Cooperation, punishment and immigration

Cooperation, Punishment and Immigration[*]

Paolo Pin[†] Brian W. Rogers[‡]

May 2015

Abstract

We study the incentive to cooperate in a society comprised of citizens and immigrants. The level of cooperation is governed by a steady state under population dynamics, along with the behavior of individual citizens and immigrants. We provide an equilibrium characterization, exhibiting a uniquely determined positive level of cooperation in society. We then use this framework to study the impact of government programs aimed at punishing immigrants who defect. When agents produce offspring, we show that a consequence of such punishment is that, while the incentive for immigrants to defect decreases, there is an equilibrium substitution effect whereby citizens realize an increased incentive to defect.

JEL codes: C73 Stochastic and Dynamic Games; Evolutionary Games; Repeated Games – D85 Network Formation and Analysis: Theory – J61 Geographic Labor Mobility; Immigrant Workers.

[*] An extended abstract of this paper has circulated under the title "Cooperation in a Society with Differential Treatment of Immigrants".

[†] Dipartimento di Economia Politica e Statistica, Università degli Studi di Siena, Italy. Email: paolo.pin@unisi.it.

[‡] Department of Economics, Washington University in St. Louis. Email: brogers@wustl.edu

1 Introduction

This paper explores theoretically the relationships between immigration and the economic behavior of a country's population. There is currently much debate over the nature of the myriad effects and consequences of immigration. We are here especially interested in showing, in the context of a simple model, how policies aimed at immigrants have the potential to influence the outcomes of existing citizens.

There is relatively little research measuring the net economic impacts of immigration.[1] It is clear that there are multiple channels through which immigration affects the economic outcomes of a country, but we will be mainly concerned with one particular dimension – how the incentives of citizens change in response to changes in the level of immigration – and so the conclusions we draw should be interpreted with this limitation in mind. Our framework is not rich enough to speak to the overall effects of immigration policy. Of particular interest to the issue we study is the estimate of Borjas (2003) that, among high school drop outs, the decrease in wage attributable to immigration was 9 percent.[2]

We develop a framework that allows us to analyze immigration and behavior in an equilibrium context. We imagine that each individual makes a binary choice between cooperating and defecting, where payoffs are based on an underlying prisoners' dilemma stage game with one's (endogenously determined) partners. While abstract, this choice is meant to capture an individual's general behavior regarding his participation in society and his interactions with others. For example, one could view these actions as entering the formal economy and generally obeying the laws of the land, or instead entering the black market and conducting activities that are illegal or deemed to be socially costly.

There are two channels through which individuals enter the economy: through birth within the country and through immigration from abroad.[3] Both kinds of agents

[1] Immigration increases labor supply allowing domestic resources that are complementary to labor to be used more efficiently, increasing profits (see Hanson 2007 and Borjas 2003). On the other hand, the increased labor supply pushes wages down, at least in some sectors. Immigration also lowers prices, raising real income (see Cortes 2008).

[2] Immigrants also contribute to the tax base and demand costly services. According to Hanson (2007), the net fiscal effect for the United States appears to be positive for high-skill immigrants, and negative for low-skilled immigrants, at least in the short run, but existing data is not of sufficiently quality to measure this precisely. Dustmann and Frattini (2014), studying the U.K., find generally net positive fiscal effects of immigration, especially among recent immigrants. Dustmann et al. (2013) find, further, that there is a small negative effect on immigration on low wages, but a positive effect on high wages.

[3] Consistent with, e.g., United States law, we assume that all individuals born in the country, whether to citizens or to immigrants, become citizens.

confront exactly the same choice at birth. However, the incentives that they face, and therefore their optimal choices, may be different. The wedge in their incentives is driven by two factors. The first factor is that an agent born within the country inherits the social relationships of its parent. The value of these relationships is endogenous and depends on the (optimal) manner in which relationships are managed in the population. In equilibrium, it is the case that offspring of cooperators have a richer set of relationships than offspring of defectors, and we think of this fact as capturing differential inheritance of social (or "network") capital due to the behavior of the parents.

The second factor is the possibility that the government may expend resources to monitor and punish immigrants who defect. This is, in fact, the policy instrument on which we focus, and we assume that punishment takes the form of expulsion. Our main goal is to analyze the impacts of such a policy. We do not model the costs of enforcing the policy, so one cannot draw welfare conclusions directly from our analysis, but, rather, our goal is to understand the policy's impact on behavior. Our results may be summarized as follows.

First, we characterize the existence of a non-trivial equilibrium, in which either all agents cooperate, or agents mix in such a way that many, but not all, agents cooperate.

The coexistence of cooperative and defective behaviors is descriptive of some systems, at least in a stylized sense. For instance, according to the U.S. Department of Justice, in 2013 there were 23.2 violent crime and 131.4 property crime victimizations per 1000 individuals.[4] That is, criminal activity is certainly present, but it involves a minority of people. One can also think in terms of online commerce, in which there can be an incentive to cheat one's trading partner, but business is still conducted, with the general expectation of honest transactions, despite occasional infractions.[5]

The intuition that guides the equilihbrium characterization is that there is a "natural" level of cooperation controlled by parameters. A key aspect of behavior is that the only way to accumulate relationships over time is through cooperation. If the cooperation level is higher than this natural level, then it becomes tempting to defect, even at the cost of retaining relationships. If the cooperation level is lower, then, as cooperators are relatively scarce, defectors cannot meet enough cooperators to obtain high payoffs; instead it becomes optimal to cooperate so as to gradually accumulate relationships.

Our main result is to characterize the impact of expelling immigrants who defect.

[4] This data is from http://www.bjs.gov/content/pub/pdf/cv13.pdf, accessed October 9, 2014.

[5] For example, the FBI reports 262,813 consumer internet fraud complaints for 2013, out of many millions of transactions. See the Internet Crime Complaint Center, which is run jointly by the FBI and the National White Collar Crime Center, at http://www.ic3.gov/media/annualreport/2013_IC3Report.pdf.

Increasing the intensity of this policy decreases the defection rate, as may be expected. However, there is an equilibrium effect of this policy whereby the incentives for *citizens* to defect increase at the institution of the policy. This arises because as immigrants shift towards cooperation, defecting becomes more tempting for citizens, as the change in immigrants' behavior drives cooperation above the natural level. In our model, offspring born to defectors optimally defect in the presence of the expulsion policy.

A crucial mechanism that generates this effect is that of inheritance. In equilibrium, cooperators maintain valuable relationships. When they die, any offspring inherit this social capital and, as mentioned, this has an important bearing on their incentives and consequently their behavior. Specifically, it arises endogenously that offspring tend to adopt the same behavior as their parents.

Finally, we enrich the model by allowing for heterogeneity in preferences. While such a formulation is certainly more realistic, our main motivation for studying this extension is to argue that the conclusion that expelling immigrants has negative consequences for a fraction of citizens is robust. In particular, this result takes a natural form in that the effect of more intense expulsion on citizens' behavior is smooth, rather than having a discontinuous effect when it is first initiated.

We interpret this latter result in light of the debate on immigration policy alluded to above. It is true in our model that the overall impact of harsher immigration policy is to improve aggregate behavior in the economy. However, this effect is smaller than would be predicted if one failed to account for equilibrium effects. In particular the net impact on the defection rate is diminished by a substitution effect whereby more citizens find it optimal to defect as immigrants shift to cooperating. This result suggests that if one is interested in decreasing the returns to defection, more effective policies target the payoff parameters of the prisoners' dilemma that governs interactions, rather than on the punishment of immigrants. In this sense, while we do not attempt to study optimal policies, we argue that certain classes of policies are unlikely to be optimal in a more general analysis.

