European Journal of Political Economy 70 (2021) 102032

Contents lists available at ScienceDirect

European Journal of Political Economy

journal homepage: www.elsevier.com/locate/ejpe

Trust and punishment☆

Gonzalo Olcina [a,b], Vicente Calabuig [a,b,*]

[a] University of Valencia, ERI-CES, Spain
[b] Faculty of Economics, Avda. dels Tarongers, s/n, 46022, Valencia, Spain

ARTICLE INFO

JEL classification:
D01
D02
Z13

Keywords:
Cultural transmission
Trust and investment
Punishment institutions
Reciprocity preferences

ABSTRACT

This paper explores the impact of institutions on the evolution of preferences (culture) and on economic outcomes. Punishment institutions determine the capacity and the individual cost of punishing opportunistic behavior, while preferences are endogenous and can be influenced by a cultural transmission process that is conditioned by the existing punishment institutions. We investigate the interaction and evolution between the preferences for reciprocity or rewarding of the allocator and the preferences to punish hostile behaviour by the investor in a trust game with a costly punishment phase. Our main result provides a rationale for the existence of a strong positive relationship between the quality of society's punishment institutions and the levels of trust achieved. Trust only flourishes in the shadow of efficient sanctioning institutions. The driving force behind this result is that strong institutions are necessary to make the socialization of children towards punishment preferences worthwhile. However, there is no relationship between the level of trust and the distribution of social preferences. Even when there is a high proportion of punishers in the population, a high level of trust will not be obtained if there is not a well-functioning punishment institution.

1. Introduction

This paper explores the influence of punishment institutions on the evolution of preferences (culture) and on the level of trust and economic outcomes in a society.

A large body of literature[1] suggests that the level of trust in a society strongly influences its economic success. Trusting behaviour is the intention to accept vulnerability: to voluntarily place resources at the disposal of another party without any legal commitment from the latter. Hence, according to this widely accepted definition (Coleman, 1990), trust results in a behaviour that can be directly measured. Trust favours cooperative behaviour and facilitates mutually advantageous exchanges in the presence of incomplete contracts. But trust, its level and distribution across societies, as an explanation for cooperation needs further clarification. Which, then, are the determinants of trust? Why is the level of trust so different across countries and societies?

One possible explanation could be the different preference or values distributions across societies. In fact, variation in economic preferences is systematically related to both individual and aggregate economic outcomes (see, for instance, Falk et al., 2018). Another explanation puts the emphasis on well-functioning institutions. Again, variation in the quality of institutions that govern

☆ We would like to thank the editor and the two anonymous reviewers for their constructive comments which helped to improve the manuscript. The authors also wish to acknowledge the financial support received from the Spanish Ministry of Economy and Competitiveness projects ECO2014-58297-R and ECO2017-87245-R.

* Corresponding author. Faculty of Economics, Avda. dels Tarongers, s/n, 46022, Valencia, Spain.
E-mail address: Vicente.Calabuig@uv.es (V. Calabuig).

[1] See for instance, Arrow (1972), Putnam (1993), Knack and Keefer (1997), Tabellini (2010), Algan and Cahuc (2009), among others).

https://doi.org/10.1016/j.ejpoleco.2021.102032
Received 25 October 2019; Received in revised form 9 March 2021; Accepted 21 March 2021
Available online 3 April 2021
0176-2680/© 2021 Elsevier B.V. All rights reserved.

the functioning of markets and societies is also related to aggregate economic outcomes (see, for instance, Algan and Cahuc, 2014; Tabellini, 2008b, Acemoglu et al., 2006). But, what really makes a difference? Is it a good culture - preferences - or is it the presence of good institutions that make for good economic outcomes?

This paper presents a theoretical analysis on when and how punishment institutions shape the distribution of preferences in a society, crucial for the proper functioning of these institutions, and how they shape the level of trust, measured by the level of aggregate investment achieved in the society. We assume that institutions are exogenously given and embody the capacity and the individual cost of punishing opportunistic behaviour, i.e., how much damage can the punisher impose and at what cost to the punisher. Preferences are endogenous and they can be influenced by the existing institutions. A potential channel for the influence of institutions on preferences is cultural transmission. If preferences are (at least partially) intergenerationally transmitted and parental transmission (and others, for example, school) is intentional, then institutions characterized by different features will affect the incentives for socialization and, in turn, will change the preference distribution in the long run.[2]

To analyse this problem in a theoretical framework, we use an extended trust game with a costly punishment phase. In this social dilemma known as the *trust game*, one player (the investor) has the option of investing or not investing in a project, which is administered by the other player (the allocator). To invest leads to a higher joint surplus, but the allocator controls the proceeds of investment. The trust game is a metaphor for many social bilateral relationships. A prominent example is seen in the labour market. In many employment relations, the labour contract is highly incomplete and assigns usually significant authority to the employer.

We work with an overlapping-generations model of cultural transmission of preferences where both the distribution of preferences in the population and behaviour in the long run are determined endogenously. We assume that there are two kinds of preferences in each population (allocators and investors): selfish and reciprocal individuals. The latter are the potential punishers. We investigate the interaction and evolution between the preferences for rewarding by the allocator and the preferences to punish hostile behaviour by the investor in a trust game with a costly punishment phase. The distribution of preferences in each population evolves according to a process of cultural transmission which combines direct transmission from parents with oblique transmission from the society (Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985; Bisin and Verdier (1998; 2001)). However, in our model, cultural transmission takes place within the specific class or group (investors and allocators). We denote this situation as a segmented society. This is a good representation of a society with very low social mobility.[3]

Our model yields different long run outcomes or cultures in a society, characterized by an aggregate amount of investment (a level of trust), the rewards set by allocators (level of trustworthiness) and a distribution of preferences, which are determined by the features that characterize the punishment institutions, the capacity and cost of punishing and, in some cases, by the initial conditions of the dynamics (the history of the society).

The first result we obtain is that if there is a high institutional capacity for punishment in the society, a Fully Cooperative Culture will eventually be reached from any initial condition of the dynamics. This society will exhibit the highest levels of trust (investment) and trustworthiness (rewards). This culture will be characterized by a high proportion of reciprocal individuals in the investors population who are willing to punish unfair rewards and, any distribution of preferences in the allocators population. No punishment would be observed, because of the existence of a credible threat of punishment of opportunistic behaviour.

Secondly, if there is a low institutional capacity of punishment in the society, an Inefficient Separating Culture will eventually be reached from any initial condition of the dynamics. This culture will be characterized by a high proportion of selfish allocators and also a very high proportion of reciprocal investors. But, given the existing low institutional capacity for punishment, reciprocal investors are not willing to punish unfair rewards and prefer not to invest. Consequently, this society displays low levels of aggregate investment (trust), since a small fraction of selfish investors exists that invest while the majority of investors do not invest. This economy shows the lowest levels of trust and investment, of trustworthiness and rewards and again, there is not observable punishment.

Finally, if there is an intermediate capacity for punishment and/or an intermediate or high cost of implementing it, the society could end up in a multiplicity of cultures: Fully Cooperative or Inefficient Separating Culture, depending on the initial conditions. Moreover, there is a third feasible long run outcome: the Quasi Cooperative Culture. This culture is characterized by a high proportion of reciprocal allocators in the population. This economy shows high levels of investment, intermediate levels of trustworthiness and rewards (because of the high proportion of reciprocal allocators) and surplus destruction because reciprocal investors punish the unfair rewards of selfish allocators.

Summarizing, our results show that there is a strong positive causal relationship between the quality of the punishment institutions and the levels of trust and investment achieved in the society. Paradoxically, there is no relationship between the distribution of preferences in the population and the levels of trust exhibited in the society. A high proportion of reciprocators in the investor population is compatible, with both high levels of trust and investment (Fully Cooperative Culture) and, with very low levels (Inefficient Separating Culture). Even if there is a high proportion of reciprocators - potential punishers - in the population, they will not punish opportunistic behaviour if a sufficiently well-functioning institution does not exist.

The intuition behind these results is the following. On the one hand, the existence of strong punishment institutions will generate a significant number of punishers, since it increases the incentives to socialize on preferences for negative reciprocity. This, in turn,

[2] See Fehr and Hoff, 2011 for other non-intergenerational psychological mechanisms that support the endogeneity of preferences.

[3] There is strong evidence showing that many advanced economies display low (and diminishing since 1980) levels of social mobility (Davis and Mazumder, 2017). In less advanced economies where important precapitalist features persist, social mobility is even smaller (see for instance, India and the caste system). However, an interesting extension of our work would be to analyse a society where a positive probability exists of children becoming members of the other social group. This would affect the socialization efforts of parents.

will increase the effectiveness of punishment of future generations. Hence, provided the capacity of punishment is high enough, both types of allocators will set fair returns given the credibility of the threat of punishment. Society would have reached a fully-cooperative culture. On the other hand, if institutions are "weak", cultural transmission operates the other way around: they facilitate the diffusion of selfish preferences, especially among allocators. Summing up, high levels of trust will only appear under the shadow of strong punishing institutions.

Our paper is related to some important strands of literature. It is closely related to the literature on cultural transmission and socialization (Bisin and Verdier, 2000, 2001) and more particularly, to the work on the endogenous determination of preferences and its interaction with institutions (Hauk and Saez, 2002; Olcina and Penyarrubia, 2004, among others). There is also a strand of literature that analyses the coevolution of culture and institutions (see for instance, Tabellini, 2008b; Bisin and Verdier, 2017).

Our analysis focuses on the impact of exogenous formal institutions on preferences and trust. A vast body of literature exists that studies empirically this relationship. This literature relies on experimental evidence (see, for example, Herrmann et al., 2008) and on historical natural experiments where institutions vary in locations with common geographical, cultural, and other socioeconomics characteristics (see, for example, Alesina and Fuchs-Schündeln, 2007). All this research convincingly shows that there is a clear and long-standing influence of institutions on the distribution of preferences in a society (see Tabellini 2008a; Alesina and Giuliano, 2015).

In particular, and more closely related to the problem addressed in our paper, this empirical investigation specifically shows a strong positive correlation across countries and regions between trust and the quality of the legal system and institutions. These correlations are robust when using different measures of institutional quality such as the rule of law, the enforcement of contracts, the strength of property right protection, corruption, etc. (see, for example, Tabellini, 2008a; Algan and Cahuc, 2014).

Our paper presents two important differences with respect to this literature. Firstly, as far as we know, our paper offers the first theoretical analysis of the influence of institutions on trust, by going beyond the correlation between them and providing a causal explanation. The channel that explains the impact of institutions on trust is purposeful cultural transmission of preferences. Secondly, most of these empirical studies define trust as a cultural trait of the preferences or as a personal norm: generalized trust (as defined in the World Values Survey) or sometimes, generalized morality. We work instead with the definition of trust as a behaviour, using Coleman's approach. This allows us to distinguish and make differentiated predictions on the relationship between institutions and trust and on the relationship between institutions and preferences. Our theory predicts a positive relationship between institutions and trust (behaviour) but no relationship between trust and a population's preference distribution. Moreover, our model with a segmented society yields precise predictions on the relationship between the quality of institutions and the distribution of preferences in both the allocator and investor populations.

There are other theoretical studies that, although they do not study the impact of institutions on preferences and trust, share with our paper some specific modelling choices. For instance, Huck and Kosfeld (2007) study a local-interaction model of agents who participate in a neighbourhood watch programme where active participation is costly. As in our approach, punishment institutions are only effective if individuals are willing to engage in an individually costly implementation of these mechanisms.

Herold (2012) presents an indirect evolutionary approach using a haystack model to analyse the evolution of reciprocity preferences for rewarding and preferences for punishing in competition with purely self-interested preferences. Similarly, an important element in our model is the interaction and evolution between the preferences of the allocator for rewarding and the preferences of the investor to punish hostile behaviour in a trust game with a costly punishment phase.

The rest of the paper is organized as follows. Section 2 presents the extended trust game. Section 3 introduces the social preferences and analyzes the behaviour of reciprocal agents. Section 4 studies the equilibria of the extended trust game with heterogeneous preferences in both populations within a generation. Section 5 summarizes the mechanism of cultural transmission of preferences. In section 6, we characterize the steady states of the cultural dynamics and their basins of attraction. Our main results on long run cultures are stated in Section 7. Finally, section 8 discusses the predictions of our model and concludes.

2. The trust game with punishment

We consider a strategic situation in which a player, an investor, randomly drawn from a continuum of investors of mass 1, is matched with another player, an allocator, randomly drawn from a continuum of allocators of mass 1, to play the following extended trust game. The investor (player 1, he) has to decide whether to invest (action I) or not (action N). For simplicity, we assume that if he chooses not to invest, the game ends and both players obtain a payoff of zero. If the investor chooses to make the investment (I), a total surplus of size 1 is produced. We will assume for simplicity that the cost of investment is zero.[4] The allocator (player 2, she) decides on the division of the surplus. Therefore, after observing I, she has to decide which reward b she is willing to pay to the investor, where $0 \leq b \leq 1$. Finally, the investor after observing b, has the option of punishing the allocator at some unitary cost z for him, where $0 \leq z \leq 1$. In particular, the investor can choose to destroy a proportion λ of the allocator's payoff $(1-b)$, where $0 \leq \lambda \leq \lambda^*$ and, $\lambda^* \in (0,1]$. The final material payoffs for the players in case of investment (and punishment) are: $b - z\lambda(1-b)$ for the investor and $(1-b)(1-\lambda)$ for the allocator.

The parameters λ^* and z capture the two relevant features of an institution for punishment: a) how much the investor can damage the allocator and b) how much it costs the investor. In our model, the parameter λ^* can be viewed as the maximal punitive

[4] We could have worked with a surplus of size H > 0 from investment, an outside option for each player l ≥ 0 and a cost of investment c > 0, and similar results would be obtained assuming that (H/2) - c >l.

sanction that the legal system provides to an agent to punish non-compliant behaviour of the other party, while z is the unit cost of implementing the punishment allowed by the law. As we will see in order to have well-functioning punishment institutions both elements are required: a high capacity of legally enforced punishment and a low cost of implementing the law. We will denote parameter λ^* in what follows as the investor capacity for punishment.

An interpretation of these parameters applied to the labour market context is the following. The parameter λ^* is the maximal punishment that a worker can inflict on the firm. For example, if $\lambda^* = 1$, the worker can destroy all the surplus of the firm. It depends on the workers' ability for money burning (sabotage, strikes, etc.) which in turn might depend on the workers' degree of unionization, their ability to organize collectively, their legal rights in the society. It also differs across different types of jobs depending on the strategic position of the worker in the production process. On the other hand, the (unitary) cost of this activity of money burning (parameter z) varies with the firm's capability of finding out and sanctioning this activity in the case of sabotage, or with the difficulty of coordination and organization in the case of strikes.

