Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

The Scope of Cooperation: values and incentives

Guido Tabellini

Working Paper n. 328

October 2007

IGIER – Università Bocconi, Via Salasco 5, 20136 Milano –Italy
http://www.igier.uni-bocconi.it

The Scope of Cooperation: values and incentives*

Guido Tabellini

IGIER, Bocconi University; CEPR; CES-Ifo; CIFAR

First version: January 2007; this version: October 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 values to transmit to their offspring, and this choice is influenced by the quality of external enforcement and the pattern of likely future transactions. The equilibrium displays strategic complementarities between values and current behavior, which reinforce the effects of changes in the external environment. Values evolve gradually over time, and if the quality of external enforcement is chosen under majority rule, there is histeresis: adverse initial conditions may lead to a unique equilibrium path where external enforcement remains weak and individual values discourage cooperation.

*I thank Alberto Bisin, Richard Holden and seminar participants at IGIER, the University of Naples, the NBER Summer Institute, the IMF research department for their comments, and Carlo Prato for excellent research assistance. I am grateful to CIFAR and Bocconi University for financial support.

1 Introduction

What determines the range of situations in which individuals cooperate? 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 values persist in some environments and not in others? How do values 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 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

[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] A large literature in the natural sciences and evolutionary psychology discusses the role of emotions in regulating and motivating human behavior, suggesting an evolutionary explanation of our moral capacities. See for instance Barkow, Cosmides and Tooby(1992), Pinker (1997), Massey (2002) and other references quoted in Kaplow and Shavell (2007). See also the evidence in Fehr, Fischbacher and Gachter (2002).

social norms has generally escaped the discipline of methodological individualism, the paradigm of economics. This is unfortunate because, as stressed by Kaplow and Shavell (2007), values and moral rules are malleable and entail an element of rational choice. The principles that children learn in the family or at school, or the codes of conduct that regulate adult individual behavior, are the outcome of purposeful choices and rational deliberations. As such, they lend themselves to be studied with the traditional methods of economics.[3]

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 values 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 is 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

[3] Besides the pathbreaking work of Gary Becker (see Becker 1993, 1996), recent contributions by economists have started undermining this division of labor. Guiso, Sapienza and Zingales (2006) and Fernadez (2007a,c) discuss much of this recent (mainly empirical) literature. See also Algan and Cahuc (2006), (2007), Barro and McCelary (2006) and Giuliano (2007). Important theoretical contributions include Bisin and Verdier (2001), Akerlof and Kranton (2000), (2006), Benabou and Tirole (2006a). 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 (2006c), 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.

(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, and entail respect for abstract individuals and their rights. 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, I build on the work of Bisin and Verdier (2001), Bisin, Topa and Verdier (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 backward looking, because the parents' values influence their educational choices. Thus, values evolve gradually over time and before reaching the steady state they reflect historical features of the external environment. But the equilibrium is also forward looking, since parents adapt their educational choices to the future environment of their children. This creates a strategic complementarity between values 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, values and behavior mutually reinforce each other, and this strenghtens the effects of changes in the environment, such as the external enforcement of cooperation.

In equilibrium, the diffusion of norms of generalized morality also reflects the pattern of likely future transactions relative to the pattern of moral ties between individuals. Localization of economic activity within a small community hurts the diffusion of values that sustain generalized cooperation, as parents have weaker incentives to teach values that are unlikely to be relevant. But extreme globalization can also be detrimental to values, if it increases the likelihood of situations where the parents' codes of good conduct are not applicable or have weak implications.

The endogeneity of values has additional implications if, as in Bisin and Verdier (2000, 2004) or Benabou and Tirole (2006a), 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 can hurt those who cheat. If external enforcement is the outcome of policy chocies, the endogeneity of values gives rise to 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 values 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 values and lack of cooperation. In both cases the equilibrium is unique, although its features are determined by initial conditions. For intermediate initial conditions, the model has multiple equilibria and the economy might converge to one or the other steady state, depending on 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, 2007, Rajan and Zingales 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, they display values and beliefs that are consistent with norms of limited morality, and are less likely to punish corrupt politicians - see the evidence in Tabellini (2005, 2007). The idea that culture is the missing link between distant history and current institutional performance is also supported by data on the attitudes of 2nd generation US citizens. Generalized trust is higher if the ancestors came from countries that over a century ago had better political institutions - cf. Tabellini (2007).

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

[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.

conduct are jointly determined, and their evolution is dictated by initial and possibly random historical circumstances.

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 discusses some extensions.

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 uniformly distributed on the circumference of 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. The probability of a match with someone located at distance y is $g(y)$, and it only depends on distance, not on the specific location. No restriction is placed on how $g(.)$ depends on y, except that the probabilities of all matches between 0 and S sum to 1, $\int_0^S g(y) = 1$.

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 cheat 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 3.5, 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:

$$T(\pi(y)) = [(w - l)\,\pi(y) + l] > 0 \tag{1}$$

We can think of this expression as the temptation not to cooperate. The right hand side of (1) is strictly positive: it is always better not to cooperate. The function $T(\pi(y))$ is increasing or decreasing in $\pi(y)$, the probability that the opponent will play C, depending on whether $w \gtrless l$.