The rest of the paper proceeds as follows. Section 2 discusses the academic literature on which we build. Section 3 presents the framework, including our model of population dynamics and our specification of utility functions for agents. Section 4 characterizes equilibria for a baseline case where there is no inheritance and when the government makes no attempt to expel immigrants. Section 5 presents our main results regarding the impact of immigration policy. Section 6 concludes and offers thoughts on how our model and results might be extended to remedy some of the shortcomings of our analysis. Robustness of some assumptions are discussed in Appendix A, while proofs are collected in Appendix B, and a simulation exercise is discussed in Appendix C.

2 Our contribution

The fact that we model the social choices facing agents through a base game of the prisoners' dilemma variety relates our work to the large literature that seeks to explain pro-social behavior through repeated interactions. This question dates at least to Fudenberg and Maskin (1986) who formalized the folk theorem.

Many researchers have by now been motivated by the empirical observation that cooperative behavior is widespread even in situations where punishment schemes are limited.[6] Our work is different from this strand of literature for several important reasons. Principally, we assume that agents make once-per-lifetime decisions about their behavior, which shuts down the possibility of non-stationary punishments. Instead, punishment comes through the threat to sever links, which brings us in touch with studies that focus on voluntary separation.[7] Recent contributions to this line of work, that focus explicitly on the role of separation of partners, are Izquierdo et al. (2010, 2014).[8]

Even though our setting has the features of endogenous termination of relationships and anonymity, our work bears little in common with this literature by virtue of the fact that our interactions take place through an endogenous network in which, importantly, individuals typically manage multiple relationships concurrently, rather than having a single partner at any given moment of time. Fosco and Mengel (2011) study imitation dynamics in an evolving network, showing, as do we, that cooperation and defection coexist. As a result, the central tradeoff for our agents is that cooperation allows for the gradual accumulation of many profitable relationships, whereas defection results in a series of more profitable, yet transient relationships.

The most closely related analysis is Immorlica et al. (2010, 2013), on whose matching model we build.[9] That paper studies equilibrium cooperation in a homogeneous

[6] Dall'Asta et al. (2012) study, in a general network topology, the conditions for clusters of sustained cooperation. When instead connections are not fixed, Kandori (1992) demonstrates that cooperation can be sustained by use of community enforcement strategies. Further, while that construction relies on public histories, Kandori (1992) and Ellison (1994) demonstrate that cooperation is still possible under anonymity if players use contagion strategies, and Vega-Redondo (2006) introduces a local information passing to obtain a similar result.

[7] See Kranton (1996), Ghosh and Ray (1996), Datta (1996), Watson (1999, 2002), and Fujiwara-Greve and Okuno-Fujiwara (2009).

[8] A related idea is that cooperation can be sustained via ostracism, whereby an defector's behavior can spread through her local network, resulting in punishment. See Ali and Miller (2013) and Jackson et al. (2012).

[9] The earlier paper is a short and preliminary version of the working paper. While we add to this framework in several ways, we also make one simplification, which is that link formation is unilateral.

population and so cannot speak to our questions of interest, all of which relate to the difference between immigrants and citizens. Since our model introduces heterogeneity in the population, our steady state derivation is significantly more complex. Equilibrium computations are also complicated by the fact that there are different incentives for different agents that must be accounted for. We also introduce the notion of inheritance which, as it turns out, is essential for uncovering the effects of immigration, and immigration policy, on equilibrium outcomes. The main results of Immorlica et al. (2013) demonstrate existence of equilibrium and characterize some of its basic properties. Here, since our objective is to understand the consequences of punishing immigrants, our main results concern certain comparative statics of equilibrium behaviors in the population (Propositions 3, 4 and 5). In particular, the results on how citizens' behavior changes in response to the punishment of immigrants has no counterpart in Immorlica et al. (2013).

We also contribute to the economic literature investigating the impact of immigration. On this, and for additional references, see the surveys of Borjas (1994) and Hanson (2010), the books Borjas (2008) and Borjas (2014), as well as the references in the Introduction. The consensus emerging from this literature is that immigration, even when it is illegal, has relatively small net impact on the economy, but it is likely to be a positive impact. Nonetheless, it almost certainly has a negative effect on those in the lower socio-economic tier, i.e., those competing for low-skill, low-wage jobs.

This literature, while of clear importance, has not for the most part investigated the effect of illegal immigration on incentives.[10] One exception is Kemnitz and Mayr (2012) which studies, in part, the effect of punishing illegal immigrants on the rate of immigration. In our model, the inflow of immigrants is exogenous, allowing us to focus instead on the effects of punishment on citizens' incentives. Mastrobuoni and Pinotti (2014) estimate a very different model on behavior and immigration in which citizenship and immigration status is linked to the criminal behavior.

Finally our paper makes a connection to the economic literature on identity. See Akerlof and Kranton (2000) for a review of this work. There is a relationship between our model and the idea of cultural transmission in the identity literature, present in all the papers surveyed by Bisin and Verdier (2012). In those models, there is a concern for children's welfare that is imposed on parents, or otherwise there is an exogenous element by which parents care about their children's actions. Through various mechanisms, this results in the transmission of traits and, by extension, culture, from one generation to the next, so that children tend to have similar traits as their parents. In our paper

[10] There is a literature discussing the effects of policy on the level and kinds of immigration where the origin annd destination countries are explicitly modeled; on this see, e.g., Djajić (1987) and Levine (1999).

there is a similar outcome whereby children take similar actions as their parents. In our model this transmission happens endogenously, deriving from the inheritance of relationships, which we think of as social capital or network capital, as discussed in the empirical work of Shenk et al. (2010), from the parent.

3 The framework

We build a model of a nation's evolving population in which agents cooperate or defect, seeking to maximize lifetime exepcted payoffs. We study equilibrium behavior under a steady state of the population dynamics. Studying comparative statics of the steady state equilibrium allows us to analyze the potential consequences of a policy aimed at influencing the incentives of immigrants.

There is a single society, that evolves in discrete time, t, modeled as a continuum mass of individuals $\mathcal{S}_t$. At every moment of time the elements of $\mathcal{S}_t$ are the nodes of a directed network $g_t = (\mathcal{S}_t, \mathcal{K}_t)$. $\mathcal{K}_t \subset \mathcal{S}_t \times \mathcal{S}_t$ are the directed links of this network. A link $(i, j) \in \mathcal{K}_t$, is called an out-link for i and an in-link for j. An agent's out-degree (in-degree) at time t is said to be the number of his out-links (in-links) at time t.

A link $(i, j) \in \mathcal{K}_t$ represents the play of a prisoner's dilemma between i and j at time t, given by

$$\begin{array}{c|cc|cc|}
 & C & & D & \\
\hline
C & 1, & 1 & -b, & 1+a \\
\hline
D & 1+a, & -b & 0, & 0 \\
\hline
\end{array} \; ,$$

with $a, b > 0$ and $a - b < 1$. Notice that the game is fully symmetric. The orientation of the link plays a role only in the evolution of $\mathcal{K}_t$, described below. The evolution of g_t over time depends on several factors, including exogenous stochastic events as well as strategic choices of the individuals.

3.1 Population dynamics: agents

Let us first focus on the evolution of $\mathcal{S}_t$. At each period there is a fixed inflow, of mass η, of agents, referred to as *immigrants*. There is also an endogenous mass of offspring of existing nodes that enter at each period, called *citizens*. Every entering agent chooses at birth to be a cooperator or a defector, which determines his behavior in the prisoners' dilemma interactions. This decision is taken only once and commits an agent to that behavior for the duration of its life.[11] Given that an agent either

[11] Under certain conditions one can show that an agent would never have an incentive to revise his action. This implies that, in this case, the equilibrium we describe survives an extension to the case of no

cooperates for his life or defects for his life, there is no role for punishments in the traditional repeated game sense using non-stationary strategies. Instead, incentives are provided through how links are maintained, as we will describe below.