Suppose now that all players have self-regarding preferences (only concerned with their own payoffs) and that there is complete information. We obtain a Subgame Perfect Equilibrium of the game by backward induction. In the last subgame, selfish investors do not punish because it is costly and does not increase their payoff. Given that the allocator will not be punished, she will set a reward $b = 0$ to the investor and therefore, if there was a small investment cost no matter how small, the optimal action for the investor would be not to invest at the beginning of the game.[5] This is an inefficient outcome in which both players obtain a payoff of zero. In this extended trust game, both the promise of rewarding by the allocator and the threat of punishment by the investor are not credible.

In this paper we will assume that there is heterogeneity of preferences, and that in addition to self-regarding people there is also a significant fraction of the population that exhibits social preferences. In the next section, we introduce this type of preferences.

3. Social preferences

Overwhelming evidence generated by experiments in the laboratory and also everyday's experience, suggest that fairness and reciprocity motives affect the behaviour of many people. By reciprocity we mean the willingness to reward friendly behaviour and the willingness to punish hostile behaviour, even though this may be at a cost. We will assume that in society there are two groups. In each group, there is a certain proportion of people with selfish preferences and the remaining proportion with reciprocal preferences. In particular, we assume that in the investor (allocator) population there is a proportion x_t (y_t) of self-interested (selfish) agents in period t who are motivated exclusively by their own monetary payoffs and a proportion $1 - x_t$ ($1 - y_t$) of agents who have preferences for reciprocity.

We assume that reciprocal players display the inequality aversion preferences introduced by of Fehr and Schmidt (1999). Inequality averse individuals are not just concerned with their monetary payoff, they also aspire to get a fair final payoff compared to the other individual's payoff and, hence, any distribution of payoffs that differs from what is considered a fair division will generate disutility for them. We assume that a fair division is half of the obtained final payoff [6] and that the disutility derived from getting a final payoff smaller than this division is proportional, captured by the parameter $\alpha \geq 1$, to the distance between the other player's payoff and his actual payoff. The parameter α measures the loss from disadvantageous inequality for this type of player. However, inequality averse individuals also dislike advantageous inequality. Their disutility derived from getting a payoff higher than the fair division is proportional, captured by the parameter $1 > \beta > 0.5$, to the distance between his payoff and the payoff obtained by the other party. The parameter β measures the loss from advantageous inequality for this type of player.

Summarizing, we denote the utility of player j of type i as U_{ji} where $j \in \{1, 2\}$, 1 denotes investor and 2 allocator and $i \in \{s, r\}$, where s denotes selfish and r reciprocal.

Let us denote π_j the material payoff of player j and π_k the material payoff of player $k \neq j$. The utility function of a reciprocal player is $U_{jr}(\pi_j, \pi_k) = \pi_j - \alpha \max\{\pi_k - \pi_j, 0\} - \beta \max\{\pi_j - \pi_k), 0\}$. The utility function of a selfish player is: $U_{js}(\pi_j, \pi_k) = \pi_j$, that is, when $\alpha = \beta = 0$.

In our paper, the distribution of preferences in both populations of economic agents in each period is endogenously determined by the decision made by adult players. Note that agents do not know the true type of player they are matched with in period t. However, we will assume that the preference distributions x_t or y_t in both groups are common knowledge.

Reciprocal players have very different policies compared to those of selfish players: specifically, the rewarding policy of the allocators and the punishment policy of investors. Reciprocal allocators compensate generously the cooperative action of the investors and this is why we denote them as fair-minded allocators. The reciprocal investors are willing to punish a low rewarding policy by the allocators, provided that the unit cost of punishment is low enough. Let us derive these policies in turn.

3.1. Punishment policy of reciprocal investors

In contrast to the behaviour of selfish investors, the threat of punishment is credible in the case of reciprocal investors, provided that the unit cost of punishing z is smaller than a critical value which is increasing in α, the parameter that captures their aversion

[5] Note that for simplicity, we do not consider this cost of investment in the rest of the paper. None of the results are altered by this further simplification. However, note that when $c = 0$, there is another Subgame Perfect Equilibrium where the investor invests and the allocator sets $b = 0$.

[6] Here we assume that a fair reward is the egalitarian payoff but the results do not change qualitatively if we allow for another definition of "fair" outcome.

to disadvantageous inequality.

Lemma 1. *Assume $z \leq \overline{z} = \frac{\alpha}{1+\alpha}$, if the reciprocal investor is offered a reward $b \geq 1/2$ by the allocator, he will not punish her. If the allocator sets a reward $b < 1/2$, the investor will punish her with a punishment that depends inversely on the compensation offered by the allocator. In particular, if the investor is offered a reward of b, where $0 \leq b \leq b^* = \frac{1-\lambda^*+\lambda^* z}{2-\lambda^*+\lambda^* z}$, he will punish the allocator choosing λ^*, the maximal possible punishment. If the investor gets an offer b, where $b^* < b < 1/2$, he will punish the allocator choosing $\lambda = \frac{1-2b}{(1-z)(1-b)} < \lambda^*$.*

Proof. See Appendix A. ■

The intuition behind this result is that, by punishing, the reciprocal investor, reduces inequality against himself and this positive effect more than compensates for the reduction in his material payoffs. Note that, in principle, the reciprocal investor will punish his allocator until an egalitarian payoff vector is obtained. But if the return offered by the allocator is very small, an egalitarian payoff is not feasible because there is a limit to the amount of punishment λ^*. In this case, the investor punishes until this limit is reached.

Note that if $z > \overline{z} = \frac{\alpha}{1+\alpha}$, the punishing behaviour of a reciprocal investor is indistinguishable from the behaviour of a selfish investor. Hereafter, we will assume that $z \leq \overline{z}$.

3.2. Rewarding policy of fair-minded allocators

The following lemma states the optimal rewarding policy of fair-minded allocators.

Lemma 2. *The optimal strategy of a fair-minded allocator is to set a reward of half of the surplus ($b = 1/2$) to the investor.*

Proof. See Appendix A. ■

To set a fair reward, $b = 1/2$, is optimal given that a selfish investor never punishes and a reciprocal investor only punishes when $b < 1/2$ as stated in Lemma 1. Therefore, this strategy implies that none of the two types of investor will punish her. Note that starting from an unequal advantageous distribution for the allocator, to give one monetary unit to the investor reduces her material payoff by one unit and consequently her utility. But, it also reduces the advantageous inequality by two units. Therefore, as $\beta > 0.5$, her utility increases by more than one unit. The total effect is an increase in her utility.[7]

3.3. Rewarding policy of selfish allocators

Firstly, it is worth noting that the rewarding policy of a fair-minded allocator does not change when there is incomplete information. However, the rewarding policy of a selfish allocator is indeed affected by the distribution of preferences in the investor population.

To compute the optimal rewarding policy of selfish allocators, note that a particular decision made by the investor might also change her beliefs about the investor's type. We will denote by $\mu(s/I)$ the updated probability that the allocator assigns to the investor being selfish after observing action I. The selfish allocator then has three options: i) to set a low return b, such that $0 \leq b < b^*$, ii) to set an intermediate return b, such that $b^* \leq b < 1/2$ or, finally, iii) to set a generous reward, $b = 1/2$.

We show in Appendix A that the second option is dominated by the other two. It is clear that if the allocator wishes to set a low return, the best option is to set $b = 0$. Thus, we have to compare whether the allocator prefers to offer $b = 0$ or $b = 1/2$. The answer obviously depends on the beliefs regarding the proportion of selfish investors in the population. In the following lemma we state this result:

Lemma 3. *If $\lambda^* \geq 1/2$, a selfish allocator will apply the following rewarding policy:*

a) Set $b = 0$ if $\mu > \frac{\lambda^*-1/2}{\lambda^*}$
b) *Set $b = 1/2$ if* $\mu \leq \frac{\lambda^*-1/2}{\lambda^*}$

Proof. See Appendix A ■

Note that this lemma only holds for $\lambda^* \geq 1/2$. If the investors' capacity of punishment λ^* is smaller than $1/2$, that is, the investor can only destroy at most less than half of the surplus, the optimal behaviour of a selfish allocator is always to set $b = 0$. Thus, her behaviour would be the same as in the absence of the punishment option.

We are ready to obtain the equilibria of this extended trust game with incomplete information in the next section.

4. Equilibria of the trust game with punishment within a generation

In this section, we characterize the Perfect Bayesian Equilibria of the incomplete information sequential game played in each period. First, we obtain the efficient equilibrium of the game. All types of investors choose to invest and all types of allocators pay

[7] If $\beta \leq 0.5$, the optimal rewarding policy of the fair-minded allocator would be indistinguishable from the selfish one.

high (fair) returns and, therefore, there is no punishment or surplus destruction.

Lemma 4. *Assume* $\lambda^* \geq \frac{1}{2}$. *For every* y_t *and* $x_t \in [0, \hat{x}]$, *where* $\hat{x} = \frac{\lambda^* - 1/2}{\lambda^*}$, *there is a Fully Cooperative Equilibrium (FCE) in which both types of investors choose to invest and both types of allocators set a reward of* $b = 1/2$.

Proof. See Appendix B. ■

The equilibrium payoff for all types of players is $1/2$. This equilibrium can only happen with a relatively high fraction of reciprocal investors. Specifically, it exists for a proportion of selfish investors smaller than or equal to a critical value, $\hat{x} = \frac{\lambda^* - 1/2}{\lambda^*}$. Reciprocal investors credibly punish the unfair behaviour of allocators. Thus, if their proportion is high enough, selfish allocators are better off setting high returns and avoiding punishment. Note that the critical fraction of selfish investors $\hat{x}$ depends positively on the investors' capacity for punishment λ^* and can take values between 0 and $1/2$ for $\lambda^* \in [1/2, 1]$. The higher the investors' capacity for punishment λ^*, the smaller the critical proportion of reciprocal investors needed for this equilibrium to exist. Moreover, note that for λ^* smaller than $1/2$, a fully cooperative equilibrium does not exist.

Next we show an equilibrium in which all types of investors choose to invest, but selfish allocators set a reward of $b = 0$ and there is surplus destruction because of punishment.

Lemma 5. *For* x_t *such that* $max\{0, \hat{x}\} \leq x_t \leq 1$ *and for* $y_t \in [0, \hat{y}]$, *where* $\hat{y} = \frac{1}{(1+2(z\lambda^*(1+\alpha)+\alpha(1-\lambda^*))} > 0$, *there is a Quasi Cooperative Equilibrium (QCE) in which both types of investor choose to invest, selfish allocators set* $b = 0$ *and fair-minded allocators set* $b = 1/2$. *Reciprocal investors punish unfair rewards* $b < 1/2$.

Proof. See Appendix B. ■

Note that this is an inefficient equilibrium because reciprocal investors punish the low rewards set by selfish allocators, destroying the maximal possible amount of surplus. This equilibrium exists for a society with a high proportion of selfish investors and a low proportion of selfish allocators. A high proportion of selfish investors is needed for selfish allocators to offer low returns and a low proportion of selfish allocators is needed for reciprocal investors to invest, given that punishment is costly.

Next, we obtain an equilibrium in which the two types of investors choose differently. Selfish investors choose to invest while fair-minded investors decide not to invest. The two types of allocators also choose different rewarding policies.

Lemma 6. *For any* x_t *and for* $y_t \geq \hat{y}$, *there is an Inefficient Separating Equilibrium (ISE) in which a selfish investor chooses to invest and a reciprocal investor chooses not to invest. Selfish allocators set* $b = 0$ *and fair-minded allocators set* $b = 1/2$. *There is no punishment in equilibrium.*

Proof. See Appendix B. ■

Note that, paradoxically, in this inefficient equilibrium, it is the selfish investors who invest while reciprocal investors do not invest. Selfish investors play cooperatively because of the presence of enough fair-minded allocators who pay high returns and reciprocal investors choose not to invest because of the presence of a significant fraction of selfish allocators who pay low returns. As punishment is costly, if the number of selfish allocators is high enough, it is not worthwhile for the reciprocal investor to invest.[8] This implies that this equilibrium only exists for an interval $[\hat{y}, 1]$.

Note also that there is no separating equilibrium in which a selfish investor does not invest and a reciprocal investor decides to invest. In this case, allocators would offer high returns ($b = 1/2$), and selfish investors would deviate, imitating the behaviour of their reciprocal companions.

Finally, there is also a very inefficient equilibrium in which both types of investors decide not to invest. This equilibrium exists for any distribution of preferences in the population of investors but only for a population of allocators composed entirely of selfish agents ($y_t = 1$). We relegate the formal lemma to Appendix B. We do not consider this equilibrium in our forthcoming dynamic analysis for two reasons. Firstly, because we will always assume interior initial conditions of the dynamics and, secondly, because it can be easily shown that these equilibria are unstable steady states of the dynamics.

Note that for $\lambda^* > \frac{1}{2}$ and for $x \leq \hat{x}$ and $y \geq \hat{y}$, there is multiplicity of equilibria (FCE and ISE). Hereafter we will assume that all players select the FCE, because it Pareto-dominates the ISE. If the society settles in this region, all agents expect and coordinate themselves in the Pareto dominant equilibrium.

We can divide the space of pairs of preference distributions (x, y) into three regions, according to the expected and actually played equilibrium. In Fig. 1 we show these equilibrium regions.

5. Cultural dynamics in a segmented society

Human behaviour is governed by preferences that are transmitted through generations and acquired by learning and other ways of social interaction. The transmission of preferences which is the result of social interaction between generations is called cultural transmission. We will draw from the model of cultural transmission of Cavalli-Sforza and Feldman (1981) and Bisin and Verdier (2001).

[8] A similar result is obtained in Fehr et al. (2007) in a principal agent model. The presence of fair principals induces selfish agents to choose high effort levels, while the presence of selfish principals induces the fair agents to provide low effort levels.

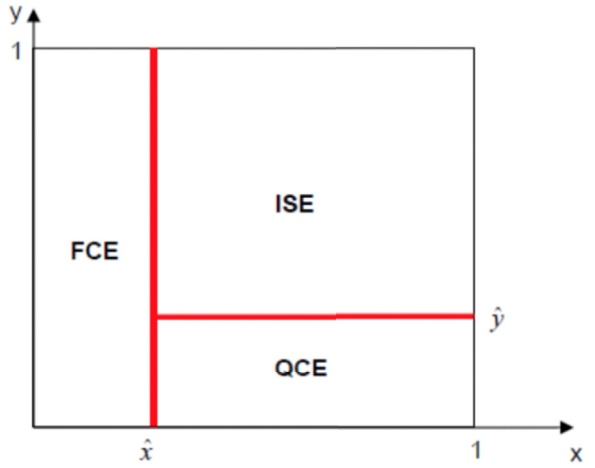


Fig. 1. Equilibrium Regions for parameters $\alpha = 2$ and $\beta = 0.6$ and $\lambda^* > 1/2$.