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:

$$T(\pi(\tilde{y})) = de^{-\theta \tilde{y}} \tag{2}$$

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

$$\tilde{y} = \{\ln[d] - \ln[(w - l)\,\pi(\tilde{y}) + l]\}\,/\theta \tag{3}$$

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 is below some thresholds, while it fails above those thresholds. Specifically define the distance thresholds Y' and Y:

$$Y' = [\ln d - \ln l]\,/\theta \tag{4}$$

$$Y = [\ln d - \ln w]\,/\theta \tag{5}$$

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 (5) 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 $1 > n > 0$.[5]

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

$$Y^0 = [\ln d - \ln w] / \theta^0 \tag{6}$$
$$Y^1 = [\ln d - \ln [(w - l) n + l]] / \theta^1 \tag{7}$$

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 players, 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 players cooperate over a strictly larger range of matches. Those with $\theta = \theta^0$ continue to behave as described above: they

[5] 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 $d^1 > d^0$.

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).[6] 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;
- if the fraction of good players (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

[6] 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.

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 values that sustain cooperation, as captured by the parameter θ^k. Our goal is to study how parents 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 $1 - \delta - f$. [7]

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

[7] 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 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.

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 (6) and (7), 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, and recalling that the probability of a match with someone located at distance z is denoted $g(z)$, we can write V_t^{pk} as:

$$V_t^{pk} = U_t^k - d \int_{Y_t^k}^{S} e^{-\theta^p z} g(z) dz \tag{8}$$

where U_t^k is the expected equilibrium material payoffs of a kid of type k, while the second term on the RHS of (8) 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 (8), θ^p, is that of the parent, while the relevant distance thresholds according to which the 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:

$$U^k = \left[\int_0^{Y_t^k} g(z)[c\pi_t(z) - l(1 - \pi_t(z))]dz + \int_{Y_t^k}^{S} g(z)(c + w)\pi_t(z)dz \right] \tag{9}$$

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). 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, 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:

$$n_t = n_{t-1}(\delta + f_t) + (1 - n_{t-1})\delta = \delta + n_{t-1} f_t \tag{10}$$

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 (9) and (8). 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 (9) and (8) it can be written as:

$$f_t/\varphi = (U_t^1 - U_t^0) + d\int_{Y^0}^{Y_t^1} e^{-\theta^1 z} g(z)dz \tag{11}$$

Consider the right hand side of (11), 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 (30) 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 (4)). 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 (30) in subsection 3 of the Appendix, we can rewrite the parents' optimality conditions,

(11) as:

$$\begin{aligned} f_t &= \varphi d \left[-e^{-\theta^1 Y_t^1} \int_{Y^0}^{Y_t^1} g(z)dz + \int_{Y^0}^{Y_t^1} e^{-\theta^1 z} g(z)dz \right] \\ &= \varphi d \left\{ -e^{-\theta^1 Y_t^1} + E[e^{-\theta^1 y} \mid Y_t^1 \geq y \geq Y^0] \right\} \Pr(Y_t^1 \geq y \geq Y^0) \end{aligned} \quad (12)$$

Note that f_t denotes a probability and that by (12) $f_t > 0$. Thus, implicit in (12) is a restriction on parameter values (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 (12) 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 (12) 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 players 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 players 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 (10) and simplifying, the equilibrium is thus given by the vector (Y_t^{1*}, n_t^*) that solves the following two equations:

$$Y_t^1 = [\ln d - \ln [(w-l) n_t + l]] / \theta^1 \equiv Y(n_t) \quad (13)$$

$$n_t = \delta + n_{t-1} F(Y_t^1) \equiv N(Y_t^1, n_{t-1}) \quad (14)$$

The first equation defines the maximum distance Y_t^1 that sustains cooperation by the good players, as a function of the proportion of other good players 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 players 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 players, 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 (13) and (14) implicitly define the equilibrium vector (Y_t^{1*}, n_t^*) as a function of n_{t-1} :

$$Y_t^{1*} = G^Y(n_{t-1}) \tag{15}$$

$$n_t^* = G^n(n_{t-1}) \tag{16}$$

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

$$Y_s^{1*} = Y(n_s^*) \tag{17}$$

$$n_s^* = \frac{\delta}{1 - f_s} \tag{18}$$

where $f_s = F(Y_s^{1*})$ 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 values 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 5 of the appendix proves that a sufficient condition for this to happen is:

$$\frac{1}{\varphi} > l - w \tag{A1}$$

which says that the marginal cost of effort for the parents, $1/\varphi$, must be higher than the strategic complementarity in the prisoner's dilemma game, captured by $(l - w)$. In the remainder of the paper we assume that (A1)holds, so that the equilibrium (Y_t^{1*}, n_t^*) is unique.[8]

Subsection 6 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 $1/\varphi > l - w$, 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 (17)-(18). If $1/\varphi > l - w$, 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 the 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.

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, $(n_s^*,\ Y_s^{1*})$.

3.4.1 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,

[8] Note that, even if condition (A1) fails, the game remains supermodular. Hence, although there might be multiple equilibria, all comparative statics results apply to the extremal equilibria - (Amir 2003). Of course, the matching game described in section 2 and played in each period by the kids has multiple equilibria even if (A1) holds. But here we are restricting attention to the Pareto superior equilibrium of the matching game.