Agents are partitioned according to how they entered the network and on the choice they make. Specifically, there are three classes, with each node belonging to exactly one class: *Cooperators*, including both citizens and immigrants ($\mathcal{C}_t$), *Immigrant defectors* ($\mathcal{I}_t$) and *Citizen defectors* ($\mathcal{D}_t$). Thus, $\mathcal{S}_t = \mathcal{C}_t \cup \mathcal{I}_t \cup \mathcal{D}_t$. As will be clear below, agents will generally have different incentives depending on whether they enter (i) as an immigrant, (ii) as an offspring of a defector, or (iii) as an offspring of a cooperator. Accordingly, we introduce notation to describe the behavior of entering agents as follows. Let the probability with which an entering agent at time t chooses to cooperate be denoted $p_{I,t}$ for an immigrant, $p_{D,t}$ for an offspring of a (immigrant or citizen) defector, and $p_{C,t}$ for the offspring of a cooperator. Agents choose to be defectors with the complementary probabilities.[12]

Agents exit the system either through death or through expulsion of immigrant defectors. A proportion $(1-\delta)$ of agents die at every period, independently of their behavior and immigration status. A proportion μ of dying agents have a single offspring. A proportion ν of immigrant defectors are expelled at every period.[13] Figure 1 summarizes the dynamics by representing the flows of agents through the system across classes.

3.2 Population dynamics: links

Turning now to the links, $\mathcal{K}_t$ evolves in the following way. First, whenever an agent exits the system, either through death (without an offspring) or expulsion, all links incident to that agent are removed. An offspring, on the other hand, inherits the in– and out–links of its parent. After the prisoners' dilemma are played, every agent unilaterally severs a subset of its in- and out-links of its choice. All links for which neither of the two incident nodes die (including, potentially, the replacement of a parent with its offspring), nor choose to sever the link, survive to the next period.

commitment, in which an agent is free to use an arbitrary non-stationary strategy. On this see Appendix A.

[12] One could, e.g., apply a purification argument through a small amount of heterogeneity in preferences in order to generically obtain strict preferences. See Section 5.3.

[13] The important aspect of punishment is that it is a cost imposed specifically on immigrants. If the punishment were temporary such as, e.g., prison, then the population dynamics would have to be adjusted to account for reentry.

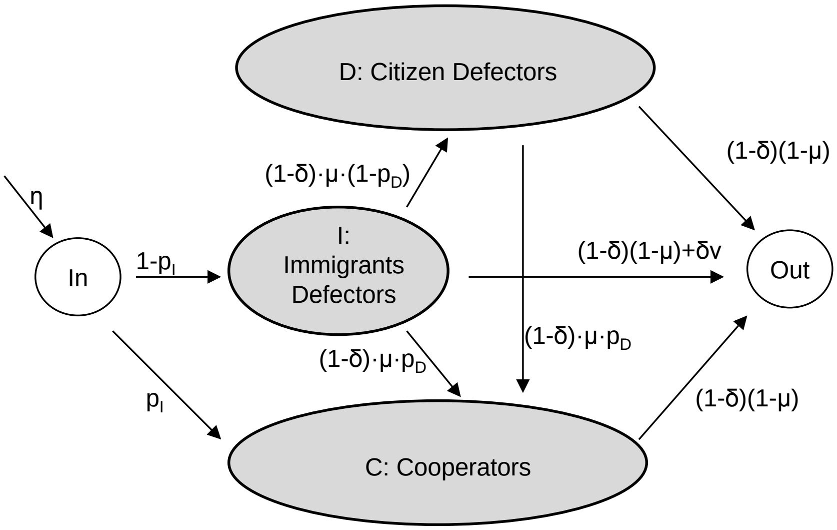


Figure 1: Representation of population dynamics.

We assume that every agent has a budget of k out-links that he maintains.[14] Any agent who, through the loss of links at the previous period, or because it is new in the population, has fewer than k out-links, searches and re-matches the remaining out-links with new partners chosen uniformly at random from the entire population. Notice that out-degrees k are therefore homogeneous, while in-degrees generally vary. What is qualitatively important for our results are that (i) there is a bound on the number of links an agent can expect to maintain, and (ii) it takes time to build up a network of partners, which comes through the fact that in–links are obtained only gradually through the search of other nodes. Similar results could be obtained for specifications with these properties, even if the graph was undirected.

Let us now be more precise about the timing that determines the transition from g_{t-1} to g_t at each period.

(a) The set of offspring (defined from the previous period, see item (g) below) enter and are added to $\mathcal{S}_{t-1}$. Each offspring born to a defector decides with probability

[14] One can imagine that links are costly, and it is the link initiator who pays the cost, with a budget that allows for k out-links to sponsor.

$p_{D,t} \in [0,1]$ to be a cooperator; otherwise it becomes a defector. In the same way each offspring born to a cooperator decides with probability $p_{C,t} \in [0,1]$ to be a cooperatorr; otherwise it becomes a defector.

(b) A set of mass η of new immigrants enter and are added to $\mathcal{S}_{t-1}$. They are ex–ante homogeneous. Each of them decides to cooperate with probability $p_{I,t} \in [0,1]$; otherwise it becomes a defector.

(c) Every entering agent, both immigrants and offspring, casts k out–links to agents in the society. Similarly, all existing agents cast available out–links (k less the number of out–links maintained from the previous period). All partners are chosen uniformly at random.

(d) Payoffs are realized, and actions observed, from the play of the bilateral prisoners' dilemma game, along every link.

(e) Every agent unilaterally severs any of its in– and out–links that it desires.

(f) A proportion $(1-\delta) \in (0,1)$ of $\mathcal{S}_t$ is randomly selected to die, independent of which behavior the agent is taking, and whether the agent is an immigrant or citizen.

(g) Of the agents who die, a proportion $\mu \in [0,1]$ is randomly selected to generate an offspring. The offspring inherits the network position of her parent, i.e., the same set of in– and out–links in $\mathcal{K}_t$. Every offspring is a citizen, whether or not the parent was an immigrant.

(h) Finally, a proportion $\nu \in [0,1]$ of surviving immigrant defectors is randomly selected to be expelled (which is equivalent to death without an offspring).[15]

3.3 Steady state

We analyze a steady state of these dynamics in which the mass of each class is constant, so that $\mathcal{C}_t = \mathcal{C}$, $\mathcal{I}_t = \mathcal{I}$, and $\mathcal{D}_t = \mathcal{D}$. We denote by $q = \frac{|\mathcal{C}|}{|\mathcal{S}|} \in [0,1]$ the corresponding proportion of cooperators in society. Given the focus on steady state, we take the behavior of entering agents to be constant over time, so that $p_{C,t} = p_C$, $p_{I,t} = p_I$, and $p_{D,t} = p_D$. For the main analysis, we shall also set $p_C = 1$, i.e. offspring of cooperators

[15] In principle, we could consider a different model in which the probability ν of being expelled is independent from the probability δ of dying. This alternative assumption would change the probability of passing from $\mathcal{I}$ to *Out* (refer to Figure 1) from $(1-\delta)(1-\mu) + \delta\nu$ to $(1-\delta)(1-\mu)(1-\nu) + \nu$, or equivalently, it would change ν into $(\delta + \mu - \delta\mu)\nu$. In this sense, the alternative assumption simply results in a rescaling of ν that depends on δ and μ.

always choose to cooperate. In Appendix A we show that this is a weak assumption and discuss its robustness.