We consider overlapping generations of agents who only live two periods (as a child and as an adult).[9] The agents can be either investors or allocators and we assume that each type of agent belongs to a group. In the first period, the agent is a child and is educated in certain preferences. In the second period, the agent (as an adult with well-defined preferences), is randomly matched with an adult player from the other group, to play the extended trust game described in section 2. In this second period, any adult player has one offspring and has to make a (costly) decision regarding his/her child's education, trying to transmit his/her own preferences.

The distributions of preferences (x_t, y_t) will evolve according to a purposeful and costly socialization process that we go on to describe. Let $\tau_j^i \in [0,1]$ be the educational effort made by a parent of class j of type i, where remember that $j \in \{1,2\}$, 1 denotes investor and 2 allocator and $i \in \{s,r\}$, s denotes selfish and r denotes reciprocal.

The socialization mechanism works as follows. Consider a parent with i preferences. His child is, firstly, directly exposed to the parent's preferences and is socialized to these preferences with probability τ_j^i chosen by the parent (vertical transmission); if this direct socialization is not successful, with probability $1-\tau_j^i$, he is socialized to the preferences of a role model picked at random in a population composed exclusively of members of the same "social group" (oblique transmission). In this latter aspect, we depart from the usual approach in which oblique transmission takes place in society at large and instead, in our model, children acquire preferences through imitation and learning of cultural models prevailing in their social and cultural environment, that is, in their family and in their social group. We call this process the socialization process in a segmented society.

The transition probabilities[10] P_j^{ik}, determined by this socialization mechanism, can be easily computed and the dynamic evolution of the distribution of preferences is given by the following equation on differences,[11] where v_t is the proportion of selfish types in each social group (v_t is x_t for investors and y_t for allocators): $v_{t+1} = v_t + v_t(1-v_t)[\tau_{jt}^s - \tau_{jt}^r]$

Direct purposeful transmission is justified because parents are altruistic towards their children, but they have some kind of imperfect altruism: their socialization decisions are not based on the purely material payoff expected for their children but on the payoff as perceived by their parents according to their own preferences. This particular form of myopia is called imperfect empathy (Bisin and Verdier, 2001).

Direct transmission is costly. Let $C(\tau_j^i) = \left(\tau_j^i\right)^2/2q$ denote the cost[12] of the education effort τ_j^i, where $j \in \{1,2\}$, $i \in \{s,r\}$ and $q > 0$.

Therefore, a parent from group j of type i chooses the education effort $\tau_j^i \in [0,1]$ at time t, which maximizes

$$P_{jt}^{ii}(\tau_j^i, v_t)V_j^{ii}(v_{t+1}^E) + P_{jt}^{ik}(\tau_j^i, v_t)V_j^{ik}(v_{t+1}^E) - (\tau_j^i)^2/2\ q,$$

where P^{ij} are the transition probabilities and V_j^{ik} is the utility to a parent of group j with preferences i if his child is of type k. Note that the utility V_j^{ik} depends on v_{t+1}^E, which denotes the expectation of the proportion of selfish players in period $t+1$ in their own social group. In this paper, we assume that parents have adaptive or backward looking expectations, believing that the proportion of

[9] As is customary in this class of models we will assume that reproduction is asexual, with a parent per child and thus the population remains constant.

[10] P_j^{ik} denote the probability that a child of a parent of a social group j with preferences i is socialized to preferences k.

[11] We relegate to Appendix C the particular details of this process.

[12] While it is possible to obtain similar results with any increasing and convex cost function we will assume, for simplicity, the quadratic form.

selfish players will be the same in the next period as in the current period, that is, $v^E_{t+1} = v_t$.[13]

Maximizing the above expression with respect to τ^i_j, $j \in \{1,2\}$ and $i \in \{s,r\}$, give us the following optimal education effort functions[14]:

$\tau^s_j(v_t) = q \cdot \Delta V^s_j(v_t) \cdot (1 - v_t)$, $\tau^r_j(v_t) = q \cdot \Delta V^r_j(v_t) \cdot v_t$ Here $\Delta V^s_j = V^{ss}_j - V^{sr}_j$ and $\Delta V^r_j = V^{rr}_j - V^{rs}_j$. That is, ΔV^i_j is the net gain from socializing a child to a parent's own preferences. It can also be interpreted as the cultural intolerance of parents with respect to cultural deviation from their own preferences.

Substituting the optimal education efforts in the differences equation that characterizes the dynamic behaviour of x_t and y_t, we obtain the following dynamical system:

$$\begin{cases} x_{t+1} = x_t + x_t(1 - x_t) \cdot q \cdot [\Delta V^s_j \cdot (1 - x_t) - \Delta V^r_j \cdot x_t] \\ y_{t+1} = y_t + y_t(1 - y_t) \cdot q \cdot [\Delta V^s_j \cdot (1 - y_t) - \Delta V^r_j \cdot y_t] \end{cases} \tag{1}$$

for initial conditions $x_o \in (0,1)$ and $y_o \in (0,1)$.

Next, we characterize the potential stable steady states or attractors of the dynamical system defined by the above non-linear differences equation system from any interior initial condition. In this paper there is a two-speed dynamics. Changes in preferences are gradual while changes in behaviour are instantaneous to maintain equilibrium play. Therefore, in each period, individuals from both social groups coordinate in the PBE of the extended trust game and, assuming adaptive expectations, they believe that this equilibrium will be played by the next generation.

In the following section, we compute the levels of cultural intolerance, the socialization efforts and the stable steady states in the different regions characterized by different Perfect Bayesian equilibria.

6. The dynamical system and the incentives to socialize

Evolution of the economy in the different equilibrium regions is fully determined by the evolution of the population preference distributions x and y. It is given by the sequence $\{x_t, y_t\}_{t=0}^{\infty}$ that is governed by the two-dimensional non-linear discrete dynamical system (1). In this section, we calculate the levels of cultural intolerance ΔV^i_j for the different PBE regions and characterize the dynamical system and its possible stable steady states. Remember that when $\lambda^* > \frac{1}{2}$ and $y \geq \hat{y}$ and $x \leq \hat{x}$ there is multiplicity of equilibria (ISE and FCE) but we will assume that players coordinate in the efficient one, that is, the FCE.

6.1. The dynamical system in the FCE region

In this region, the proportion of selfish investors in the population is relatively small ($x \leq \hat{x}$) and parents (investors and allocators) play and expect to be played the Full Cooperative Equilibrium, in which both types of investors choose to invest and both types of allocators set $b = 1/2$.

Note that $V^{ik}_j = 1/2$ for all types of parents. Therefore, the net gains for any type of parent of any social group obtained from transmitting their own preferences ΔV^{ik}_j, that is, their levels of cultural intolerance, are zero. But this, in turn, implies that there are no incentives at all for socialization for any parent. Therefore, all the optimal education effort functions are zero and, consequently, the distribution of preferences in both populations will remain unchanged, that is $x_{t+1} = x_t$ and $y_{t+1} = y_t$. This result is stated in the following lemma.

Lemma 7. *If $\lambda^* > 1/2$, any pair of distributions (x,y) such that $x \leq \hat{x}$ is a steady-state of the cultural dynamics.*

In other words, in this case there are no incentives to socialize and each of the two populations remains locked in the initial distribution of preferences.

In the following subsections, we show that at most there can be a unique stable steady state in the QCE region denoted by (x', y') and the same is true for the ISE region denoted by $(\overline{x}, \overline{y})$.[15]

6.2. The dynamical system in the QCE region

In this region, where $x \geq \hat{x}$ and $y \leq \hat{y}$, there is a relatively high proportion of selfish investors and a sufficiently high proportion of fair-minded allocators. In this equilibrium both types of investor choose to invest and selfish allocators set $b = 0$ and fair-minded allocators set $b = 1/2$.

All types of parents have positive levels of cultural intolerance and consequently, they will have incentives for active socialization. For example, let's consider a selfish investor parent. If his child turns out to be selfish, his expected payoff is $(1 - y)/2$, while if his

[13] Another possibility would be to assume rational or forward-looking expectations, that is, $v^E_{t+1} = v_{t+1}$. None of the main results will change qualitatively. For simplicity, we work with adaptive expectations to avoid some possible discontinuities that might generate multiplicity or non-existence of perfect foresight paths in the neighbourhood of $\hat{y}$.

[14] In order to have interior solutions the parameter q must be chosen small enough so that in equilibrium $\tau^i_j < 1$.

[15] The reader can check that $(x = 1, y = 1)$ and $(x = 1, y = 0)$ are both unstable steady states.

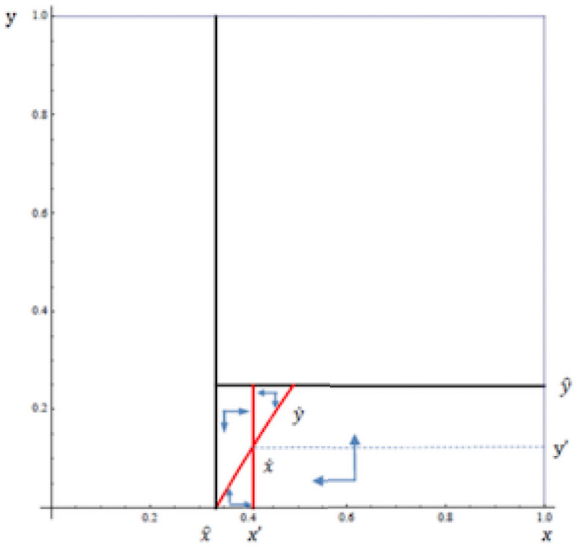


Fig. 2. Dynamical system in the QCE region.

child turns out to be reciprocal, his expected payoff will be $(1-y)/2 - yz\lambda^*$, since his child will punish a selfish allocator. Therefore, $yz\lambda^* > 0$ is the level of cultural intolerance of the selfish investor parent.

We can check that the levels of cultural intolerance in each group depend on the preference distribution of the other group and are positive in the QCE region, and, the optimal socialization effort functions will also depend on both preference distributions (x, y). The precise calculation of the levels of cultural intolerance and the optimal education effort functions for both types of parents is relegated to Appendix C.

The evolution of the economy in the QCE region is fully determined by the dynamical system *(1)*. The analysis of the dynamical system will require the derivation of its phase diagram based on the characterization of the x-demarcation curve and the y-demarcation curve and the forces that operate on the system when the variables are not in the steady state equilibrium.

The x-demarcation curve is the geometric locus of all pairs (x_t, y_t) under which x_t is in a steady-state, that is, $x_{t+1} - x_t = 0$. Similarly, the y-demarcation curve is the geometric locus of all pairs (x_t, y_t) under which y_t is in a steady-state, that is, $y_{t+1} - y_t = 0$. Along these curves the distribution of preferences in the investor (allocator) population remains constant over time.

A steady state equilibrium is characterized by an intersection of both demarcation curves. The following lemma characterizes the possible stable steady states of the dynamics in this region.

Lemma 8. *The dynamical system (1) has a stable steady state* (x', y') *in the QCE region, where* $x' = \frac{z}{\alpha(1-z)}$ *and* $y' = \frac{1/2-(\frac{\alpha(1-z)-z}{\alpha(1-z)})\lambda^*}{\beta(1-\frac{\lambda^*(\alpha(1-z)-z)}{\alpha})}$, *if* $x' > \hat{x}$ *and* $y' < \hat{y}$. *Otherwise, there is not a stable steady state in the QCE region.*

Proof. Appendix C. ■

The phase diagram qualitative analysis yields that there is at most one stable steady state or attractor (x', y') of the dynamical system in the QCE region. To supplement this phase diagram analysis we conduct a local stability analysis of this fixed point in Appendix D. Note that if $x' > \hat{x}$ *or* $y' > \hat{y}$, the dynamics will lead to the FCE region or, alternatively, to the ISE region.

In Fig. 2, we represent the phase diagram of the nonlinear difference equation system in two variables and the steady state equilibrium. The directional arrows indicate the intertemporal movement of x_t and y_t.

6.3. The dynamical system in the ISE region

We now run the analysis for the region where the ISE is played, that is, when $x > max\{0, \hat{x}\}$ and $y \geq \hat{y}$, in other words, when the proportion of selfish allocators is relatively high. In this equilibrium the selfish investor chooses to invest and reciprocal investors choose not to invest. The selfish allocator sets $b = 0$ and the fair-minded allocator sets $b = 1/2$.

In this case, all types of parents have strictly positive levels of cultural intolerance and consequently, they will have incentives for active socialization. For example, a selfish investor parent expects that if his child is selfish, like him, he will obtain an expected payoff of $(1-y)/2$ and if his child turns out to be reciprocal he will obtain a payoff of zero. Therefore, the selfish investor parent has a positive level of cultural intolerance which depends on the proportion of fair-minded allocators $(1-y)$, and on the gains from cooperation.

We can check that the levels of cultural intolerance in each group also depend on the preference distribution of the other group and are also strictly positive. Note also that the optimal socialization effort functions, τ_j^i, depend on both preferences distribution

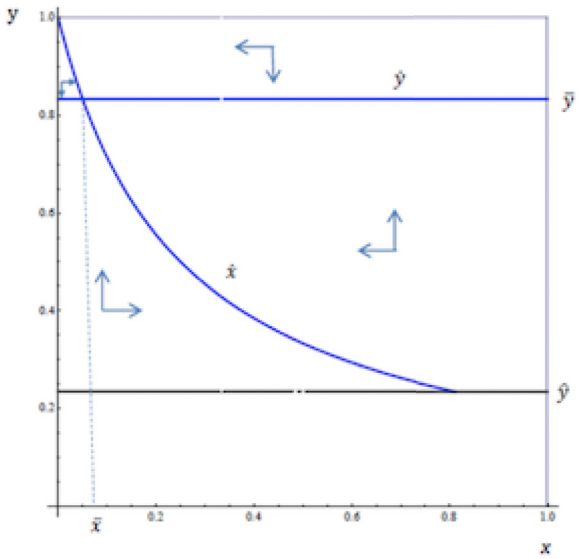


Fig. 3. Dynamical system in the ISE region.