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.[9] As l is reduced, the curve $Y_t^1 = Y(n_t)$ in Figure 2 shifts to the right - cf (13). Intuitively, for a given n_t, the good players 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. (12). Thus, the scope of cooperation immediately 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 players (n_0 rises), which further increases the scope cooperation 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, with both a larger fraction of good players and where 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.

3.4.2 Economic geography

Next, consider a change in the matching technology, as captured by the probability of a match with someone located at distance y, $g(y)$. With reference to Figure 2, this immediately shifts the curve $N(Y_t^1, n_{t-1})$, through the parents' incentives to educate their kid, while the curve $Y(n_t)$ remains unaffected.

[9] 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.

Specifically, consider a uniform increase in $\Pr(Y_s^1 \geq y \geq Y^0)$, which occurs without changing $E[e^{-\theta^1 y} \mid Y_s^1 \geq y \geq Y^0]$ and for given values of Y_s^1 and Y^0. In words, there is an increase in the probability of matches in the interval $[Y^0, Y_s^1]$, where the two types of players behave differently. Suppose that this occurs at the beginning of period 0 and before parents choose their effort. By (12), equilibrium effort f_0 in period 0 increases. Intuitively, the interval $[Y^0, Y_s^1]$ is where the difference between the two types is relevant, and hence where effort pays off from the parent's perspective. Hence, any increase in the probability of matches in this region induces more effort. This in turn increases the fraction of good players, n_0, which also brings about an immediate expansion of the upper threshold of cooperation by the good players, Y_0^1. From here onwards, a dynamic process of adjustment to the new steady state takes place, similar to the one described above, which eventually leads to a higher fraction of good players, n_s, and to a larger upper threshold of cooperation, Y_s^1.[10]

A less localized economy An increase in the probability of matches in the interval $[Y^0, Y_s^1]$ can occur for different reasons, suggesting different interpretations. One possibility is that the probability of nearby matches, with someone at distance below Y^0, drops. In other words, the economy has become less localised. Thus, the model suggests that norms of generalized morality become more diffuse if transactions are less localized, or viceversa that a more localized economy breeds limited morality - recall that the distance threshold that sustains cooperation by all types, Y^0, can be interpreted as the scope of application of norms of limited morality.

Abandoning a literal interpretation, this result can contribute to explain a difference in traditional economic organizations between East and West often stressed by economic historians. Whereas in Western Europe impersonal exchange took place in anonymous markets supported by "public order" or "private order" institutions obeying formal procedures, in East Asia markets were organized through a web of kin-based social structures linked by personal relations (Grief 2005). These historical differences are bound to reflect a variety of economic and political forces, but culture is also likely to play a role. Generalized morality, and in particular generalized respect for the individual and his rights, probably facilitated the evolution of the political

[10] As can be seen by (12), a change in the matching technology $g(y)$ that increases $E[e^{-\theta^1 y} \mid Y_s^1 \geq y \geq Y^0]$ leaving $\Pr(Y_s^1 \geq y \geq Y^0)$ unaffected has similar effects.

and economic institutions supporting Western style markets (Grief 2005). In contrast, Eastern economic organizations were also supported by a culture of loyalty to the local community or to a network of relatives and friends, something not dissimilar to the concept of limited morality captured by the model (Kumar and Matsusaka 2006). But what is the source of these different cultural traits? It is tempting to answer with reference to the results mentioned above. In the XVIIth century, population density was much higher in China and India than in Europe. Moreover, long distance travel was easier within Europe than within the far East, because of both geography and relative availability of means of transportation (Kumar and Matsusaka 2006). As a result, trade was more frequently local in Asia than in Europe. As suggested by the model, the greater localization of economic activity might have discouraged the diffusion of abstract norms of generalized morality in the East compared to the West.

The adverse effects of globalization Alternatively, the probability of matches in the region $[Y_s^1, Y^0]$ can go up because very distant matches (above Y_s^1) have become less likely. This too is beneficial to the diffusion of good values, because the more frequent interactions inside the community of reference for the good players strengthens their incentive to transmit these values to their offspring.

Taken literally, this says that globalization (the equivalent of more frequent very distant matches) might reduce the scope of cooperation, because it destroys the values that induce individuals to cooperate. More generally, this result can be interpreted as saying that the diffusion of good values is hurt by economic forces that induce individuals to move outside of the community with which older generations identify. This is consistent with recent evidence by Miguel et al. (2002) on Indonesia: social capital and community values were hurt in districts left behind in the process of industrialization and with severe out-migration. It is also consistent with observations by economic historians like Polany (1957), that the industrial revolution destroyed moral values in the UK.

Summarizing, the general insight of the model is that the evolution of values depends on the patterns of economic interactions relative to the pattern of moral ties between individuals. 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 diffusion of trustworthiness within the community. 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, and this dampens the incentives to spread morality.

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:

$$Y_t^1 = [\ln d + \ln n_t - \ln [(w - l) n_t + l]] / \theta^1 \equiv Y(n_t) \qquad (19)$$

Comparing (19) with the previous expression in (13), we have added the term $\ln n_t$ that was missing in (13). Intuitively, if more good players 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 players (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).[11]

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 (11) and (12). 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

[11] 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.

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 players, n_t). For this reason, a norm of reciprocity also reinforces the effects of changes in the external environment on equilibrium outcomes.

