	0.03	0.03	0.03	0.04	0.04

Robust standard errors in parentheses, clustered by country of origin of ancestors
* significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include the following controls:
Gender; family income in constant dollar (base=1986); dummy variables if completed high school, if completed college, if working, if unemployed, for age over 65, for age under 25, if married, for having at least one child, if catholic, if protestant, if jewish, if father attended primary school, if mother attended primary school, if father attended college, if mother attended college, for living in urban area; number of grandparents born outside US; dummy variables for survey's decade (1980s, 90s or after year 2000); dummy variables for region of residence (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain or Pacific).

Figure 1



prob. of match: $\sim e^{-\alpha y}$

Figure 2

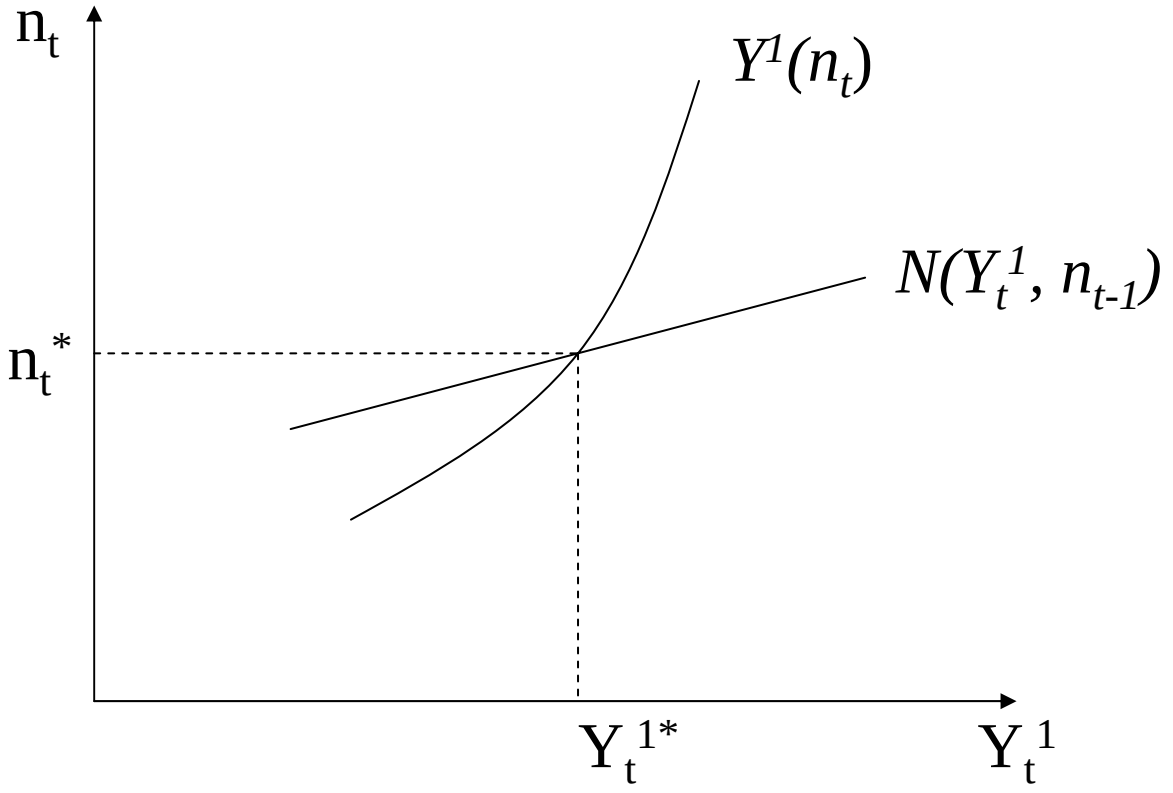


Figure 3 – Trust and ancestors' country of origin