We discuss below our notion of equilibrium that captures optimal choices in the population. We now focus on the steady state implications that an arbitrary pair (p_I, p_D) has on q.

Consider first the steady-state condition of fixed population size, i.e.,

$$\eta = (1-\delta)(1-\mu)|\mathcal{S}| + |\mathcal{I}|\delta\nu,$$

which balances the inflow to society with the total outflow from society (see Figure 1). Inflow is accounted for exclusively by immigration, since offspring merely replace a dying node, whereas outflow occurs through death in the whole population, as well as expulsion of immigrant defectors. We can write the size of the immigrant defector population as

$$|\mathcal{I}| = \eta(1-p_I)\sum_{t=0}^{\infty}(\delta(1-\nu))^t = \eta\frac{1-p_I}{1-\delta(1-\nu)},$$

which expresses the subpopulation as its per-period inflow multiplied by the expected lifetime of each agent. This allows us to determine $|\mathcal{S}|$ by substituting into the steady-state condition:

$$|\mathcal{S}| = \eta\frac{1-\delta\nu\frac{1-p_I}{1-\delta(1-\nu)}}{(1-\delta)(1-\mu)} = \eta\frac{1-\delta+p_I\delta\nu}{(1-\delta)(1-\mu)(1-\delta+\delta\nu)} \quad .$$

Observe that the size of the society in steady state does not depend on p_D (or on p_C if we allowed it to vary), because those choices are made by citizens, and do not affect their survival probabilities.

We next solve for the masses of $\mathcal{C}$ and $\mathcal{D}$, which are given by the condition

$$\begin{cases} |\mathcal{I}|(1-\delta)\mu(1-p_D) &= |\mathcal{D}|\Big((1-\delta)\mu p_D + (1-\delta)(1-\mu)\Big) \\ \eta\cdot p_I + |\mathcal{I}|(1-\delta)\mu p_D + |\mathcal{D}|(1-\delta)\mu p_D &= |\mathcal{C}|(1-\delta)(1-\mu) \end{cases} .$$

The first line above equates the total inflow to $\mathcal{D}$ with the total outflow from $\mathcal{D}$, while the second line equates total inflow to $\mathcal{C}$ with total outflow from $\mathcal{C}$, where again Figure 1 summarizes the relevant flows. The solution is

$$\begin{cases} |\mathcal{C}| &= \frac{1}{1-\mu}\cdot\frac{1}{1-\mu(1-p_D)}\left(\mu\cdot p_D|\mathcal{I}| + p_I\frac{1-\mu(1-p_D)}{1-\delta}\eta\right) \\ \\ |\mathcal{D}| &= \frac{1}{1-\mu}\cdot\frac{1}{1-\mu(1-p_D)}\left(\mu(1-\mu(1-p_D)-p_D)|\mathcal{I}|\right) \end{cases} .$$

We can now present the steady state proportion of cooperators in terms of exogenous parameters and the cooperation probabilities (p_I, p_D) by simplifying $\frac{|\mathcal{C}|}{|\mathcal{S}|}$ from the above to obtain

$$q = 1 - \frac{(1-\delta)(1-\mu)(1-p_I)}{(1-\mu+\mu p_D)(1-\delta+\delta\nu p_I)} \quad . \tag{1}$$

Equation (1) is one of the main building blocks of our analysis. Let us discuss its implications for how the steady state cooperation level varies with the parameters that describe population dynamics. As a first check it may be noted that, as expected, when $\mu = 1$ and/or $\delta = 1$ (so that $\mathcal{C}$ is absorbing, as a cooperator either lives forever or else necessarily produces a cooperating offspring when dying) we obtain that $q = 1$ (in steady state, all agents cooperate). Next, note that the partial derivatives of q with respect to p_I, p_D and ν are all weakly positive, meaning that the steady state level of cooperation rises as entering nodes cooperate with higher probability (p_I and p_D) or when immigrant defectors are removed more frequently (ν). Notice that p_I and p_D are endogenous, and so will have to be determined from incentives below. Also, when $\mu = 0$ (no offspring) there is no effect of p_D, since without offspring there is no role for the choice of newborns. Finally, when $\nu = 0$ (no expulsion) the effect of δ (death rate) disappears, since in this case agents from each subpopulation die at the same rate, so that while the death rate affects turnover, it does not affect the relative frequencies of types of agents.

3.4 Solution concept

Equation (1) shows how q depends on the endogenous choices p_I and p_D through the steady state conditions. In what follows we discuss how p_I and p_D are determined by maximizing expected payoffs, which in turn depend on q, so that the equilibrium and steady state conditions are interdependent in characterizing a steady state equilibrium.

There are two aspects to agents' strategies: link management and the choice between cooperation and defection. Naturally, how best to manage links depends on the behaviors of other agents, and, on the other hand, expected payoffs from cooperation versus defection depend on the way one expects its links to evolve.

3.4.1 Optimal linking decisions

Recall that an agent observes the behavior of a given partner only after the first round of interaction (see points (c) and (d) in Section 3.2). But then, given that each agent makes a once-per-lifetime choice at birth between being a cooperator or a defector, optimal linking decisions become simple to characterize, since the future play of a given partner is perfectly predictable. In fact, this is one of the main benefits of studying once-per-lifetime behavior, as otherwise optimal linking decisions would potentially be extraordinarily complex.

We introduce notation to describe how an agent manages its relationships with others. Let an agent using behavior X maintain any given link with an agent using behavior Y with probability $\sigma_{XY} \in [0,1]$ in each period, for $X, Y \in \{C, D\}$. That is, such a link is severed by the agent with probability $1 - \sigma_{XY}$.[16]

We will argue that links between cooperators are always maintained, but links involving a defector are always severed and re-matched. To this end consider first the case of $q \in (0,1)$. There is a pathological possibility that cooperators sever links with each other. That is, $\sigma_{CC} = 0$ is a best response to itself, since if other cooperators are severing links, then a given cooperator will lose all such links independently of her behavior. However, we require that agents overcome this basic coordination problem by assuming that they do not use weakly dominated linking strategies.[17] Once $\sigma_{CC} = 0$ is ruled out, it is immediate that cooperators have strict incentives to maintain links with each other, whereas at least one agent involved in every other link has a strict incentive to sever it.

Taking now $q = 1$, a cooperator is indifferent about how to manage his out-links: he can always re-match with another cooperator. We argue that $\sigma_{CC} = 1$ is nevertheless the unique natural behavior in our model. First, if one desires robustness of the linking strategy to small changes in q, then by the argument in the previous paragraph cooperators should maintain links with each other. A separate rationale is that if we incorporated a small search cost in the model, then maintaining a link to a cooperator is strictly better than searching for a new cooperator (thus one may think that cooperators maintaining links is a natural norm).

Finally, for completeness we note that when $q = 0$ there is a further case of indifference: a linked pair of defectors are indifferent about keeping/severing their link. Notice, however, that for any linking strategy, payoffs to defecting are identically zero, and so we choose to maintain the linking strategies described above for simplicity and consistency.

We remark that the fact that the mutual defect stage game payoff is zero is thus not without loss, as it equates the value of a (D, D) link with the value of no link. However, it is straightforward to generalize our results to accommodate a general payoff term from mutual defection, subject to constraints on optimal linking decisions remaining unchanged.

We summarize this discussion with the following:

[16] In general, the linking strategy could depend arbitrarily on the agent's complete history since birth. We use the simpler formulation since optimal linking decisions are, in fact, quite simple in our context.

[17] This assumption is natural and ubiquitous in network formation models, following the seminal work of Jackson and Wolinsky (1996).