4 Endogenous Government Policies

If the payoffs of the matching game result from policy choices, different player types prefer different policies. Good players generally prefer better enforcement of cooperation, compared to the bad players. When public policies are chosen under majority rule, this creates an additional strategic complementarity. If the good players 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 players 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. Thus, politics and culture interact with feedback effects going in both directions.

A similar point is illustrated with respect to welfare state policies in models by Bisin and Verdier (2000, 2004), where individual tastes for private vs public consumption or leisure vs work are endogenous, and by Benabou and Tirole (2006), 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. Since we have an explicitly dynamic economy, not only there can be multiple equilibria, but there is also histeresis: initial conditions matter, because they 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 The enforcement regime in political equilibrium

To simplify the algebra, throughout this section we make two assumptions. First, the probability of being matched with someone at distance y is the same for any distance: $g(y) = \eta$. Second, the material payoffs of the matching game are such that $l = w = w_t$, where w_t is a policy variable chosen by the government, as described below. With this formulation, 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 players, and is given by:

$$Y_t^k = [\ln d - \ln w_t] / \theta^k, \qquad k = 0, 1 \tag{20}$$

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 policy variable w_t can be interpreted as external enforcement by the government. A higher value of w_t corresponds to a larger temptation to cheat and a smaller loss from cooperating against a cheating opponent, and hence worse external enforcement. To simplify the algebra, we assume that government policy only matters for matches outside of the safe range $y < Y^0 \equiv [\ln d - \ln w] / \theta^0$. Thus, for any match inside the safe range $y \leq Y^0$, external enforcement always satisfies $w_t = w > 0$ irrespective of government policy. For more distant matches such that $y > Y^0$, the government is free to choose any w_t inside the interval $[w, W]$, where $W > w$ are given parameters.[12]

Government policy is set under majority rule in each period. The timing of events is as follows. First, parents choose their educational effort. Then,

[12] Thus, matches beyond Y^0 can have worse external enforcement compared to matches inside Y^0, but they cannot have strictly better enforcement.

the kids' type becomes known and the kids vote over the enforcement regime (parents don't participate in the vote). 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 players is already determined. Thus, current enforcement only affects current expected payoffs in the matching game.

What is the policy preferred by the two types? It is easy to verify that the good players always prefer the strongest possible external enforcement, corresponding to $w_t = w$, since this reduces their loss from being cheated (see subsection 7 of the Appendix). We call this policy outcome the *strong enforcement regime*.

The bad players, instead, face a tradeoff: on the one hand, worse external enforcement (a higher w_t) increases the benefit of cheating; on the other hand, it makes the good players more cautious, and this in turn shrinks the range of matches over which the bad players can take advantage of a cooperating opponent. Subsection 7 of the Appendix proves that the first effect dominates at the lower bound w, if the psychological cost d is sufficiently large relative to the material payoffs of the prisoner's dilemma game (see condition A2 in the appendix). Hence, under this condition, the bad players always prefer a weaker external enforcement than technologically feasible, $w_t > w$. The appendix also shows that the optimal policy from the point of view of the bad players is time invariant, since it does not depend on n_t. We call this policy outcome the *weak enforcement regime*, and we denote it as $w_t = \bar{w}$, where $\bar{w} > w$.

Given these results, the political equilibrium in any period t is straightforward and it is summarized in the following:

Lemma 7 *Suppose that condition (A2) in the appendix holds. If $n_t > 1/2$ then the strong enforcement regime prevails in period t: $w_t = w$. If $n_t < 1/2$ then the weak enforcement regime prevails in period t: $w_t = \bar{w}$, with $\bar{w} > w$. If $n_t = 1/2$ then either regime can prevail.*

4.2 Equilibrium dynamics

The two enforcement regimes in Lemma 7 entail different incentives to inculcate trustworthiness. Since the good players bear a larger loss from being cheated if external enforcement is weak, parents exert more effort to educate their kid in the strong than in the weak enforcement regime.

Specifically, let f and $\bar{f}$ denote educational effort under strong and weak enforcement respectively. Since under strong enforcement $w_t = w$, effort in the strong enforcement regime is $f = F(Y^1)$, where $F(.)$ is still given by (12) above with $w = l$ in it. Subsection 8 of the appendix proves that, under condition (A2) in the appendix, effort under weak enforcement is:[13]

Lemma 8 $\bar{f} = f - \Delta > 0$ *with* $\Delta > 0$

This set up induces a strategic complementarity in the education decision of the parents. If parents expect $n_t > 1/2$, then they anticipate better enforcement and, by Lemma 8, they exert more effort to inculcate trustworthiness in their kid. This in turn increases the fraction of good players, and might bring about a political equilibrium where they are a majority. Viceversa, if parents expect $n_t < 1/2$, they reduce effort, which might shift future political majorities. For some parameter values, this can give rise to multiple steady states.

Specifically, suppose that parents expect strong enforcement. Then the steady state fraction of good players is given by (18) in the previous section, reproduced here for convenience (with f_s replaced by f):

$$n_s^* = \frac{\delta}{1-f} \tag{21}$$

If $n_s^* > 1/2$, this steady state reproduces itself in a political equilibrium. Suppose instead that parents expect weak enforcement. Then, by Lemma 8, the steady state fraction of good players is:

$$\bar{n}_s^* = \frac{\delta}{1-f+\Delta} \tag{22}$$

If $\bar{n}_s^* < 1/2$, this steady state too reproduces itself in a political equilibrium. Thus, both steady states are possible in equilibrium if $n_s^* > 1/2 > \bar{n}_s^*$, or, by (22) and (21), if:

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

Note that, since $f > \Delta$, the left hand inequality 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 $\bar{Y}^1$, do not depend on n_t). Thus,

[13] Note that parental effort no longer depends on time, in either regime, since with $w = l$ effort no longer depends on n_t.

under the same conditions discussed in the previous sections, both steady states are dynamically stable and the adjustment is monotonic.[14]. Which steady state is reached in equilibrium depends on the initial conditions and on parents' expectations, as we now discuss.