(x, y) and are positive. The precise calculation of the levels of cultural intolerance and the optimal education effort functions for both types of parents is relegated to Appendix C.

As before, the evolution of the economy into the ISE region is determined by the dynamical system (1). The analysis of the dynamical system will also require the derivation of its phase diagram based on the characterization of the demarcation curves in this region. A steady state equilibrium is characterized by an intersection of both demarcation curves. The following lemma characterizes the possible stable steady states of the dynamics in this region.

Lemma 9. *The dynamical system (1) has a stable steady state* $(\overline{x}, \overline{y})$ *in the ISE region, where* $\overline{x} = \frac{\beta - 1/2}{\alpha}$ *and* $\overline{y} = \frac{1}{2\beta}$, *if* $\overline{x} > \hat{x}$ *and* $\overline{y} > \hat{y}$. *Otherwise, there is not a steady state in the ISE region.*

Proof. Appendix C

The phase diagram qualitative analysis yields that there is at most one stable steady state or attractor $(\overline{x}, \overline{y})$ of the dynamical system in the ISE region. To supplement this phase diagram analysis we perform a local stability analysis of this fixed point in Appendix D. Note that if $\overline{x} < \hat{x}$ *or* $\overline{y} < \hat{y}$, the dynamics will lead to the FCE region or, alternatively, to the QCE region.

In Fig. 3, we represent the phase diagram of the nonlinear difference equation system in two variables and the steady state equilibrium. The directional arrows indicate the intertemporal movement of x_t and y_t.

6.4. Regions in the space of institutional parameters (λ^*, z)

In the previous subsections we have characterized the possible steady states of the dynamical system in different regions of the space of preference distributions (x, y) by means of phase diagram analysis. Next, we present these results in terms of the space of institutional parameters $\lambda^*$and z, as we are interested in the evolution of behaviour and preference distribution stressing the importance of punishment institutions. Therefore, we split the space (λ^*, z) for $\lambda^* \in [0, 1]$ and $z \in [0, \overline{z}]$, into regions in which the different attractors or stable steady states of the dynamical systems can take place.

Firstly, it should be remembered that whenever $\lambda^* > 1/2$, FCE will always be an attractor of the dynamics whenever $x_0 \leq \hat{x}(\lambda^*)$.

Secondly, as follows from Lemma 8, QCE can be a local attractor if $x'(z, \alpha) > \hat{x}(\lambda^*)$ and $y'(z, \lambda^*, \alpha, \beta) < \hat{y}(z, \lambda^*, \alpha)$. First, we define the function $\lambda_x^*(z; \alpha)$ by equating $x'(z, \alpha) = \hat{x}(\lambda^*)$ and second, the function $\lambda_y^*(z; \alpha, \beta)$ by equating $y'(z, \lambda^*, \alpha, \beta) = \hat{y}(z, \lambda^*, \alpha)$. Thus, for all pairs (λ^*, z) that lie in between the functions $\lambda_x^*(z; \alpha)$ and $\lambda_y^*(z; \alpha)$, the QCE is an attractor.

It is worth noting that both functions are increasing in the space (λ^*, z). And for all z, belonging to the interval $[0, \overline{z}]$, the function $\lambda_x^*(z; \alpha)$ is greater than $\lambda_y^*(z; \alpha, \beta)$, therefore the area in between the two functions is non-empty. The complete analysis of these two functions is relegated to Appendix F.

Finally, we obtain the region in which ISE is an attractor. As follows from Lemma 9, there is an steady state where the ISE is played if $\overline{x}(\alpha, \beta) > \hat{x}(\lambda^*)$ and $\overline{y}(\beta) > \hat{y}(z, \lambda^*, \alpha)$. First, we define the critical value of the capacity of punishment $\overline{\lambda}^*(\alpha, \beta)$, obtained by equating $\overline{x}(\alpha, \beta) = \hat{x}(\lambda^*)$.This value $\overline{\lambda}^*(\alpha, \beta)$, that does not depend on z, determines the values of λ^* such that only if $\lambda^* < \overline{\lambda}^*(\alpha, \beta)$ can the ISE be an attractor of the dynamics. Second, we define the function $\overline{\lambda}_y^*(z; \alpha, \beta)$ by equating $\overline{y}(\beta) = \hat{y}(z, \lambda^*, \alpha)$.This implies that for all pairs (λ^*, z) above this function ISE cannot be an attractor of the dynamics. The function $\overline{\lambda}_y^*(z; \alpha, \beta)$ is increasing in the space

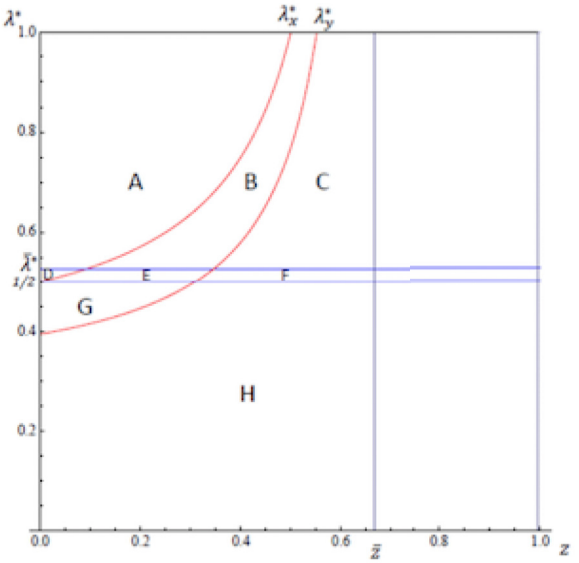


Fig. 4. Regions in the (λ^*, z) space.

(λ^*, z). Only when the pairs (λ^*, z) are located below both $\overline{\lambda}^*(\alpha, \beta)$ and $\overline{\lambda}^*_y(z; \alpha, \beta)$ simultaneously is the ISE an attractor. It can be checked that the binding restriction is $\overline{\lambda}^*(\alpha, \beta)$. The complete analysis of this value $\overline{\lambda}^*(\alpha, \beta)$ and the function $\overline{\lambda}^*_y(z; \alpha, \beta)$ is relegated to Appendix F.

Fig. 4 in section 7 depicts these functions $\overline{\lambda}^*(\alpha, \beta)$, $\lambda^*_x(z; \alpha)$ and $\lambda^*_y(z; \alpha)$ in the (λ^*, z) space, yielding the partition of this space into different regions. This figure is depicted for a given value of α and β (in this example $\alpha = 2$ and $\beta = 0.6$).

We are now ready to state our main results.

7. Main results: cultures in the long run

We adhere to the notion of culture, used by Rob and Zemsky (2002), as a stable, or self-reproducing pattern of behaviour in a group (or a society). Therefore, we identify it as a stable steady state or attractor of the preference dynamics.

Definition 1. *A Culture is any stable steady state of the dynamical system (1) where the same Perfect Bayesian Equilibrium of the trust game with punishment is played.*

We denote a Fully Cooperative Culture *(FCC) as* any stable steady state of the cultural dynamics where a Fully Cooperative Equilibrium is played. An analogous definition applies for the Quasi Cooperative Culture (QCC) and the Inefficient Separating Culture *(ISC)*.

Our model yields different long run outcomes, cultures, depending basically on the values that characterize the punishment institutions, λ^* and z. The results also depend, on the parameters α and β that characterize the degree of inequality aversion of reciprocal types and, in some cases, on the initial conditions of the dynamics (the history of the society).

As we previously mentioned, Fig. 4 shows a partition of the space (λ^*, z) into different regions.[16] The regions defined in this figure illustrate the main results of this paper. We designate the different regions by capital letters.

7.1. A fully cooperative culture

Our first result characterizes when the society achieves the maximal amount of trust and trustworthiness, reaching an efficient outcome. This will occur when the Fully Cooperative Culture is either a global or a local attractor of the dynamics. The next proposition shows that this will happen only for a high capacity of punishment λ^*.

Proposition 1. *Assume* $\lambda^* > 1/2$, $z \leq \overline{\overline{z}}(\alpha)$, $\alpha \geq 1$ *and* $\beta > 0.5$, *then FCC is an attractor of the dynamic system (1).*

Our main interest is to characterize when a culture is a global attractor of the dynamics, that is, when the society converges to this culture from any interior initial condition of the dynamics, $x_o \in (0, 1)$ and $y_o \in (0, 1)$. The next proposition shows when FCC is a global attractor.

[16] Remember that this figure is depicted for a given value of α and β (in this example $\alpha = 2$ and $\beta = 0.6$).

Proposition 2. *Assume* $\lambda^* > \overline{\lambda}^*(\alpha,\beta) > 1/2$, $z \leq \overline{z}(\alpha), \alpha \geq 1$ *and* $\beta > 0.5$, *then FCC is a global attractor of the dynamic system (1),*

i) *for pairs* (λ^*, z) *such that* $\lambda^* > \lambda_x^*(z;\alpha)$ *for any* z. *That is, for pairs* (λ^*, z) *in Region A or*
ii) *for pairs* (λ^*, z) *such that* $\lambda^* < \lambda_y^*(z;\alpha)$ and $z < z_y$, *where* z_y *is such that* $\lambda_y^*(z;\alpha) = 1$,*or for pairs* (λ^*, z) *such that* $z > z_y$. *That is, for pairs* (λ^*, z) *in Region C.*

Proof. The proof of these propositions follows from the qualitative analysis of the phase diagrams corresponding to the different regions which can be found in Appendix E. ■

Proposition 2 shows one of our main results. If investors have a high capacity of punishment ($\lambda^* > \overline{\lambda}^*(\alpha,\beta)$), the society converges from any initial condition to the fully cooperative culture provided the unitary cost of punishment, z, is low enough (region A), or alternatively, this cost is relatively "high" (region C). Only for "intermediate" values of z, is FCC only a local attractor, that is, the society converges to FCC only for a certain set of initial conditions.

Let us provide some intuition for regions A and C. For an initial condition of the dynamics such that the proportion of selfish investors is initially quite high, $x_o > \hat{x}$, all agents coordinate either on the ISE or in the QCE depending on the proportion of selfish allocators (y_o), but in both cases the incentives to socialize of the reciprocal investors are higher than those of selfish investors. Therefore, the proportion of selfish investors x decreases over time until eventually it reaches the critical value $\hat{x}$.

Note that in this situation, the preference distribution in the FCC is characterized by $x \leq \hat{x}$ (i.e. a sufficiently high fraction of reciprocal investors), but the final distribution of preferences in the allocators population depends only on the initial condition of the dynamics.

The fully cooperative culture FCC is only a local attractor for intermediate values of z (region B). For these institutional values the economy will converge to FCC if the initial preference distribution is such that there are many reciprocal investors, $x_o < \hat{x}$, or many selfish allocators, $y_0 > \hat{y}$. Nevertheless, if the society starts in an initial situation with a not sufficiently high proportion of reciprocal investors, $x_o > \hat{x}$, and a high proportion of fair-minded allocators, $y_0 < \hat{y}$, then it will get stuck in an inefficient QCC culture.

For values of λ^* such that $\overline{\lambda}^*(\alpha,\beta) > \lambda^* > 1/2$, FCC is again only a local attractor either when $x_o < \hat{x}$, (regions E and F) or when $y_0 < \hat{y}$ and the cost of punishment z is very low (region D).

In summary, if there is a high institutional capacity of punishment in the society, a Fully Cooperative Culture will eventually be reached from any initial condition. This culture will be characterized by a high proportion in the investor population of reciprocal individuals who are willing to punish unfair rewards and any composition of the allocator population, only depending on the initial condition of the dynamics. This economy will exhibit the highest levels of trust (investment) and trustworthiness (rewards). No punishment would be observed, although the behaviour of selfish individuals would be indistinguishable from that of reciprocators because of the existence of a credible threat of punishment of opportunistic behaviour.

7.2. An Inefficient Separating Culture

The Inefficient Separating Culture is an inefficient culture with a very high proportion of reciprocal investors that prefer not to invest and a minority of selfish investors that invest. The society will converge to this culture for a low capacity for punishment, $\lambda^* \leq \overline{\lambda}^*(\alpha,\beta)$ and will display the lowest levels of trust and trustworthiness. The next proposition states when the ISC is either a local or a global attractor of the dynamics.

Proposition 3. *Assume* $\lambda^* \leq \overline{\lambda}^*(\alpha,\beta)$, $z \leq \overline{z}(\alpha), \alpha \geq 1$ *and* $\beta > 0.5$, *then ISC is an attractor of the dynamic system (1).*

The following proposition characterizes when ISC is a global attractor of the dynamics.

Proposition 4. *Assume* $\lambda^* \leq 1/2$, $z \leq \overline{z}(\alpha), \alpha \geq 1$ *and* $\beta > 0.5$, *then ISC is a global attractor of the dynamic system (1) for pairs* (λ^*, z) *such that* $\lambda^* < \lambda_y^*(z;\alpha)$ *for any* z, *that is, for pairs* (λ^*, z) *in region H.*

Proof. The proof of these propositions follows on from the qualitative analysis of the phase diagrams corresponding to the different cases which can be found in Appendix E.

Let us provide some intuition for region H, where the society converges from any initial condition to the inefficient culture ISC. Even if the initial condition of the dynamics starts with a low proportion of selfish investors (low x_o) and selfish allocators (low y_o) and, although the incentives to socialize of selfish investors are high, the incentives to socialize of selfish allocators are higher than those of fair-minded allocators, $\tau_2^s > \tau_2^r$. Therefore, the proportion of selfish allocators y increases and eventually, $y > \hat{y}$. The dynamics reaches the ISE region.

Note that the ISC is an inefficient culture with a very high proportion of reciprocal investors that do not invest. Only a minority of selfish investors invest, because of the presence of a minority of rewarders in the population of allocators. [17]

The ISC culture is only a local attractor for some values of λ^* smaller than $1/2$ but nevertheless, relatively high (notice that $\lambda_y^*(z=0) < 1/2$) and for low values of z, that is, in region G. In this case, depending on the initial condition, the society can

[17] Note that if β tends to 0.5, then $\overline{x}$ is very close to zero and $\overline{y}$ is very close to one.

converge to the Quasi Cooperative Culture. Interestingly, only the initial condition on the preference distribution of the population of allocators matters. If the initial proportion of selfish allocators is very high, $y_0 > \hat{y}$, the society converges to the ISC. If this proportion is low, $y_0 < \hat{y}$, then the society converges to the *QCC*.

For values of λ^* such that $\overline{\lambda}^*(\alpha, \beta) > \lambda^* > 1/2$, ISC is again a local attractor either when $x_o > \hat{x}$ and $y_0 > \hat{y}$ (regions D and E) or when the cost of punishment z is very high (region F).