Observation **1** (Optimal link severance)**.** *It is essentially without loss of generality to take $\sigma_{XD} = 0$ and $\sigma_{XC} = 1$, for $X \in \{C, D\}$. I.e., links with defectors are always severed and links with cooperators are always maintained. Thus, it is (only) the links involving mutual cooperation that survive across periods.*

3.4.2 Optimal choice of cooperation versus defection

The derivation of payoffs in this subsection shares important elements with the development in Immorlica et al. (2013).

We denote by $\delta_N = \delta + (1 - \delta)\mu$ the turnover rate among the *network* of cooperators. It is this probability with which a given cooperator either survives one more period, or dies but is replaced by an offspring who inherits the same position in the network. Naturally, when there is no inheritance ($\mu = 0$), we have that $\delta_N = \delta$. Let $n_{XY}^{out}(t)$ denote the expected number of out-links from an age t agent using behavior X to agents using behavior Y, for $X, Y \in \{C, D\}$, where the expectation is taken at the time the agent is born.

The expected number of out–links from a cooperator to other cooperators is:

$$\begin{aligned} n_{CC}^{out}(t) &= \delta_N \ n_{CC}^{out}(t-1) + q(k - \delta_N \ n_{CC}^{out}(t-1)) \\ &= kq + (1-q)\delta_N \ n_{CC}^{out}(t-1) \\ &= k \ q \frac{1 - (\delta_N(1-q))^{t+1}}{1 - \delta_N(1-q)} \quad , \end{aligned}$$

where the last equality is solved recursively setting $n_{CC}^{out}(0) = k \ q$. The first equality reflects the fact that out-links to cooperators in a given period consist of surviving maintained out-links from the previous period, as well as those new out-links that happen to connect with a cooperator. Notice that this calculation, as well as the subsequent ones, rely on Observation 1.

Clearly $n_{CD}^{out}(t) = k - n_{CC}^{out}(t)$, $n_{DC}^{out}(t) = kq$ and $n_{DD}^{out}(t) = k(1-q)$.

The proportion of cooperators that have occupied their position in the network for t periods is $s(t) = (1 - \delta_N)\delta_N^t$.[18] Similarly, we denote the age distribution of defectors by $s_D(t)$. We then have that the expected per-period inflow of links from cooperators

[18] By this we mean the number of periods for which the agent has either been alive, or has replaced a dying parent, i.e., from the perspective of other agents, it is the number of periods for which the agent's position in the network has remained occupied.

and defectors to a given agent are, respectively,

$$\begin{aligned} r_C &= \sum_{t=0}^{\infty} q\ s(t)\ (k - \delta_N\ n^{out}_{CC}(t-1)) = k\frac{q(1-\delta_N^2)}{1-\delta_N^2(1-q)} \quad , \\ r_D &= k\sum_{t=0}^{\infty}(1-q)\ s_D(t) = k(1-q) \quad . \end{aligned}$$

In each case, the aggregate mass of search being done by agents of a particular behavior (cooperate or defect) is given by an expectation over age, given the behavior-dependent age distribution, of the mass of agents of that age taking the particular behavior, multiplied by the number of out-links such agents are expected to form in that period, which comes from the above calculations on out-links. Notice that, in the case of cooperators, the calculation is complicated by the fact that expected out-link search varies non-trivially with age, whereas for defectors it is instead constant (and equal to k).

The expected number of in–links from cooperators to an age-t cooperator is then

$$\begin{aligned} n^{in}_{CC}(t) &= \delta_N\ n^{in}_{cc}(t-1) + r_C \\ &= r_C\frac{1-\delta_N^{t+1}}{1-\delta_N} \\ &= k\frac{q(1+\delta_N)}{1-\delta_N^2(1-q)}(1-\delta_N^{t+1}) \end{aligned}$$

where the second line is solved explicitly setting $n^{in}_{CC}(0) = r_C$. The evolution of expected in-links is dictated by the fact that in each period, a cooperator expects r_C new in-links from cooperators, which are added to the surviving in-links from the previous period. Note that, because of death, the number of expected in-links has a finite upper bound.

Clearly $n^{in}_{CD}(t) = n^{in}_{DD}(t) = r_D$ and $n^{in}_{DC}(t) = r_C$.

The expected stage payoff for a player that has age t is

$$\begin{aligned} \pi_C(t) &= 1\cdot(n^{out}_{CC}(t) + n^{in}_{CC}(t)) - b\cdot(n^{out}_{CD}(t) + n^{in}_{CD}(t)) \quad , \\ \pi_D(t) &= (1+a)\cdot(n^{out}_{DC}(t) + n^{in}_{DC}(t)) \quad , \end{aligned}$$

where the above simply sum the stage game payoffs across the expected set of in- and out-links an agent has with cooperators and defectors at age t.

Recall that there are three classes to which a node can belong: Cooperators, Immigrant defectors and Citizen defectors. The agent's choice at birth, along with whether it entered the system as an offspring or an immigrant, determines the agent's expected lifetime payoffs, as follows:[19]

$$
\begin{aligned}
u_C(q) &= \sum_{t=0}^{\infty} \delta^t \pi_C(t) && (2)\\
u_D(q) &= \sum_{t=0}^{\infty} \delta^t \pi_D(t) && (3)\\
u_I(q) &= \sum_{t=0}^{\infty} (\delta(1-\nu))^t \pi_D(t), && (4)
\end{aligned}
$$

where the distinction in the two roles of defecting derives from the fact that immigrants, and only immigrants, face the possibility of being expelled. Notice here that δ plays two roles: it affects population dynamics through the turnover rate independently of preferences, and it also affects preferences for a given population dynamics.

Every entering agent chooses between cooperation and defection so as to maximize its expected lifetime payoff. An immigrant thus compares $u_C(q)$ with $u_I(q)$, while a citizen compares $u_C(q)$ with $u_D(q)$, potentially mixing in the case of indifference.

3.4.3 Steady state equilibrium and stability

We can now be precise and observe that, given exogenous parameters (δ, μ, ν), Equation 1 determines the steady state level of cooperation as a function of the strategic variables (p_I, p_D), while, given preference parameters (a, b, δ), the optimizing behavior of entering agents determines (p_I, p_D) as a function of q through Equations 2, 3 and 4. With this in mind we now define the notion of equiliibrium as follows:

Definition **1** (Equilibrium concept)**.** *An equilibrium is a sixtuple $(p_I, p_D, \{\sigma_{XY}\}_{X,Y\in\{C,D\}})$ such that: (i) $q(p_I, p_D)$ is a steady state; (ii) $(p_I, p_D, \{\sigma_{XY}\}_{X,Y\in\{C,D\}})$ are best responses given q, which requires*

- *$\sigma_{CC} = \sigma_{DC} = 1$ and $\sigma_{CD} = \sigma_{DD} = 0$, from Observation 1,*
- *if $u_C(q) > u_D(q)$ then $p_D = 1$, while if $u_C(q) < u_D(q)$ then $p_D = 0$, from Equations 2 and 3,*
- *if $u_C(q) > u_I(q)$ then $p_I = 1$, while if $u_C(q) < u_I(q)$ then $p_I = 0$, from Equations 2 and 4.*

[19] A new cooperator could enter as the offspring of a cooperator, in which case it inherits the parent's network of cooperators. The calculation below corresponds instead to the case of no inherited network (as pertains to an immigrant or the offspring of a defector), as the former case is taken care of by the requirement that $p_C = 1$.

All of our results concern equilibria in the sense of Definition 1. Thus it should be emphasized that our analysis relies on an assumption that agents use stationary strategies, form links according to a specific set of rules, and that society is in a steady state. All of these requirements mean that, while we identify certain effects that may arise in equilibrium, we cannot predict that those effects necessarily arise in society.