Strong enforcement is a political equilibrium in period t if, given that it is expected, we have $n_t > 1/2$. By (10), this condition can be stated as:

$$n_t = \delta + n_{t-1} \; f > 1/2 \tag{23}$$

Similarly, weak enforcement is a political equilibrium in period t if, given that it is expected, $n_t < 1/2$, namely if :

$$n_t = \delta + n_{t-1}(f - \Delta) < 1/2 \tag{24}$$

Combining (23) and (24), we obtain two thresholds, that define which equilibria exist in period t, depending on the fraction of good players in period $t-1$. Specifically, let:

$$\hat{n} = \frac{1-2\delta}{2f} \tag{25}$$

$$\hat{N} = \frac{1-2\delta}{2(f-\Delta)} \tag{26}$$

with $\hat{N} > \hat{n}$. Then we have:

Lemma 9 *If $n_{t-1} < \hat{n}$, then in period t the unique equilibrium has weak enforcement. If $n_{t-1} > \hat{N}$ then in period t the unique equilibrium has strong enforcement. If $\hat{N} \geq n_{t-1} \geq \hat{n}$, then both the weak and the strong 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 $\hat{n}$, then even if parents expect strong enforcement, we have $n_t < 1/2$. Hence strong enforcement cannot be a political equilibrium. Conversely, if n_{t-1} is so high that it exceeds the threshold $\hat{N}$, then even if parents expect weak enforcement we have $n_t > 1/2$, which rules out weak enforcement as an

[14]Stability of both steady state requires $f < 1$, $f - \Delta < 1$. A monotonic adjustment path requires $f > 0$ and $f - \Delta > 0$, which follows from the optimality condition of parental effort.

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 $[\hat{n}, \hat{N}]$. Manipulating (25)-(26) and (21)-(22), a sufficient condition for this to happend is:

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

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 $\hat{n}$ and $\hat{N}$ define three regions with different dynamics. If the economy starts from an initial condition $n_0 < \hat{n}$, then the equilibrium is unique. The economy remains for ever in the weak enforcement equilibrium and it converges to the weak enforcement steady state. Conversely, if the economy starts from an initial condition $n_0 > \hat{N}$, then the equilibrium is again unique. The economy remains for ever in the strong enforcement equilibrium and it converges to the strong enforcement steady state. If the initial condition is in between these two tresholds, $n_0 \in [\hat{n}, \hat{N}]$, 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 10 *If condition (A3) holds, then the economy has two steady states, one with strong external enforcement and where the good players are a majority; and one with weak 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* $[\hat{n},\hat{N}]$*, then the equilibrium is unique. For* $n_0 < \hat{n}$ *(for* $n_0 > \hat{N}$*), the economy remains always under the weak (strong) enforcement regime and eventually reaches the weak (strong) 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

Proposition 10 highlights the importance of mutually reinforcing effects between culture and politics when both are endogenous. On the one hand, the effectiveness of law enforcement shapes the incentives to transmit moral values that support honest behavior. On the other hand, the quality of law enforcement is also endogenous, and reflects deliberate policy choices. A society with weak moral values, or where respect for the law and for others is lacking, is also more tolerant of lax law enforcement. As a result, otherwise identical societies may end up along very different paths if they start from different initial conditions.

Thus, 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 make citizens more tolerant of ineffective government. Better institutions are available, and nothing prevents the country from adopting them, but this does not happend in a political equilibrium. Whether lax law enforcement refers to tax evasion, or to free riding on public transportations, or to cheating on the public welfare system, weak law enforcement is tolerated and perhaps even preferred by a majority of citizens. 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 law 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.

Similar arguments would also apply if policy outcomes refer to an agency conflict between the government and citizens at large, rather than the enforcement of cooperation between citizens. Ousting a corrupt politician requires a minimal amount of cooperation by citizens, who need to be informed, to bear the cost of voting, and perhaps to vote according to general social welfare rather than their own particularistic benefit (cf. Ferejohn 1986). These forms of political cooperation may not be sustainable where limited morality prevails. If government abuse and nepotism in turn induce the diffusion of adverse cultural traits, then we have yet another loop where political and cultural outcomes have mutually reinforcing effects. Preliminary evidence in Tabellini (2007) supports this idea. Italian voters in regions where generalized morality is more diffused, and that were ruled by better political institutions over two centuries ago, are more willing to punish incumbent politicians under criminal investigation.

The interaction between culture and politics has relevant implications also for groups formation and redistributive policies. It is well known, for instance, that in Africa public policies often provide targeted benefits to ethnic groups. Opportunistic politicians have an incentive do so if ethnic ties are strong and if individuals identify with ethnic groups, rather than with groups formed along other economic or social dimensions. But group identity is not exogenous, on the contrary, it is likely to be strengthened by any policy that targets the group. Indeed, Miguel and Posner (2006) found that in Africa ethnic identity is stronger amongst the individuals who are more likely to be exposed to public policies. The approach of this paper could be extended to examine the historical reasons that make some groups influential, and to study the joint evolution of redistributive politics and group identity.