In summary, an inefficient culture ISC will prevail in the presence of a low institutional capacity for punishment in the society. This culture will be characterized by a high proportion of selfish allocators and also a very high proportion of reciprocal investors. But, given the existing low institutional capacity for punishment, reciprocal investors are not willing to punish unfair rewards and prefer not to invest. This economy shows the lowest levels of trust and investment, of trustworthiness and rewards and again, no punishment is observed.

7.3. An inefficient Quasi Cooperative Culture

Finally, let us characterize the institutional conditions under which the society will converge to a culture in which both types of investors invest but there will be punishment and surplus destruction because selfish allocators set low rewards.

Proposition 5. *Assume $z \leq \overline{z}(\alpha), \alpha \geq 1$ and $\beta > 0.5$, QCC is a local attractor of the dynamic system (1), for pairs (λ^*, z) such that $\lambda_x^*(z;\alpha) > \lambda^* > \lambda_y^*(z;\alpha)$, for any z, that is, for pairs (λ^*, z) in regions B, E and G.*

Proof. The proof of this proposition follows on from the qualitative analysis of the phase diagrams corresponding to the different cases which can be found in Appendix E. ■

Note that if there is an initial condition of the society with a relatively greater number of selfish investors and not as many selfish allocators, $x_0 > \hat{x}$ and $y_0 < \hat{y}$, then, under these institutional values (λ^*, z), the economy gets stuck in the inefficient QCC.

The Quasi Cooperative Culture might appear in two very different institutional scenarios: with a high capacity of punishment $\lambda^* > 1/2$ and intermediate values of the unit cost of punishment z (regions B and E) or with a relatively low capacity of punishment $\lambda^* < 1/2$ and low values of the unit cost of punishment z (region G). But, in both cases, QCC is characterized and supported by a high proportion of fair-minded allocators in the population. This economy shows high levels of investment, intermediate levels of trustworthiness and rewards (because of the high proportion of fair-minded allocators) and surplus destruction because reciprocal investors punish the unfair rewards of selfish allocators.

7.4. Modifying the degrees of inequality aversion (α and β)

The parameter α, that measures the degree of disadvantageous inequality aversion of the agents, is important because it determines the strength of the disposition of reciprocal investors to punish. More precisely, it determines the upper bound, $\overline{z}$, in the unitary cost of punishment such that for smaller costs they are willing to punish. But the parameter α also determines the level of cultural intolerance of reciprocal investors. Specifically, the higher α is, the higher these levels are, both in the ISE region and in the QCE region. This obviously increases the relative socialization incentives of reciprocators. Both effects increase the basin of attraction of the global attractors of the dynamics because eventually, the proportion of selfish investors x would fall below the critical value $\hat{x}$.

This is expressed formally in the following lemma:

Lemma 10. *The sets of pairs (λ^*, z) for which both FCC and ISC are global attractors become larger as α increases.*

Proof. See Appendix G.

Note that when α tends to infinity, then $\overline{z}$ tends to 1. In this limiting case, for $\lambda^* > 1/2$ and any $z < 1$, FCC is a global attractor and for $\lambda^* \leq 1/2$ and any $z < 1$, ISC is a global attractor. The QCC disappears as an attractor of the dynamics.

The parameter β measures the degree of advantageous inequality aversion of the agents. Note that changes in β in our feasible set, that is, for $1 > \beta > 0,5$, do not change the behaviour of fair-minded allocators which, as we already know, is to set a fair reward, $b = 1/2$. However, the parameter β also determines their levels of cultural intolerance. Specifically, the higher β is, the higher these levels are, both in the *ISE* region and in the QCE region. This obviously increases the relative socialization incentives of fair-minded allocators. In this case, the basin of attraction of the local attractor QCC increases and the basin of attraction of the global attractors of the dynamics diminishes. In summary, changes in parameter β, operate in the opposite direction to changes in parameter α.

These results are expressed formally in the following lemma:

Lemma 11. *The sets of pairs (λ^*, z) for which both FCC and ISC are global attractors become smaller as β increases.*

Proof. See Appendix G.

8. Discussion and concluding remarks

Our main result provides a rationale for the existence of a strong positive relationship between the quality of a society's punishment institutions (measured by their capacity for and their cost of punishing hold up) and the level of trust achieved (measured by the aggregate amount of investment when there is a risk of being held up). Our model also predicts a positive relationship between the

quality of a society's punishment institutions and the level of trustworthiness achieved (measured by the rewards set by allocators). Trust and trustworthiness only flourish in the shadow of efficient sanctioning institutions. Nevertheless, sanctioning institutions need the presence of a high proportion of reciprocators in the investor population, punishers, willing to make use of them. Our model proves how the existence of strong institutions together with the intergenerational transmission of preferences result in a high proportion of punishers in the population. Therefore, the driving force behind our results is the need for strong and effective institutions to incentivize the socialization of children to preferences for punishment.

However, our theory predicts that there is no relationship between the distribution of preferences in the population and the levels of trust exhibited in a society. A high proportion of reciprocators in the investor population is compatible both with high levels of trust and investment (FCC) and with very low levels of trust and investment (ISC). Even if there is a high proportion of reciprocators - potential punishers - in the population, they will not punish opportunistic behaviour if a sufficiently well-functioning institution does not exist. Consequently, they choose not to invest because of the risk of being held up.

The efficient FCC can exist with any distribution of preferences in the allocator population because, given the greater strength of the punishment institutions the behaviour of selfish and fair-minded allocators is indistinguishable. But our theory predicts that an inefficient culture (ISC) will display a high proportion of selfish allocators. However, QCC cultures display a high proportion of fair-minded allocators. In other words, in the previous two cultures, the strength of the punishment institutions (for example, workers' unions in the labour market) would be negatively correlated with the proportion of selfish allocators (for instance, employers) in the population.

We have shown throughout the paper the key role of the investors' capacity for punishment to achieve trust and efficiency in the economy, but what happens when there is no punishment institution? Suppose that we remove from our investment game the punishment option, that is, only the presence of a fraction of reciprocal agents operates. Note first that the FCE equilibrium is no longer a perfect Bayesian equilibrium of this extended trust game, because obviously selfish allocators will always set low returns. But the remaining set of previously obtained equilibria are still equilibria in the new game.

However, the QCE equilibrium is now not a stationary point of the dynamics under any initial condition: both types of investors choosing to invest is not a long run outcome of cultural evolution. The only possible steady state is the inefficient equilibrium ISE. Therefore, positive reciprocity is not enough to achieve an efficient culture. A significant capacity for punishment of investors is also needed in order to make the threat of punishment a powerful enough tool to enhance efficiency and cooperation.

Our theory allows for different promising future extensions.

Obviously, our game represents an asymmetric relationship, in the sense that only allocators can hold up on investors, while the latter lacks this capacity with respect to the allocators. A more symmetric situation would be the object of future research. Nevertheless, we claim that a situation in which investors put at risk much more than allocators in their bilateral relationship, again abstracting from reputational considerations, is very widespread in real-life investor-allocator relationships.

In the present analysis, we have assumed there to be no social mobility at all in the society (a segmented society). Although we believe that this could be a good approximation to many contemporary societies, an interesting extension of our work would be to analyse societies that display positive levels of social mobility, where there is a positive probability of children becoming members of a different social group. Note that this will only affect the direct socialization incentives of parents.

On a final note, we have focused on the impact of exogenously given institutions on culture. Therefore, our theory is silent about the creation of these strong punishment institutions. However, in real life, both culture and institutions are endogenous and there is a two-way relationship between them. A society with a certain culture, characterized by a preference and behaviour distribution, will choose certain institutions. In turn, these institutions lead to the transmission across generations of certain cultural values and behaviours. In other words, culture and institutions coevolve. The study of this interaction in our framework belongs to future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendixes

Appendix A. Proof of Lemma 1

Optimal punishing policy of reciprocal investors.

If the reciprocal investor receives a compensation $b \geq 1/2$, he will not punish the allocator. Effectively, the utility of punishing for this investor would be:

$U^{p}_{1r}(b - z\lambda(1-b), (1-b) - \lambda(1-b)) = b - z\lambda(1-b) - \beta[b - z\lambda(1-b) - ((1-b) - \lambda(1-b))] = (b - z\lambda(1-b) - 2\beta b + \beta - \lambda\beta(1-b)(1-z))$.

It can be checked that this expression is maximized when $\lambda = 0$.

Next, we assume that $0 \leq b < 1/2$.

Note that in this case this investor will punish the allocator provided that the unitary cost of punishing z is smaller than $\frac{\alpha}{1+\alpha}$, since $U_{1r}^{p} > U_{1r}^{np}$, where $U_{1r}^{p}(b - z\lambda(1-b), (1-b) - \lambda(1-b)) = (b - z\lambda(1-b)) - \alpha[(1-b) - \lambda(1-b) - (b - z\lambda(1-b))] = (b(1+2\alpha) - \alpha + \lambda(z(1+\alpha) - \alpha)(b-1))$ (the utility of punishing by the reciprocal investor) and $U_{1r}^{np}(b, (1-b) = b - \alpha((1-b) - b) = (b(1+2\alpha) - \alpha)$, (the utility of not punishing by this investor).

We call $\overline{\lambda}$ the proportion of punishment that equals the material payoff of both players after punishment, that is, $(1-b) - \overline{\lambda}(1-b) = b - z\,\overline{\lambda}(1-b)\Big)$. The value of $\overline{\lambda}$ is $\frac{1-2b}{(1-z)(1-b)}$.

If $\lambda > \overline{\lambda}$, the payoff of the allocator will be smaller than the payoff of the investor. To maximize his utility the reciprocal investor has to set λ as small as possible because this punishment generates advantageous inequality and this option is dominated by punishing until both players get the same payoff, that is, setting $\lambda = \overline{\lambda}$.

Note that, for instance, for $b = 0$, the optimal amount of punishment is $\overline{\lambda} = \frac{1}{1-z}$, but this is greater than 1. We assume that the maximal proportion of punishment is λ^*, with $\lambda^* \leq 1$. So, for these low offers the optimal punishment is the maximal one, λ^*. We equate $\frac{1-2b}{(1-z)(1-b)}$ to λ^*, in order to obtain the threshold reward that will trigger the maximal level of punishment λ^*. This level is $b^* = \frac{1-\lambda^*+\lambda^* z}{2-\lambda^*+\lambda^* z}$.

In summary, the optimal punishment policy of the reciprocal investor is:

For $0 \leq b \leq b^* = \frac{1-\lambda^*+\lambda^* z}{2-\lambda^*+\lambda^* z}$, the optimal proportion of punishment will be $\lambda = \lambda^*$.

For $b^* < b < 1/2$, the optimal proportion of punishment will be $\lambda = \frac{1-2b}{(1-z)(1-b)} < \lambda^*$.

For $b \geq 1/2$, the reciprocal investor will not punish.

Proof of Lemma 2

Optimal rewarding policy of a fair-minded allocator.

If a fair-minded allocator offers a compensation of $b = 1/2$, her utility would be: $U_{2r}(1/2, 1/2) = 1/2$, given that a selfish investor never punishes and a reciprocal investor only punishes when $b < 1/2$ as stated in Lemma 1.

Rewarding with $b = 1/2$ dominates any reward $b > 1/2$, because the latter reduces the allocator's material payoff and increases disadvantageous inequality.

If the fair-minded allocator decides to offer a compensation of b, where $0 \leq b < 1/2$ and she is punished by a proportion λ her utility would be:

$U_{2r} = (1-b)(1-\lambda) - \beta[(1-b)(1-\lambda) - b + z\lambda(1-b)]$.

Equivalently, $U_{2r} = b[(2\beta - 1) - \lambda(\beta - 1) + \beta z\lambda] + (1-\lambda)(1-\beta) - \beta z\lambda$.

This expression is increasing in b given that $\beta > 1/2$ for any $\lambda \in [0, \lambda^*]$. Then the optimal reward is $b = 1/2$.

Proof of Lemma 3

Optimal rewarding policy of a selfish allocator with incomplete information.

The expected payoff of offering $0 \leq b < b^*$ is $\mu[(1-b)] + (1-\mu)[(1-b) - \lambda^*(1-b)] = (1-b)[(1-(1-\mu)\lambda^*]$, given that the selfish investor does not punish and the reciprocal investor applies the maximal punishment λ^*. It is easy to check that setting $b = 0$ dominates any other $b > 0$ whenever $b < b^*$. So, the expected payoff of offering $b = 0$ is $[1 - (1-\mu)\lambda^*]$.

On the other hand, offering $b = 1/2$, generates a payoff of $1/2$ to the allocator, since none of the types of investors punishes her.

Then, for the allocator, offering $b = 0$ is better than offering $b = 1/2$ when $\mu > \frac{\lambda^* - 1/2}{\lambda^*}$. And offering $b = 1/2$ is better than offering $b = 0$ when $\mu \leq \frac{\lambda^* - 1/2}{\lambda^*}$.

Next, we have to prove that to offer $b = 0$ is better than to offer $b^* \leq b < 1/2$,when $\mu > \frac{\lambda^* - 1/2}{\lambda^*}$.

The expected payoff of the selfish allocator of offering b, such that, $b^* \leq b < 1/2$ is $\mu[(1-b)] + (1-\mu)[(1-b) - \lambda(1-b)] = \mu[(1-b)] + (1-\mu)[(1-b) - \frac{1-2b}{(1-z)(1-b)}(1-b)]$.

Assume that $\mu > \frac{\lambda^* - 1/2}{\lambda^*}$. Then, we have to prove that $[1 - (1-\mu)\lambda^*] > \mu[(1-b)] + (1-\mu)[(1-b) - \frac{1-2b}{(1-z)(1-b)}(1-b)]$.

It is easy to verify that this expression holds for $\mu = 1$. It also holds for $\mu = \frac{\lambda^* - 1/2}{\lambda^*}$.

As the function $(b - (1-\mu)\lambda^* + (1-\mu)\frac{1-2b}{(1-z)(1-b)}) > 0$ is monotonically increasing in μ, then this expression always holds. It can be verified that $b + (1-\mu)\,\frac{1-2b}{(1-z)(1-b)}\Big) > (1-\mu)\lambda^*$, since in the worst case, that is, when $\mu = \frac{\lambda^* - 1/2}{\lambda^*}$, it holds. Therefore, it is shown that to offer $b = 0$ is better than to offer $b^* \leq b < 1/2$,when $\mu > \frac{\lambda^* - 1/2}{\lambda^*}$.