Some equilibria fail a basic stability requirement and, as such, are not compelling solutions to the model. More specifically, applying a refinement proposed by Blonski (1999), we desire a solution with the property that, if the cooperation level q is slightly perturbed, then entering agents, using utility calculations based on the pertubed cooperation level, make optimal decisions that send the cooperation level back towards its original equilibrium value. It is equilibria that are stable in this sense that we seek to characterize, and perform comparative statics on, in Sections 4 and 5.

We make this intuition precise in the following:

Definition **2** (Stable equilibrium). *A stable equilibrium is an equilibrium* $(p_I, p_D, \{\sigma_{XY}\}_{X,Y\in\{C,D\}})$ *with associated steady state* q *such that there exists an* $\epsilon > 0$ *for which the following hold:*

- *If* $q < 1$*, then for every* $q' \in (q, q+\epsilon)$ *and for every* (p'_I, p'_D) *that are (part of) best responses at* q'*, Equation 1 produces* $q(p'_I, p'_D) < q'$*;*
- *If* $q > 0$*, then for every* $q' \in (q-\epsilon, q)$*, the above inequality is reversed.*

A few remarks are in order. Note first that optimal linking decisions, characterized in Observation 1, are not affected by a perturbation of q. That is, even if perturbations of linking decisions were incorporated into Definition 2, an agent's best response would still involve keeping links with coperators and severing links with defectors. For this reason, our results would not change if we augmented the stability notion in this fashion. Next, recall that if $\nu = 0$, $u_D = u_I$ and all defectors have the same expected payoff, whereas if $\nu > 0$ then $u_D > u_I$, so that it is impossible that both citizens and immigrants are indifferent between cooperating and defecting. Recalling that, from Equation 1, q is increasing in both p_I and p_D, this implies that for small enough perturbations around an equilibrium, all agents will have strict preferences between cooperating and defecting. Thus, for small enough ϵ, there is a unique $q(p'_I, p'_D)$ that the stability condition needs to consider.

Finally, this stability refinement is quite mild, in that it accounts for the direction of the best response, ignoring the magnitude of the resulting change in q across periods. In other words, there could in principle exist equilibria that satisfy Definition 2, but such that following a small perturbation, the best response of entering agents will cause a change that "overshoots" the original equilibrium, as could be the case e.g. when δ_N

is small so that a large fraction of the population is born at each time step. However, as we show below, non–trivial stable equilibria are in most cases unique, so that any stronger notion of stability would either provide the same refinement, or it would leave only the all-defect equilibrium.[20] But since, in general, one may desire a stronger stability concept, one can think of our results as characterizing a best-case outcome in this sense.

4 Stable equilibria without inheritance or punishment

We can easily write equations (2) and (3) explicitly in the case where $\mu = \nu = 0$. We have, in particular, that

$$u_C = \left(\frac{k}{1-\delta}\right)\left(\frac{2q - b(1-q)(1-\delta^2)}{1-\delta^2(1-q)} - b(1-q)\right) \quad , \tag{5}$$

$$u_D = u_I = \left(\frac{k}{1-\delta}\right)\left(\frac{(1+a)q(1-\delta^2)}{1-\delta^2(1-q)} + (1+a)q\right). \tag{6}$$

An equilibrium (Definition 1) is a probability of cooperation for each kind of entering node, and an associated steady state, such that every entering agent makes an optimal choice between defection and cooperation at birth, maintains links only with cooperators, and the steady state is consistent with these optimal choices over time. A stable equilibrium (Definition 2) is one in which, if the level of cooperation is perturbed up (down), and optimal choices are re-computed at the new steady state, then the level of cooperation will decrease (increase) as new agents enter. We now fully characterize the set of stable equilibria.

Proposition **1.** *When $\mu = \nu = 0$, there is always a stable equilibrium at $q = 0$. There is at most one other stable equilibrium, as follows.*

1. *If $a < \frac{\delta^2}{2-\delta^2}$, then there is a stable equilibrium at $q = 1$.*
2. *If $a > \frac{\delta^2}{1-\delta^2}$, then there are no other stable equilibria.*
3. *If $a \in \left(\frac{\delta^2}{2-\delta^2}, \frac{\delta^2}{1-\delta^2}\right]$, then there is another stable equilibrium if and only if*

$$b < \frac{2(\delta^4 + a(2-3\delta^2+\delta^4))}{(2-\delta^2)^2} - \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2} \quad ,$$

in which case $0 < q < 1$.

[20] If our stable equilbrium was not unique, a stronger refinement could be useful but it would not alter the comparative static results that we obtain.

The proof of Proposition 1, and those of the following results, is in Appendix B. The proof works by building a function $V(q)$ that is proportional to $u_C(q) - u_D(q)$ and analyzing its properties. The interior equilibria thus occur at the zeros of V. Applying stability in the sense of Definition 2, an interior equilibrium q^* is stable if and only if $V'(q^*)$ is negative, which implies that near q^*, $u_C(q) > u_D(q)$ for $q < q^*$ and conversely for $q > q^*$. See Figure 2 for illustrations of V to compare with the following discussion.

The technical reason for the uniqueness of a non-trivial stable equilibrium is concavity of $V(q)$ (which implies a unique zero where V is decreasing). Economically, this concavity derives from two facts. First, defectors benefit at a nearly constant rate from an increase in cooperation, since most of their payoff is dictated by the proportion of cooperators they meet with their out-links, which is linear in q. Second, the marginal returns for cooperators are decreasing in q. To see an intuition for this, note that cooperators approach an asymptotic neighborhood of other cooperators over time, and once they near this state early enough in their lives, further increases in cooperation do little to improve their lifetime utility. So, we have argued that u_D is nearly linear and that u_C is concave, which implies that V is concave. In other words, stability requires that the marginal benefit from an increase in q is higher for a defector than for a cooperator. If that is true at a given equilibrium q^*, then the utility of defection can only increase relative to the utility of cooperation with all further increases in q above q^*, ruling out the possibility of another (even unstable) equilibrium above q^*.

Having discussed uniqueness, we next note that the temptation for a cooperator to defect is increasing in the stage game payoff a. This drives the conclusion that for large enough a no cooperation can be sustained (part 2), while for small enough a, full cooperation is a stable equilibrium (part 1). To gain an intuition for the bound below which full cooperation is possible, consider the limiting case $\delta \to 1$, in which case the condition reduces to $a < 1$. In this case, a cooperator has on average twice as many links as a defector would have (a defector has essentially all out-links, whereas a cooperator has, in addition, an equal number of in-links). Thus optimality of defection requires double the per-link payoff as cooperation, i.e., that $1 + a > 2 \cdot 1$, which in turn requires $a > 1$. The intermediate result of partial cooperation (part 3) requires not only that a fall between the two bounds above, but also that b not be too large. This final condition requires that a cooperator not suffer too much when meeting a defector, which is necessary for cooperators to survive in the presence of defectors (interior q).

Observe that one implication of Proposition 1 is that, fixing a and b, as $\delta \to 1$, (i) if a and b are large, then no cooperation is possible, (ii) if b is small, then there exists a cooperative equilibrium, (iii) whether or not the cooperative equilibrium involves full cooperation or only partial cooperation is determined by the value of a, and finally (iv) when the cooperative equilibrium exists, it is quite stable in the sense that the basin of attraction expands to include all (arbitrarily low) cooperation levels.