Finally, in the model individuals were assumed to vote according to their self interest. Hence cultural traits influence political preferences only through economic behavior, because this determines how individuals are affected by external enforcement. If instead individual values also have a direct impact on political ideologies, as seems plausible, then there is an additional channel through which the external environment interacts with individual values, which could reinforce the results presented above.[15]

[15] 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.

5 Concluding remarks

I conclude by discussing several possible extensions of this basic framework, and other recent related work.

The model literally assumes that values are transmitted within the family, and that only parents make purposeful educational choices. In practice, other channels of cultural transmission, from peers, own experience, educational institutions or the media, are also likely to be important. This opens the door to other relevant choices, such as whom to select as your friend, or how intensely to experiment. It also gives a role to other motivated actors who might have economic or political reasons to influence cultural traits. Recent work by Benabou (2007), Fernandez (2007b) and Guiso, Sapienza and Zingales (2007) has started investigating the role of belief formation and manipulation in a variety of related settings.

I have also neglected reputational forces, not because they are unimportant, but to focus on values alone. As stressed for instance by Kaplow and Shavell (2007), values also interact with reputational incentives. Sustained punishment of deviant behavior is more likely to be incentive compatible, or to be a focal point of coordinated action, it the deviation is morally tainted. Thus, reputational mechanisms support law enforcement when the law is considered fair, or if the state enjoys the confidence of citiziens, but not if individual values clash with the state. Incorporating these channels in this model, perhaps exploiting the work of Dixit (2004), is feasible and might lead to new insights.

Reputation can also operate through signalling. If values were observable, then players with good values might have an advantage in this model, because they could induce their partner to cooperate over a larger range of matches. This would change the incentives of parents, and even the bad players in the model might want to transmit better values to their children, depending on the strength of external incentives. True values may be unobservable, but individuals may find ways of signalling them (other than through repeated cooperation). Levy and Razin (2006), for instance, formulate a theory of religion based on the assumption that religious rituals are observable (maybe only within a subset of the population). This creates a strategic incentive to join a religious organization, to signal one's type. 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). Combining their insights with the model of this paper seems doable and promising.

In summary, 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. These issues can be fruitfully studied with the standard tools of economic analysis, and can yield important new insights on why cooperation is easier to sustain in some social environments than in others.

6 Appendix

6.1 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 (3). Replacing $\pi(y)$ with 0 and 1 respectively, equation (3) yields $\tilde{y} = Y'$ and $\tilde{y} = Y$, as defined in (4) and (5). 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

6.2 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 $\theta = \theta^0$, (9) simplifies to:

$$\begin{aligned} U_t^0 &= c\int_0^{Y^0} g(z)dz + (c+w)n_t \int_{Y^0}^{Y_t^1} g(z)dz \\ &= c\ \Pr(z \leq Y^0) + (c+w)n_t\ \Pr(Y_t^1 \geq z \geq Y^0) \end{aligned} \tag{27}$$

The first term on the right hand side of (27) 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 players play non-cooperatively. Repeating the same steps, the expected utility of those with $\theta = \theta^1$ instead is:

$$U_t^1 = c\Pr(z \leq Y^0) + [cn_t - l(1-n_t)]\Pr(Y_t^1 \geq z \geq Y^0) \tag{28}$$

where the first term on the right hand side of (28) continues to have the same interpretation, while the second term is the expected outcome, given that only the good players cooperate in the intermediate distance range.

Note that, (7) implies:

$$de^{-\theta^1 Y_t^1} = [l + (w-l)n_t)] \tag{29}$$

Hence, (27) and (28) imply:

$$\begin{aligned} U_t^1 - U_t^0 &= -[l + (w-l)n_t)]\Pr(Y_t^1 \geq z \geq Y^0) = \\ &= -de^{-\theta^1 Y_t^1}\Pr(Y_t^1 \geq z \geq Y^0) < 0 \end{aligned} \tag{30}$$

where the last equality follows from (29).

6.3 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 (8), 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 (9) and (8), differentiating V^{pk} with respect to θ^k and rearranging, we have:

$$\frac{\partial V^{pk}}{\partial \theta^k} = \frac{\alpha e^{-\alpha Y^k}}{2\left[1-e^{-\alpha S}\right]} \frac{\partial Y^k}{\partial \theta^k} \left\{ d e^{-\theta^p Y^k} - \left[(w-l)\pi(Y^k)+l\right] \right\} \tag{31}$$

By (6) and (7), $\frac{\partial Y^k}{\partial \theta^k} < 0$. Hence, the optimal value of d^k is such that Y^k solves the expression

$$d e^{-\theta^p Y^k} = \left[(w-l)\pi(Y^k)+l\right] \tag{32}$$

for $\pi(Y^k)$ corresponding to the equilibrium probability of cooperation by a partner located at distance Y^k. But by (3), this implies $\theta^k = \theta^p$. Hence the parent strictly prefers to have a kid with his own value parameter. QED