Next, we have to prove that to offer $b = 1/2$ is better than to offer $b^* \leq b < 1/2$,when $\mu \leq \frac{\lambda^* - 1/2}{\lambda^*}$. Then, we have to prove that: $1/2 \geq \mu[(1-b)] + (1-\mu)[(1-b) - \frac{1-2b}{(1-z)(1-b)}(1-b)]$.

If we are in the worst case, that is $\mu = \frac{\lambda^* - 1/2}{\lambda^*}$ and $b = 1/2$, this expression holds. Therefore, it is shown that to offer $b = 1/2$ is better than to offer $b^* \leq b < 1/2$, when $\mu \leq \frac{\lambda^* - 1/2}{\lambda^*}$.

Appendix B.

Proof of Lemma 4

FCE Equilibrium.

Given that $x < \hat{x} = \frac{\lambda^* - 1/2}{\lambda^*}$ and that both types of investors invest, there is no updating of beliefs and the best option for both type of allocators is the reward $b = 1/2$. Every player gets a payoff of $1/2$ without any punishment. No type of investor has incentives to deviate since $1/2 > 0$.

Proof of Lemma 5

QCE Equilibrium.

The equilibrium utility of the selfish investor is $(1-y)/2$ and for the selfish allocator is $x + (1-x)(1-\lambda^*)$.The payoff for the fair-minded allocator is $\frac{1}{2}$ and for the reciprocal investor is $(1-y)/2 + y(-z\lambda^* - \alpha(1-\lambda^* + z\lambda^*)$, while his material payoff is $(1-y)/2 + y(-z\lambda^*)$).

Given that x_t is such that $max\{0, \hat{x}\} \leq x_t \leq 1$ and that both types of investors invest I, as there is no updating of beliefs, now the selfish allocator will offer $b = 0$ by Lemma 3. The fair-minded allocator will set $b = 1/2$. On the one hand, the reciprocal investor chooses to invest because $(1-y)/2 + y(-z\lambda^* - \alpha[1-\lambda^* + z\lambda^*]) \geq 0$, that is, $y \leq \hat{y}$. On the other hand, the selfish investor will also choose to invest because $(1-y)/2 > 0$.

Proof of Lemma 6

ISE Equilibrium.

The equilibrium payoff of the selfish investor is $(1-y)/2$, and 0 for the reciprocal investor. The payoff for the selfish allocator is x and, $x/2$ for the fair-minded allocator.

Suppose that the selfish investor chooses to invest and the reciprocal investor chooses not to invest, then the updated beliefs of the allocators are $\mu(s/I) = 1$. Therefore, the selfish allocator will set a reward $b = 0$ and the fair-minded allocator will set $b = 1/2$.

Incentive compatibility for the selfish player implies that $(1-y)/2 \geq 0$, that is, $y \leq 1$.

On the other hand, incentive compatibility for the reciprocal investor implies that.

$(1-y)/2 + y(-z\lambda^* - \alpha[1-\lambda^* + z\lambda^*])$ has to be smaller than 0, and this is achieved if

$y \geq \hat{y}$. Note that there exist separating equilibria because $\hat{y} \leq 1$.

Non Cooperative Equilibrium.

Lemma. *For any x_t and for $y_t = 1$, there is a Non-Cooperative Equilibrium in which both types of investor choose not to invest.*[18,19]

Proof.

The out-of-equilibrium beliefs are $\mu(s/I) = 1$. Therefore, the strategy of the selfish allocator is to set a reward $b = 0$ and for the fair-minded allocator $b = 1/2$.The selfish investor does not deviate to invest since $y = 1$. Reciprocal investors does not invest either if $y \geq \hat{y}$, therefore as $1 \geq \hat{y}$, this investor does not deviate from not investing.

Appendix C.

Transition probabilities

P_j^{ik} denote the probability that a child of a parent of a social group j with preferences i is socialized to preferences k. The socialization mechanism in group j is then characterized by the following transition probabilities:

$P_{jt}^{ss} = \tau_{jt}^s + (1-\tau_{jt}^s)\nu_t P_{jt}^{sr} = (1-\tau_{jt}^s)(1-\nu_t) P_{jt}^{rr} = \tau_{jt}^r + (1-\tau_{jt}^r)(1-\nu_t) P_{jt}^{rs} = (1-\tau_{jt}^r)\nu_t$

Given these transition probabilities it is easy to characterize the dynamic behaviour of ν_t:

$\nu_{t+1} = [\nu_t P_{jt}^{ss} + (1-\nu_t) P_{jt}^{rs}]$.

Substituting we obtain the following equation on differences:

$\nu_{t+1} = \nu_t + \nu_t(1-\nu_t)[\tau_{jt}^s - \tau_{jt}^r]$.

Computing the levels of cultural intolerance and the socialization efforts

Levels of cultural intolerance and socialization efforts in the FCE.

Notice that $V_j^{ik} = 1/2$ for all types of parents. Therefore, the net gains for any type of parent of any social group obtained from transmitting their own preferences ΔV_j^{ik}, that is, their levels of cultural intolerance, are zero. But this in turn implies that there are not incentives at all for socialization for any parent. Therefore, all the optimal education effort functions are zero and consequently, the distribution of preferences in both populations will remain unchanged, that is $x_{t+1} = x_t$ and $y_{t+1} = y_t$.

Proof of Lemma 8

[18] The allocators' beliefs in this equilibrium are such that if they observe investment they will infer that this action is more likely to come from a selfish investor.

[19] If we assume that the surplus of investing is H, instead of normalizing the surplus to 1, and the surplus of not investing is $2l$, where $H > 2l$, then this equilibrium would exist for $y_t \in [\tilde{y}, 1]$, where $\tilde{y} = \frac{H-2l}{H}$.

Levels of cultural intolerance and socialization efforts in QCE.

The equilibrium payoff of the selfish investor is $(1-y)/2$, for the reciprocal investor is $(1-y)/2+y(-z\lambda^*-\alpha(1-\lambda^*+z\lambda^*))$, for the selfish allocator is $x+(1-x)(1-\lambda^*)$ and for the fair-minded allocator is $\frac{1}{2}$.

We start computing the levels of cultural intolerance of the selfish investors parents:

$V_{1s}^{ss}=(1-y)/2$ $V_{1s}^{sr}=(1-y)/2-yz\lambda^*$.

Therefore, $\Delta V_{1s}^{s}(y)=yz\lambda^*>0$.

The same reasoning applies for reciprocal investors:

$V_{1r}^{rr}=(1-y)/2+y(-z\lambda^*-\alpha(1-\lambda^*+z\lambda^*))$,

$V_{1r}^{rs}=(1-y)/2\ -y\alpha$.

Therefore, $\Delta V_{1r}^{r}(y)=y\lambda^*(\alpha-z(1+\alpha))\geq 0$.

Then, we can compute: $\tau_1^s(x_t,y_t)=\ q\cdot\Delta V_{1s}^{s}(y)\cdot(1-x_t)=\ qy_tz\lambda^*(1-x_t)$.

And, $\tau_1^r(x_t,y_t)=\ q\cdot\Delta V_{1r}^{r}(y)\cdot x_t=\ qy_t\lambda^*(\alpha-z(1+\alpha))x_t$.

Now, we proceed to analyse the socialization decision of the selfish allocators parents. In this case:

$V_{2s}^{ss}=x+(1-x)(1-\lambda^*)$.

$V_{2s}^{sr}=1/2$.

Therefore, $\Delta V_{2s}^{s}(x)=1/2-(1-x)\lambda^*\geq 0$.

The same computation can be done for fair-minded allocators:

$V_{2r}^{rr}=1/2$.

$V_{2r}^{rs}=x(1-\beta)+(1-x)(1-\lambda^*-\beta(1-\lambda^*+z\lambda^*))$

Therefore, $\Delta V_{2r}^{r}(x)=1/2-[x(1-\beta)+(1-x)(1-\lambda^*-\beta(1-\lambda^*+z\lambda^*)]\geq 0$.

Thus, we can obtain: $\tau_2^s(x_t,y_t)=\ q\cdot\Delta V_{2s}^{s}(x)\cdot(1-y_t)=\ q[1/2-(1-x_t)\lambda^*](1-y_t)$.

And. $\tau_2^r(x_t,y_t)=\ q\cdot\Delta V_{2r}^{r}(x)\cdot y_t=q[1/2-[x_t(1-\beta)+(1-x_t)(1-\lambda^*-\beta(1-\lambda^*+z\lambda^*)]]y_t$

Notice that the optimal socialization effort functions also depend on both preferences distributions (x,y).

Demarcation curves in the QCE region. The x-demarcation curve is the geometric locus of all pairs (x_t,y_t) under which x_t is in a steady-state, that is, $x_{t+1}-x_t=0$. Along this curve the distribution of preferences in the investors population remains constant over time.

As $\Delta V_1^i(y)>0$ and $\Delta V_2^i(x)\ \geq 0$ for $i=s,r$, it follows from the system (1), that $x_{t+1}-x_t\gtreqless 0$ iff $\tau_1^s(x_t,y_t)\gtreqless\tau_1^r(x_t,y_t)$.

Thus, the $x-$demarcation curve is defined by $\tau_1^s(x_t,y_t)=\tau_1^r(x_t,y_t)$.[20]

The properties of the x-demarcation curve in the QCE regions are:

$$\Delta x=x_{t+1}-x_t\begin{cases}>0,\ \text{if } x_t<x'\\ =0,\ \text{if } x_t=x'\\ <0,\ \text{if } x_t>x'\end{cases}\tag{2}$$

where $x'=\frac{z}{\alpha(1-z)}$ which is independent of y_t.

Similarly, the $y-$demarcation curve is the geometric locus of all pairs (x_t,y_t) under which y_t is in a steady-state, that is, $y_{t+1}-y_t=0$. Along this curve the distribution of preferences in the allocators population remains constant over time.

As $\Delta V_1^i(y)>0$ and $\Delta V_2^i(x)\ \geq 0$ for $i=s,r$, it follows from the system (1), $y_{t+1}-y_t\gtreqless 0$ iff $\tau_2^s(x_t,y_t)\gtreqless\tau_2^r(x_t,y_t)$.

Thus, the $y-$demarcation curve is defined by $\tau_2^s(x_t,y_t)=\tau_2^r(x_t,y_t)$.

The properties of the y-demarcation curve

$$\Delta y=y_{t+1}-y_t\begin{cases}>0,\ \text{if } y_t<y_t(x_t)\\ =0,\ \text{if } y_t=y_t(x_t)\\ <0,\ \text{if } y_t>y_t(x_t)\end{cases}\tag{3}$$

where $y_t(x_t)=\frac{1/2-(1-x_t)\lambda^*}{\beta(1-\lambda^*(1-x_t)(1-z))}$ which is increasing in x_t.

Proof of Lemma 9

Levels of cultural intolerance and socialization efforts in ISE.

In this equilibrium the selfish investor chooses to invest (I) and the reciprocal investor chooses not to invest (N). The selfish allocator offers $b=0$ and the fair-minded allocator offers $b=1/2$. The equilibrium payoff of the selfish investor is $(1-y)/2$, for the reciprocal investor is 0, for the selfish allocator is x and for the fair-minded allocator is $x/2$.

[20] Obviously, $x=0$ and $x=1$ are also steady states but unstable.

We start computing the levels of cultural intolerance of the selfish investors parents:
$V_{1s}^{ss} = (1-y)/2$ $V_{1s}^{sr} = 0$.
Therefore $\Delta V_{1s}^{s} = (1-y)/2 > 0$.
For to reciprocal investors:
$V_{1r}^{rr} = 0$.
$V_{1r}^{rs} = (1-y)/2 \ -y\alpha$.
Therefore $\Delta V_{1r}^{r} = 0 - [(1-y)/2 \ -y\alpha] = y\alpha - (1-y)/2 > 0$.

Note that to compute V_j^{ik} we assume that a parent of type i evaluates his child's well-being using his own utility function. For example, V^s is the utility to an reciprocal player if his child is selfish. This child will not punish the allocator when he is offered $b = 0$.

We can now obtain the optimal education effort function for both types of investors:
$\tau_1^s(x_t, y_t) = \ q \cdot \Delta V_{1s}^{s}(x_t) \cdot (1-x_t) = \ q(1-y_t)/2)\,(1-x_t)$.
$\tau_1^r(x_t, y_t) = \ q \cdot \Delta V_{1r}^{r}(x_t) \cdot x_t = q(y_t\alpha - (1-y_t)/2)x_t$.
Now we proceed to analyse the socialization decision of the selfish allocators parents. In this case:
$V_{2s}^{ss} = x$.
$V_{2s}^{sr} = x/2$.
Therefore $\Delta V_{2s}^{s}(x) = x/2 > 0$.
The same for fair-minded allocators:
$V_{2r}^{rr} = x/2$.
$V_{2r}^{rs} = x(1-\beta)$.
Therefore, $\Delta V_{2r}^{r}(x) = x/2 - \ x(1-\beta) = x(\beta - 1/2) > 0$.
The same analysis as before gives: $\tau_2^s(x_t, y_t) = \ q \cdot \Delta V_{2s}^{s}(x_t) \cdot (1-y_t) = \ qx_t\frac{1}{2}(1-y_t)$.
And, $\tau_2^r(x_t, y_t) = \ q \cdot \Delta V_{2r}^{r}(x_t) \cdot y_t = \ qx_t(\beta - 1/2)y_t$.

Demarcation curves in the ISE region.

The $x-$demarcation curve is defined by $\tau_1^s(x_t, y_t) = \tau_1^r(x_t, y_t)$.

The properties of the x-demarcation curve are:

$$\Delta x = x_{t+1} - x_t \begin{cases} > 0, \text{ if } x_t < x_t(y_t) \\ = 0, \text{ if } x_t = x_t(y_t) \\ < 0, \text{ if } x_t > x_t(y_t) \end{cases}$$

where $x_t(y_t) = \frac{1-y_t}{2\alpha y_t}$ which is decreasing with y_t.

Similarly, the $y-$demarcation curve is the geometric locus of all pairs (x_t, y_t) under which y_t is in a steady-state, that is, $y_{t+1} - y_t = 0$.

Thus, the $y-$demarcation curve is defined by $\tau_2^s(x_t, y_t) = \tau_2^r(x_t, y_t)$.

The properties of the y-demarcation curve are

$$\Delta y = y_{t+1} - y_t \begin{cases} > 0, \text{ if } y_t < \overline{y} \\ = 0, \text{ if } y_t = \overline{y} \\ < 0, \text{ if } y_t > \overline{y} \end{cases}$$

where $\overline{y} = \frac{1}{2\beta}$ which is independent of x_t.

Appendix D.