Figure 2 depicts the possibilities for the qualitative shape of V. When there is a unique zero of V on $[0, 1]$, as in the top–left panel of Figure 2, then $V(1)$ must be positive, implying that $u_C > u_D$ when all agents cooperate, i.e., full cooperation is an equilibrium. Since u_C is independent of a, while u_D is increasing in a, $V = u_C - u_D$ is decreasing in a. Thus, this case obtains for a sufficiently small, noting that $V(0) < 0$ always, i.e., all-defection is always an equilibrium. When V has two zeros, as in the top–right panel of Figure 2, only the larger equilibrium is stable, and we characterize the range of parameters a and b for which this occurs, requiring an intermediate value of a and a sufficiently small value of b. The bottom–right panel of Figure 2 illustrates part 1 of Proposition 1, in which a is large enough that there is no cooperative equilibrium. Finally, the bottom–left panel of Figure 2 illustrates the stability regions in the (a, b) plane, when $\delta = 0.9$. Notice that the parameters from the other three panels are depicted in this graph, with the corresponding implications for equilibrium properties.

The conditions on payoffs introduced in Section 3 ($a, b > 0$ and $a - b < 1$) can, to some extent, be relaxed. In particular, they are important for justifying optimality of the network dynamics that we analyze but, given the dynamics, the equilibrium characterization of cooperating and defecting hold more generally.

To be more precise, $b > 0$ is used only in that it guarantees the optimality of a cooperator severing an in–link from a defector, $a > 0$ ensures that $q = 1$ is not a trivial equilibrium, and $a - b < 1$ ensures that mutual cooperation is efficient – otherwise there can be more efficient arrangements involving CD links. See Figure 2 for a graphical representation of these inequalities.

We remark that the analysis of equilibrium, as depicted in Figure 3, displays a hysterisis effect. Consider parameters for which there exists a stable non-trivial equilibrium, say $a = 0.8$, which also corresponds to the top right panel of Figure 2. Now consider a change in parameters such that this equilibrium fails to exist, as is the case in the bottom-right panel.[21] In particular consider an increase in a, say to $a = 1.0$. The outcome of the economy must now shift, discontinuously, to the unique equilibrium, in which all players defect. Then, if the shock is temporary and parameters return to their original values that support the non-trivial equilibrium, the economy cannot be expected to leave the all–defect state, as it constitutes a stable equilibrium. In general, whether the economy is in the all–defect state or in the non-trivial equilibrium, when the latter exists, is, in part, determined by the historical path of the economy.

[21] Comparative statics are discussed in the next section. There we show, for example, that an increase in either a or b would produce such a shift.

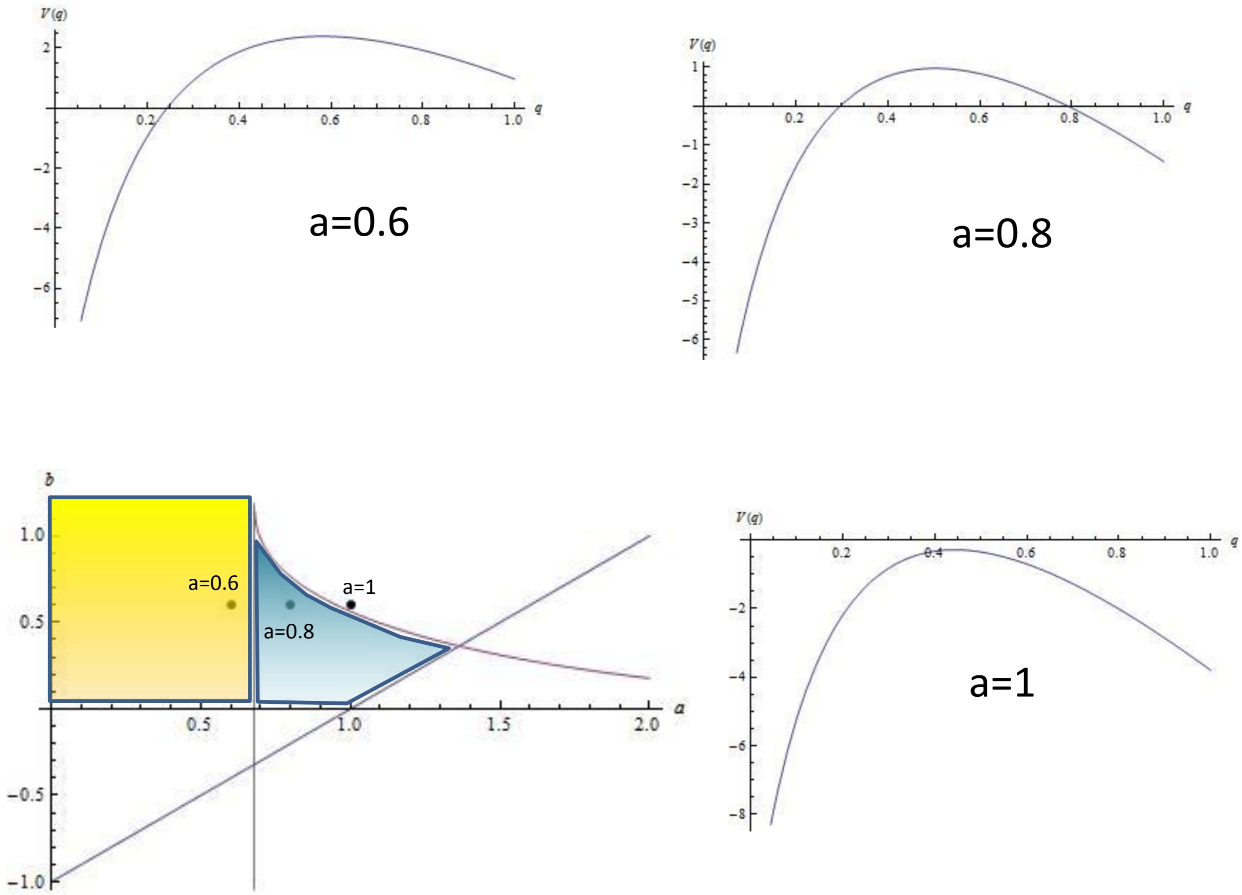


Figure 2: The function $V(q) = u_C - u_D$ for $\delta = 0.9$, $b = 0.6$, and $a \in \{0.6, 0.8, 1\}$. The zeros correspond to equilibria, as well as the positive value for $q = 1$ in the upper-left panel. In each panel, the larger equilibrium is stable. In the bottom–left panel, the regions of (a, b) pairs that generate non-trivial equilibria are displayed for $\delta = 0.9$ (lighter for stable ones with $q = 1$ – where this region which is drawn as rectangular but extends to infinity for any positive b – and darker for stable ones with $0 < q < 1$.) , where the points corresponding to the first three plots are depicted. The bottom–left plot is enlarged in Figure 4, showing more details that are used in the proof of Proposition 1.

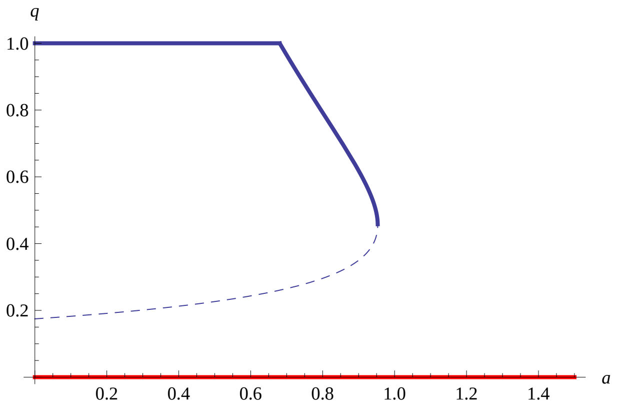


Figure 3: Equilibrium correspondence as a function of a, when $b = 0.6$ and $\delta = 0.9$. The flat line at $q = 0$ corresponds to the all–defect equilibrium. The solid curve corresponds to the stable non-trivial equilibrium, while the dashed curve corresponds to the unstable equilibrium.