6.4 Proof of Lemma 5

Differentiating the RHS of (12) with respect to Y_t^1 and simplifying, we have:

$$F_{Y_t^1} = \varphi d \theta^1 e^{-\theta^1 Y_t^1} \int_{Y^0}^{Y_t^1} g(z)dz > 0 \tag{33}$$

QED

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

Equation (14) implies that $N_{Y_t^1} = n_{t-1} F_{Y_t^1} > 0$

Differentiating the RHS of (13) with respect to n_t, we also have:

$$Y_{n_t} = \frac{1}{\theta^1} \frac{l-w}{x_t} > 0 \tag{34}$$

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 (32), (33) and (34), if:

$$\frac{1}{l-w} > \varphi n_{t-1} \int_{Y^0}^{Y_t^1} g(z)dz \tag{35}$$

Since $n_{t-1} \int_{Y^0}^{Y_t^1} g(z)dz < 1$, a sufficient for (35) to hold is:

$$\frac{1}{\varphi} > l - w \tag{A1}$$

6.6 Dynamic stability of the steady state

Applying the implicit function theorem to (15) and (16), 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

$$f_t + Y_{n_t}.N_{Y_t^1} < 1 \tag{36}$$

By (33) and (34), the term $Y_{n_t}.N_{Y_t^1}$ is proportional to φ. By (12), 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. (34), (33) and (12)). 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 $0 < \varphi < \bar{\varphi}$, (36) also holds. QED

6.7 Proof of Lemma 7

Here we show that the good players always prefer the strong enforcement regime, $w_t = w$, and we provide a sufficient condition that guarantees that the bad players prefer the weak enforcement regime, $w_t = \bar{w} > w$.

Let U_t^k denote the expected utility of players of type k as a function of w_t Exploiting the envelope theorem and adapting (28) and (8) to the new notation, we have:

$$\frac{\partial U_t^1}{\partial w_t} = -(1 - n_t) \int_{Y^0}^{Y_t^1} g(z)dz < 0$$

where Y_t^1 is the maximum threshold for cooperation of the good players under w_t. Since this expression holds for any w_t, the good players are always in favor of the lowest possible value of w_t, namely they prefer the weak enforcement regime $w_t = w$.

Next, consider the bad players. Since the threshold Y^0 is not affected by the regime, the derivative of their expected utility with respect to w_t is:

$$\frac{\partial U_t^0}{\partial w_t} = n_t \int_{Y^0}^{Y_t^1} g(z)dz + (c + w_t)n_t \frac{\partial Y_t^1}{\partial w_t} g(Y_t^1) \tag{37}$$

We assumed $g(z) = \eta$. Moreover, by (20), $\frac{\partial Y_t^1}{\partial w_t} = -1/\theta^1 w_t$. Hence, (37) simplifies to:

$$\frac{\partial U_t^0}{\partial w_t} = n_t \eta (Y_t^1 - Y^0) - n_t \eta \frac{(c + w_t)}{\theta^1 w_t} \tag{38}$$

Evaluating the right hand side of (38) at the point $w_t = w$, and simplifying, we can show that it is strictly positive if :

$$\ln d > \ln w + \frac{(c + w)\theta^0}{w(\theta^0 - \theta^1)} \tag{A2}$$

By (20), Y_t^1 only depends on time through w_t. Hence, by (38), the optimal value of w_t from the point of view of the bad players is constant. Denoting such optimal value by $\bar{w}$, under (A2) we have $\bar{w} > w$.

6.8 Proof of Lemma 8

Let an upper bar over a variable denote the corresponding variable in the weak enforcement regime. Repeating the previous analysis, parental effort in the weak enforcement regime is given by the following first order condition, adapted from (11):

$$\bar{f}\ /\varphi = (\bar{U}^1 - \bar{U}^0) + d\int_{Y^0}^{\bar{Y}^1} e^{-\theta^1 z} g(z)dz \tag{39}$$

where time subscripts have been dropped because under the simplifying assumptions of this section n_t no longer enters the right hand side of (39). We know from previous results that $\bar{f} > 0$. Moreover, in the weak enforcement 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):

$$\begin{aligned} \bar{U}^1 - \bar{U}^0 &= -\bar{w}\int_{Y^0}^{\bar{Y}^1} g(z)dz = \\ &= -w\int_{Y^0}^{Y^1} g(z)dz + (w-\bar{w})\int_{Y^0}^{Y^1} g(z)dz + \bar{w}\int_{\bar{Y}^1}^{Y^1} g(z)dz \end{aligned} \tag{40}$$

Combining (40) and (39), and exploiting (12) for $w = l$, we have:

$$\bar{f} = f - \Delta$$

where, under the simplifying assumption that $g(z) = \eta$:

$$\Delta = \varphi\eta \left\{ (\bar{w} - w)(Y^1 - Y^0) - \bar{w}(Y^1 - \bar{Y}^1) + d\int_{\bar{Y}^1}^{Y^1} e^{-\theta^1 z} dz \right\} \tag{41}$$

We also have:

$$\begin{aligned} \frac{\partial \Delta}{\partial \bar{w}} &= \varphi\eta \left\{ (\bar{Y}^1 - Y^0) + \bar{w}\frac{\partial \bar{Y}^1}{\partial \bar{w}} - de^{-\theta^1 \bar{Y}^1}\frac{\partial \bar{Y}^1}{\partial \bar{w}} \right\} = \\ &= \varphi\eta(\bar{Y}^1 - Y^0) > 0 \end{aligned}$$

where the second equality follows from $d = \bar{w}\ e^{\theta^1 \bar{Y}^1}$, and where $(\bar{Y}^1 - Y^0) > 0$ follows from the optimality condition (38) and the definition of $\bar{w}$ as the optimal value of w_t for the bad players. Note that when $\bar{w} = w$, we have $\bar{Y}^1 = Y^1$ so that $\Delta = 0$, by (41). Thus, $\bar{w} > w$ implies $\Delta > 0$. QED

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] Barkow, Jerome H., Leda Cosmides, and John Tooby (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Oxford U P

[12] Barro, R. and McCleary, R.M. (2006). "Religion and Economy," *Journal of Economic Perspectives*, Spring

[13] 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.