Proof of the stability of the steady states of the dynamic system

We have a non-linear difference equation system. In order to check the stability of the system we can use a linear approximation to this system. As long as we analyse a small neighbourhood of the fixed points, the linear approximation can give us the same equilibrium as the original system, therefore the linear approximation (the local stability analysis) could serve as a supplement to the phase-diagram analysis.

The local stability or instability of the equilibrium can be deduced from the behaviour of the matrix of partial derivatives-the Jacobian matrix of the nonlinear system-evaluated at the equilibrium.

We will denote the Jacobian evaluated at the equilibrium $(\overline{x}, \overline{y})$ by J_E and its elements by a, b, c, d:

$$J_E = \begin{bmatrix} \frac{\partial \dot{x}_t}{\partial x} & \frac{\partial \dot{x}_t}{\partial y} \\ \frac{\partial \dot{y}_t}{\partial x} & \frac{\partial \dot{y}_t}{\partial y} \end{bmatrix}_{(\overline{x},\overline{y})} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

To check the stability of the dynamical systems we have to verify that the trace of J_E is negative, that is, $a + d < 0$ and that the determinant of J_E is positive, that is, $a \cdot d - b \cdot c > 0$, evaluated in the fixed points.

Proof of the stability of the dynamic system in ISE region

Taking the continuous time limit, the dynamics in the ISE region is governed by the equations:

$\dot{x}_t = [x_t(1-x_t)(1-y_t)(\frac{1}{2} - \alpha y_t x_t)]\ \dot{y}_t = x_t[y_t(1-y_t)(\frac{1}{2} - \beta y_t)]$

Therefore the Jacobian matrix J_E would be.

$$\begin{bmatrix} x_t y_t - \frac{1}{2}y_t - x_t + 3\alpha x_t^2 y_t - 2\alpha x_t y_t + \frac{1}{2} & \frac{1}{2}x_t^2 - \frac{1}{2}x_t - \alpha x_t^2 + \alpha x_t^3 \\ \frac{1}{2}y_t - \frac{1}{2}y_t^2 - \beta y_t^2 + \beta y_t^3 & \frac{1}{2}x_t - x_t y_t + 3\beta x_t y_t^2 - 2\beta x_t y_t \end{bmatrix}$$

If we substitute the value of the fixed points $\overline{x} = \frac{\beta - 1/2}{\alpha}$ and $\overline{y} = 1/2\beta$ in this matrix, we obtain

$$\begin{bmatrix} -\frac{1}{8\alpha\beta}(2\beta-1)(2\alpha-2\beta+1) & -\frac{1}{4\alpha^2}\beta(2\beta-1)(2\alpha-2\beta+1) \\ 0 & -\frac{1}{8\alpha\beta}(2\beta-1)^2 \end{bmatrix}$$

Therefore the trace of the Jacobian evaluated in the equilibrium $(a + d)$, after simplifying and rearranging terms is, $-\frac{1}{4\beta}(2\beta - 1)$, this expression is negative if $\beta > 0.5$, therefore the trace is negative.

On the other hand, computing the determinant of the Jacobian matrix, $a \cdot d - b \cdot c$ yields the result: $\frac{1}{64\alpha^2\beta^2}(2\beta-1)^3(2\alpha - 2\beta + 1)$, an expression that is positive, since $\alpha > \beta$. Therefore, the dynamical system is locally stable.

Proof of the stability of the dynamic system in QCE region

Taking the continuous time limit, the dynamics in the QCE region is governed by the equations:

$\dot{x}_t = x_t(1-x_t)(y_t z\lambda^*(1-x_t) - y_t\lambda^*(\alpha - z(1+\alpha))x_t)\ \dot{y}_t = y_t(1-y_t)$
$((1/2 - (1-x_t)\lambda^*)(1-y_t) - ((1/2 - (x_t(1-\beta) + (1-x_t)(1-\lambda^* - \beta(1-\lambda^* + z\lambda^*)))y_t)\ \big]$.

Therefore the Jacobian matrix J_E would be.

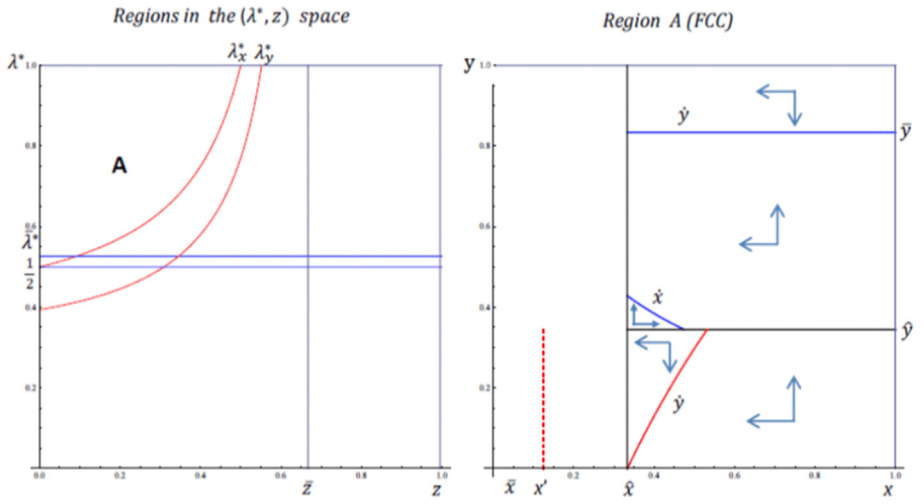
$$\begin{bmatrix} y_t\lambda^*\begin{pmatrix} z - 2zx_t - 2\alpha x_t + \\ 3\alpha x_t^2 + 2z\alpha x_t - 3z\alpha x_t^2 \end{pmatrix} & -x_t\lambda^*(x_t - 1)(z - \alpha x_t + z\alpha x_t) \\ -y_t\lambda^*(y_t - 1)(z\beta y_t - \beta y_t + 1) & \begin{array}{c} y_t - \lambda^* - \frac{3}{2}y_t^2 - 2\beta y_t + x_t\lambda^* + 2y_t\lambda^* + 3\beta y_t^2 + \\ 2\beta y_t\lambda^* - 2x_t y_t\lambda^* - 3\beta y_t^2\lambda^* - 2z\beta y_t\lambda^* - \\ 2\beta x_t y_t\lambda^* + 3z\beta y_t^2\lambda^* + 3\beta x_t y_t^2\lambda^* + \\ 2z\beta x_t y_t\lambda^* - 3z\beta x_t y_t^2\lambda^* \end{array} \end{bmatrix}$$

If we substitute the value of the fixed points $x' = \frac{z}{\alpha(1-z)}$ and $y' = \frac{1/2 - (\frac{\alpha(1-z)-z}{\alpha(1-z)})\lambda^*}{\beta(1 - \frac{\lambda^*(\alpha(1-z)-z)}{\alpha})}$ in this matrix, we obtain similar conclusions to the previous section that is, the trace of the Jacobian evaluated in the equilibrium is negative and the determinant of the Jacobian matrix is positive, thus the dynamical system is locally stable.

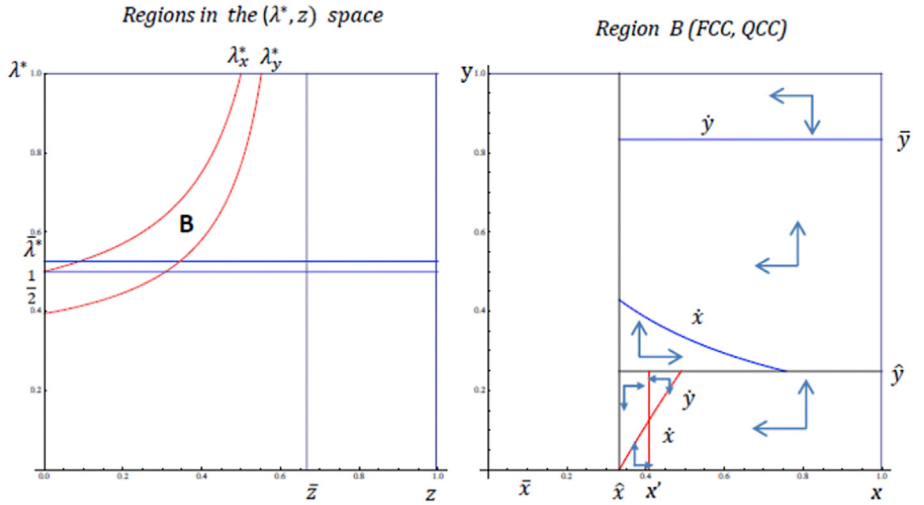
Appendix E.

Phase Diagram Analysis

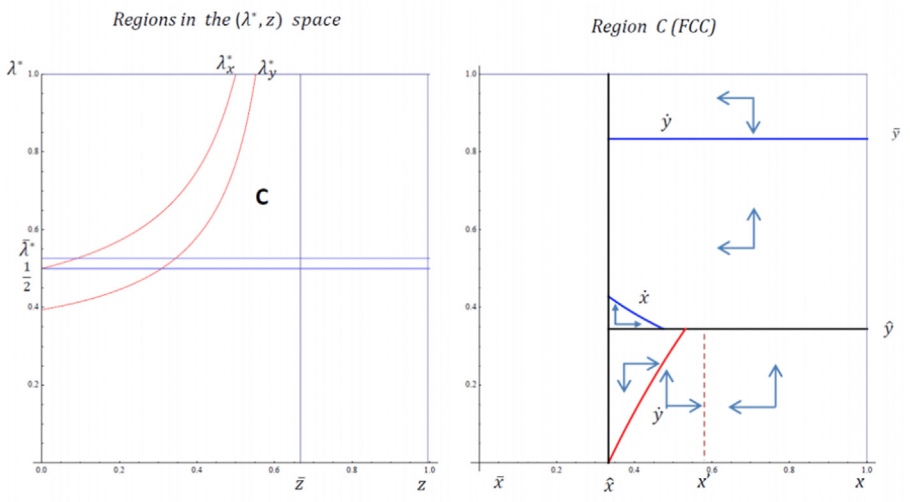
Region A is characterized by having $\hat{x}(\lambda^*) \geq x'(z)$, $y'(z, \lambda^*, \alpha, \beta) < \hat{y}(z, \lambda^*, \alpha)$ and $\overline{x}(\alpha, \beta) < \hat{x}(\lambda^*)$.



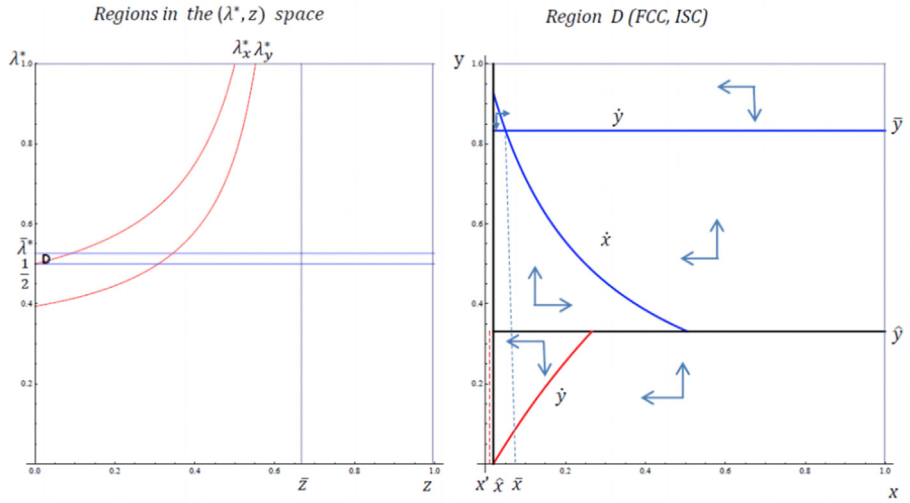
Region B is characterized by having $\hat{x}(\lambda^*) < x'(z)$, $y'(z, \lambda^*, \alpha, \beta) < \hat{y}(z, \lambda^*, \alpha)$ *and* $\overline{x}(\alpha, \beta) < \hat{x}(\lambda^*)$.



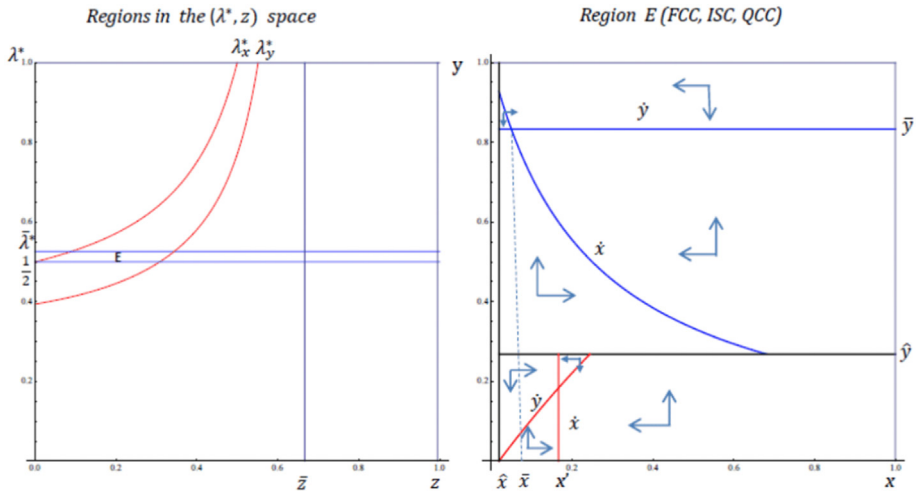
Region C is characterized by having $\hat{x}(\lambda^*) < x'(z)$, $y'(z, \lambda^*, \alpha, \beta) > \hat{y}(z, \lambda^*, \alpha)$ and $\overline{x}(\alpha, \beta) < \hat{x}(\lambda^*)$.