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

[15] Benabou, R. (2007), "Groupthink and Ideology", Schumpeter Lecture at the meetings of the European Economic Association, *Journal of the European Economic Association*, forthcoming

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

[17] Benabou, R. and Tirole, J. (2006b), "Identity, Dignity and Taboos: Belief as assets", mimeo, Princeton University

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

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

[20] Bisin, A. and Verdier, T. (2000), "A model of Cultural Transmission, Voting and Political Ideology", *European Journal of Political Economy*, 16, 5-29

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

[22] Bisin, A. and Verdier, T. (2004), "Work Ethic and redistribution: A Cultural Transmission Model of the Welfare State", mimeo, New York University

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

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

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

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

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

[28] Ferejohn, J. (1986), "Incumbent Perfromance and Electoral Control", *Public Choice,* Fall , vol 30

[29] Fernandez, R. (2007a), "Culture, Women, and Work", *Journal of the European Economic Association*

[30] Fernandez, R. (2007b), "Culture as Learning: the Evolution of Female Labor Force over a Century", mimeo, New York University

[31] Fernandez, R. (2007c), "Culture and Economics", *New Palgrave Dictionary of Economics*, 2nd edition, forthcoming

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

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

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

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

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

[37] Greif, A. (2005), "Commitment, Coercion, and Markets: The Nature and Dynamics of Institutions Supporting Exchange", C. Ménard and M. Shirley (eds.), *Handbook of New Institutional Economics,* 727–786, Springer.

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

[39] Guiso L., Sapienza P. and Zingales L. (2007), "Social Capital as Good Culture," Marshall Lecture at the meetings of the European Economic Association, *Journal of the European Economic Association*, forthcoming

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

[41] Kaplow, L. and S. Shavell (2007), "Moral Rules, the Moral Sentiments, and Behavior: Toward a Theory of an Optimal Moral System", *Journal of Political Economy*, 115 (3)

[42] Kumar, K. and J. G. Matsusaka (2006), "Village vs Market Social Capital", mimeo, University of Southern California

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

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

[45] Massey, D. S. (2002), "A Brief History of Human Society: The Origin and Role of Emotion in Social Life", *American Sociological Review*, 67 (1), 1-29.

[46] Miguel, E., P. Gertler and D. I. Levine (2002), "Did INdustrialization Destroy Social Capital in Indonesia ?", NBER working paper

[47] Micguel, E. and D.N. Posner (2006), "Sources of Ethnic Identification in Africa", mimeo, U.C. Berkeley

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

[49] Pinker, S. (1997), *How the Mind Works*, Norton.

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

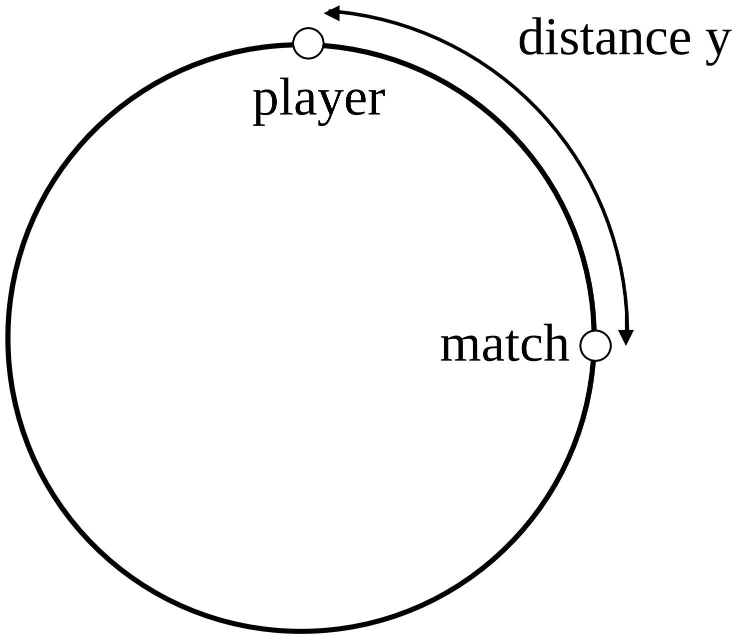
[51] Polany, K (1957), *The Great Transformation: the Political and Economic Origins of Our Time,* Beacon Press, Boston

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

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

[54] Tabellini, G. (2007), "Institutions and Culture", Presidential lecture at the European Economic Association, *Journal of the European Economic Association*, forthcoming

Figure 1



prob. of match: $g(y)$

Psychological cost
of NC: $d\ e^{-\theta y}$

Figure 2