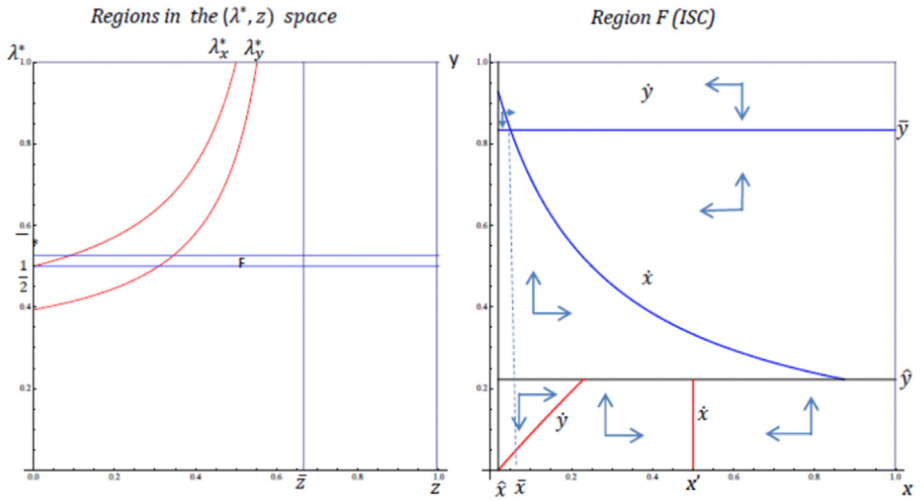
Region D is characterized by having $\hat{x}(\lambda^*) > x'(z)$, $y'(z, \lambda^*, \alpha, \beta) < \hat{y}(z, \lambda^*, \alpha)$ and $\overline{x}(\alpha, \beta) > \hat{x}(\lambda^*)$.



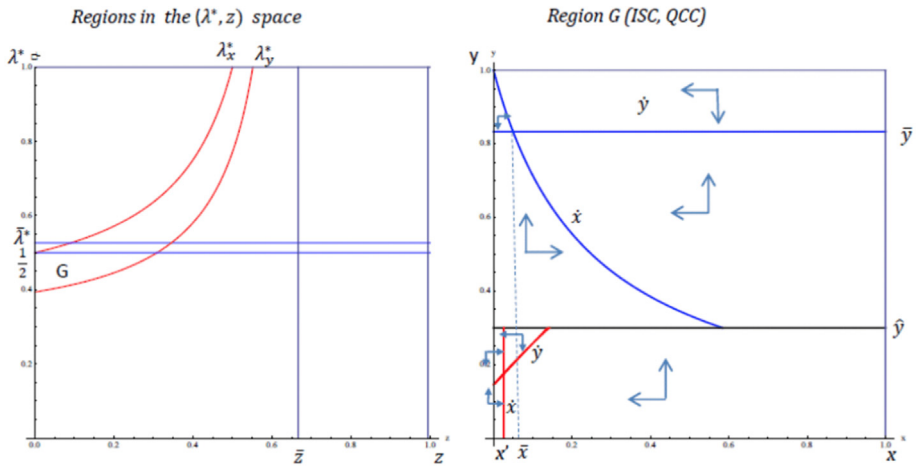
Region E is characterized by having $\hat{x}(\lambda^*) < x'(z)$, $y'(z, \lambda^*, \alpha, \beta) < \hat{y}(z, \lambda^*, \alpha)$ and $\overline{x}(\alpha, \beta) > \hat{x}(\lambda^*)$.



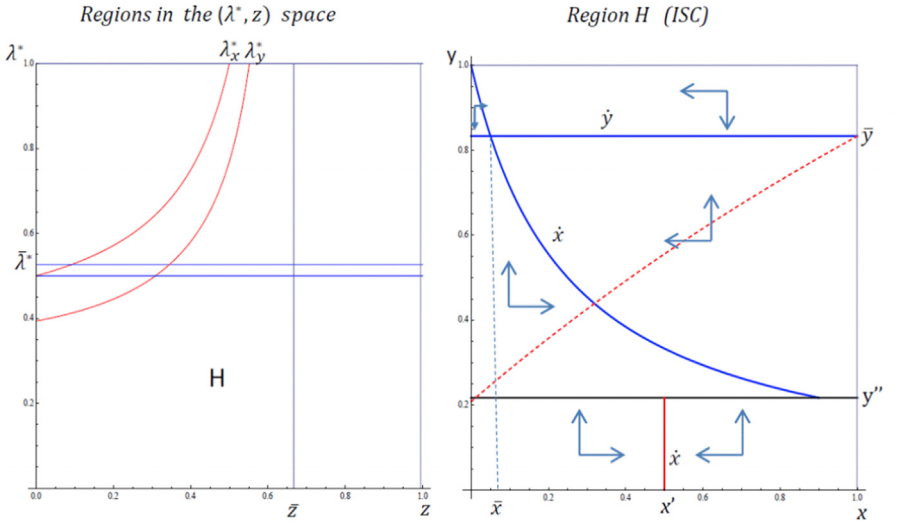
Region F is characterized by having $\hat{x}(\lambda^*) < x'(z)$, $y'(z, \lambda^*, \alpha, \beta) > \hat{y}(z, \lambda^*, \alpha)$ and $\overline{x}(\alpha, \beta) > \hat{x}(\lambda^*)$.



Region G is characterized by having $y'(z, \lambda^*, \alpha, \beta) < \hat{y}(z, \lambda^*, \alpha)$, and $\hat{x}(\lambda^*) < 0$.



Region H is characterized by having $y'(z, \lambda^*, \alpha, \beta) > \hat{y}(z, \lambda^*, \alpha)$ and $\hat{x}(\lambda^*) < 0$.



Appendix F

Characterization of the functions: $\overline{\lambda}(\alpha, \beta)$,$\overline{\lambda}_y(z; \alpha, \beta)$, $\lambda_x(z; \alpha)$ *and* $\lambda_y(z; \alpha, \beta)$.

Note that $\overline{\lambda}(\alpha, \beta)$ is obtained from equating $\overline{x}(\alpha, \beta) = \frac{\beta - 1/2}{\alpha}$ and $\hat{x}(\lambda^*) = \frac{\lambda^* - 1/2}{\lambda^*}$. Therefore, we get $\overline{\lambda}(\alpha, \beta) = \frac{\alpha}{2\alpha - 2\beta + 1}$, which depends on α and β.

$\overline{\lambda}(\alpha, \beta)$ is decreasing with respect to α (if $\beta > 0.5$) and increasing with β.

The function $\overline{\lambda}_y(z; \alpha, \beta)$, is obtained by equating $\overline{y}(\beta) = \hat{y}(z, \lambda^*, \alpha)$. Therefore, we get $\overline{\lambda}^*_y(z; \alpha, \beta) = \frac{2\alpha - 2\beta + 1}{2\alpha - 2z(\alpha + 1)}$, which is increasing in z.

When $\alpha \geq 1$ and $\beta > 0.5$, $\overline{\lambda}_y(z; \alpha, \beta)$ is located above $\overline{\lambda}(\alpha, \beta)$ for most of the combination of pairs (λ^*, z). Only for values of α and β very close to one and for a very small set of pairs (λ^*, z) such that $\lambda < 1/2$ and a very small z, the above condition does not hold,and ISC is not longer an attractor.

The function $\lambda^*_x(z; \alpha)$ is obtained from equating $x'(\alpha, z) = \frac{z}{\alpha(1 - z)}$ and $\hat{x}(\lambda^*)$. Therefore, we get $\lambda^*_x(z; \alpha) = \frac{\alpha(1 - z)}{2(\alpha(1 - z) - z)}$.

This function is increasing in z and decreasing in α. It takes value of $1/2$ when $z = 0$ and 1 when $z = \frac{\alpha}{2 + \alpha}$. We denote this latter value as z_x. Note that $z_x < \overline{z}$.

The function $\lambda_y^*(z;\alpha,\beta)$ is obtained from equating y' and $\hat{y}$,where $y' = \frac{1/2-(\frac{\alpha(1-z)-z}{\alpha(1-z)})\lambda^*}{\beta(1-\frac{\lambda^*(\alpha(1-z)-z)}{\alpha})}$ and $\hat{y} = \frac{1}{(1+2(z\lambda^*(1+\alpha)+\alpha(1-\lambda^*))}$. This defines the implicit function $\lambda_y^*(z;\alpha,\beta)$.

Or explicitly, $\lambda_y^*(z;\alpha,\beta) = \frac{1}{4(z-\alpha+z\alpha)^2}(\alpha+3\alpha^2+z^2(1+\alpha)(\alpha-\beta)-z(1+2\alpha)(1+2\alpha-\beta)-\alpha\beta-$

$\sqrt[2]{(z-\alpha+z\alpha)^2((1+z)^2\alpha^2+(1+(z-1)\beta)^2+2\alpha(1+z+\beta-z^2\beta)))}$

This function is increasing in z, and takes the value of.

$\frac{1}{4\alpha^2}\left(\alpha+3\alpha^2-\alpha\beta-\sqrt{\alpha^2\left(\alpha^2+(\beta-1)^2+2\alpha(\beta+1)\right)}\right) < (1/2)$ when $z=0$.

This function takes value 1 when $z = \frac{1}{4(\alpha+1)(\alpha+\beta+2)}(\alpha+2\beta+2\alpha^2+2\alpha\beta+$

$\sqrt{12\alpha-8\beta+17\alpha^2+12\alpha^3+4\beta^2+4\alpha^4+4\alpha^2\beta^2+4\alpha\beta+8\alpha\beta^2+20\alpha^2\beta+8\alpha^3\beta+4}-2)$.

We denote this value as z_y. It can be checked that $z_x < z_y < \overline{z}$.

Notice that for a given z, $\lambda_x^*(z;\alpha)$ is always greater than $\lambda_y^*(z;\alpha,\beta)$, so the region between them is not empty. It can be checked that when $z=0$, $\lambda_x^*(z=0;\alpha) > \lambda_y^*(z=0;\alpha,\beta)$ and also when $z=\frac{\alpha}{2+\alpha}$, $\lambda_x^*(z;\alpha) > \lambda_y^*(z;\alpha,\beta)$. Thus, as both functions are monotonically increasing in the range $z\in[0,\frac{\alpha}{2+\alpha}]$, the condition $\lambda_x^*(z;\alpha) > \lambda_y^*(z;\alpha,\beta)$ holds for all z. Therefore, the area between both functions (where QCC is a local attractor) is not empty.

Appendix G.

Proofs of Lemmata 11 and 12

$\lambda_x^*(z;\alpha)$ is decreasing with α and $\lambda_y^*(z;\alpha,\beta)$ is decreasing with α and with β.

Therefore, if there is an increase in α, both $\lambda_x^*(z;\alpha)$ and $\lambda_y^*(z;\alpha,\beta)$ decrease. However, the magnitude of the decrease of $\lambda_x^*(z;\alpha)$ is greater than that of $\lambda_y^*(z;\alpha,\beta)$, so the area between them diminishes, leading to a reduction in the region in which QCC is an attractor. Therefore, the areas in which FCC or ISC are global attractors increase.

However, when β increases, there is only a decrease in $\lambda_y^*(z;\alpha,\beta)$. Thus, the region in which QCC is an attractor expands and this reduces the area of the regions in which FCC or ISC are global attractors.

References

Acemoglu, D., Johnson, S., Robinson, P., 2006. Institutions as a fundamental cause of long-run growth. Amsterdam and Boston. In: Aghion, Phillipe, Durlauf, Steven (Eds.), Handbook of Economic Growth, Volume IA, Elsevier, North Holland, pp. 385–472.

Alesina, A., Fuchs-Schündeln, N., 2007. Goodbye lenin (or not?): the effect of communism on people. Am. Econ. Rev. 97 (4), 1507–1528.

Alesina, A., Giuliano, P., 2015. Culture and institutions. J. Econ. Lit. 53 (4), 898–944.

Algan, Y., Cahuc, P., 2009. Civic virtue and labor market institutions. Am. Econ. J. Macroecon. 1 (1), 111–145.

Algan, Y., Cahuc, P., 2014. Trust, growth, and well-being: new evidence and policy implications. Amsterdam and Boston. In: Aghion, Phillipe, Durlauf, Steven (Eds.), Handbook of Economic Growth, ume 2. Elsevier, North Holland, pp. 49–120.

Arrow, K., 1972. Gifts and exchanges. Philos. Publ. Aff. 1 (4), 343–362.

Bisin, A., Verdier, T., 1998. On the cultural transmission of preferences for social status. J. Publ. Econ. 70, 75–97.

Bisin, A., Verdier, T., 2000. Beyond the melting pot: cultural transmission, marriage, and the evolution of ethnic and religious traits. Q. J. Econ. 115, 955–988.

Bisin, A., Verdier, T., 2001. The economics of cultural transmission and the dynamics of preferences. J. Econ. Theor. 97 (2), 298–319.

Bisin, A., Verdier, T., 2017. On the Joint Evolution of Culture and Institutions. NBER Working Papers 23375. National Bureau of Economic Research, Inc.

Boyd, R., Richerson, P., 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.

Cavalli-Sforza, L., Feldman, M., 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton NJ.

Coleman, J., 1990. Foundations of Social Theory. Harvard University Press.

Davis, J., Mazumber, B., 2017. The Decline in the Intergenerational Mobility after 1980. Working Paper Series WP-2017-5. Federal Reserve Bank of Chicago.

Falk, A., Becker, A., Dohmen, T., Enke, B., Hoffman, D., Sunde, W., 2018. Global evidence on economic preferences. Q. J. Econ. 133 (4), 1645–1692.

Fehr, E., Hoff, K., 2011. Introduction: tastes castes and culture:the influence of society on preferences. Econ. J. 121, 396–412.

Fehr, E., Klein, A., Schmidt, K., 2007. Fairness and contract design. Econometrica 75 (1), 121–154.

Fehr, E., Schmidt, K., 1999. A theory of fairness, competition and cooperation. Q. J. Econ. 114, 817–868.

Hauk, E., Saez, M., 2002. On the cultural transmission of corruption. J. Econ. Theor. 107 (2), 311–335.

Herold, F., 2012. Carrot or stick? The evolution of reciprocal preferences in a haystack model. Am. Econ. Rev. 102 (2), 914–940.

Herrmann, B., Thöni, C., Gachter, S., 2008. Antisocial punishment across societies. Science 319 (5868), 1362–1367.

Huck, S., Kosfeld, M., 2007. The dynamics of neighbourhood watch and norm enforcement. Econ. J. 117, 270–286.

Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? A cross-country investigation. Q. J. Econ. 112 (4), 1251–1288.

Olcina, G., Penyarrubia, C., 2004. Hold up and intergenerational transmission of preferences. J. Econ. Behav. Organ. 54, 111–132.

Putnam, R., 1993. Making Democracy Work. Princeton University Press, Princeton, NJ.

Rob, R., Zemsky, P., 2002. Social capital, corporate culture, and incentive intensity. Rand J. Econ. 33 (2), 243–257.

Tabellini, G., 2008a. Institutions and culture. J. Eur. Econ. Assoc. 6 (2–3), 255–294.

Tabellini, G., 2008b. The scope of cooperation: values and incentives. Q. J. Econ. 123 (3), 905–950 MIT Press.

Tabellini, G., 2010. Culture and institutions: economic development in the regions of europe. J. Eur. Econ. Assoc. 8 (4), 677–716.