5 Expulsion of immigrants

We now present the findings on the policy instrument of targeting defecting immigrants with the threat of expulsion from society, which constitute the main results of the paper. First, increasing the expulsion rate increases the equilibrium level of cooperation (Proposition 3). However, the cooperation rate among citizens is decreasing in the expulsion rate (Proposition 4). This effect is strict and smooth when heterogeneity in the population is accounted for (Proposition 5).

5.1 Payoffs, policy, and inheritance

We begin with a preliminary observation that extends the functional forms of utilities to the case that accomodates inheritance and expulsion.

Lemma **2.** *In the general case, with $\nu > 0$ and $\mu > 0$, the following are true.*

(i) u_C is independent of a and ν, and linearly decreasing in b.

(ii) u_D is independent of b and ν, and linearly increasing in a.

(iii) u_I is independent of b, linearly increasing in a and decreasing in ν.

The proof is conceptually simple, as it is still possible to express u_C, u_D and u_I explicitly. In this way, even though the expressions are cumbersome, the above dependencies are easy to check.

We now show that the effects of changes to the payoff parameters produce the intuitive results in equilibrium.

Proposition **3.** *If an interior stable equilibrium q^* exists then:*

(i) the marginal effects of an increase in parameters a and b on q^* are negative, and

(ii) the marginal effect of an increase in ν on q^* is positive.

The proposition focuses on interior equilibria, since comparative statics are trivial otherwise. For the proof, as with Proposition 1, we consider a function $V(q)$ that is proportional to $u_C(q) - u_I(q)$. Instead of deriving equilibria explicitly, we apply the implicit function theorem, using the fact that in a stable equilibrium it must be that $V(q)$ is decreasing.[22]

The comparative statics with respect to δ are left out of our analysis. Apart from the fact that it is technically more demanding, from an applied point of view, the subjective discount factor modelled by δ is not a ready target of change in any policy.

5.2 Effect of immigrant expulsion on citizen behavior

While expulsion of immigrants incentivizes them to cooperate, there is an equilibrium effect whereby the increased level of cooperation makes defection relatively more attractive to citizens. In this sense, a policy of expulsion may increase the defection rate among citizens.

When $\nu > 0$ it is clear that $u_D > u_I$. Thus, it must be that $p_D \leq p_I$ in equilibrium, with at most one of them interior. The interesting case, in which the stable equilibrium is interior, is when $p_D = 0$ and $0 < p_I < 1$.[23] Simplifying equation (1) accordingly, we have

$$q = 1 - \frac{(1-\delta)(1-p_I)}{1-\delta+\delta\nu p_I} \quad .$$

We now define the correspondence $p_D^*(\nu; a, b, \delta, \mu)$ which, given all other parameters of the model, maps ν into the set of values of p_D that obtain in a stable non-trivial equilibrium. The next result shows that $p_D^*(\nu)$ is decreasing.

Proposition **4.** *Consider a set of parameters for which there is a stable equilibrium with $0 < q^* < 1$ when $\nu = 0$. Then $p_D^*(\nu)$ is monotone decreasing.*

We remark that the monotonicity of $p_D^*(\nu)$ takes a simple form. In particular, $p_D^*(0) = [0, \bar{p}_D]$ for some $\bar{p}_D > 0$, and $p_D^*(\nu) = 0$ for all $\nu > 0$. In this case, the

[22] We do not prove uniqueness of non–trivial stable equilibria for this general case, although we have not found numerical examples in which multiplicity arises.

[23] This follows from equation (1). If $p_I = 1$ then $q = 1$. If $p_I = 0$ then $q = 0$.

equilibrium without expulsion generically involves a positive rate of cooperation among citizens who are born to defectors. As soon as any policy of expulsion is implemented, no matter how weak it may be, it drives the cooperation rate among these citizens immediately to zero.

Under the hypotheses of Proposition 4 , when $\nu = 0$ it is natural to take $p_D = p_I > 0$, since immigrant defectors and citizen defectors can effectively be pooled into one population. What the result demonstrates that, when immigrant defectors are punished, behavior necessarily adjusts so that immigrant defectors remain indifferent between cooperating and defecting, which immediately implies that offspring of defectors *strictly* prefer to defect. The introduction of a wedge in incentives between immigrants and citizens results in a discontinuous shift into defecting behavior for citizens.

5.3 Heterogeneity across the agents

Proposition 4 describes a discontinuous effect of the expulsion policy on cooperation among citizens. The discontinuity derives from the simplifying assumption in our model that agents have identical preferences, so that a small change in incentives can shift the behavior of a large mass of agents. We now show that allowing for heterogeneity of payoffs smooths out this effect such that p_D is uniquely determined, even at $\nu = 0$ and, more importantly, it is strictly decreasing in ν.

To this end we augment the model by assuming that the temptation payoff, a, is drawn from a continuous distribution Φ. Entering agents thus generically have strict preferences between cooperating and defecting. An interior equilibrium is now characterized by the system given by equation (1) and the two indifference conditions

$$u_D(a) = u_C \ , \tag{7}$$

$$u_I(a) = u_C \ , \tag{8}$$

where $u_D(a)$ and $u_I(a)$ express the expected utility that an agent with temptation payoff a obtains by defecting, when she is, respectevily, a citizen or an immigrant. Equations (7) and (8) can each be solved for a unique value of a, to yield thresholds a_D and a_I, below which a citizen and an immigrant, respectively, strictly prefer to cooperate and above which they strictly prefer to defect.[24] These, in turn, generate the probabilities of interest via $p_D = \Phi(a_D)$ and, similarly, $p_I = \Phi(a_I)$. It is straightforward adopt Definitions 1 and 2 to this context, as follows.

[24] There is at most one solution for $q > 0$, which is the case of interest.

Definition **3** (Interior stable equilibrium)**.** *An interior equilibrium is a sixtuple* $(a_I^*, a_D^*, \{\sigma_{XY}\}_{X,Y\in\{C,D\}})$ *such that:*

- $\sigma_{CC} = \sigma_{DC} = 1$ *and* $\sigma_{CD} = \sigma_{DD} = 0$*, from Observation 1;*
- $q(p_I, p_D)$ *is a steady state given* $p_I = \Phi(a_I^*)$ *and* $p_D = \Phi(a_D^*)$*;*
- a_I^* *and* a_D^* *satisfy Equations (7) and (8) given* q*.*

Further, q *is stable if there is an* $\epsilon > 0$ *such that the following hold:*

- *If* $q < 1$*, then for every* $q' \in (q, q+\epsilon)$*, and for every* (a'_I, a'_D) *that solves Equations (7) and (8) given* q'*, the steady state* $q(\Phi(a'_I), \Phi(a'_D)) < q'$*;*
- *If* $q > 0$*, then for every* $q' \in (q-\epsilon, q)$*, the above inequality is reversed.*

It is easy to check that if the support of Φ contains only positive numbers, then the autarky equilibrium with $q = p_I = p_D = 0$ is always an equilibrium. However, stable non-trivial equilibria also exist, as can be shown through a *purification* argument *á la* Harsanyi (1973), such that one can construct a sequence of pure strategy equilibria that converge to the non-trivial stable equilibrium discussed in Section 4 at the limit of vanishing heterogeneity. The following result studies those non-trivial stable equilibria and extends the comparative statics finding of Proposition 4 to the context with preference heterogeneity.

Proposition **5.** *Consider a continuous distribution* Φ*, and fix parameters such that there is an interior stable equilibrium for* $\nu = 0$*. This characterizes a unique value of* $p_D^*(0)$*. If this* $p_D^*(0) > 0$*, then for any* $\nu \geq 0$ *we have that* $\frac{dp_D^*}{d\nu} < 0$*.*

This result relies on the existence of an interior equilibrium, which depends on the parameters of the model, including now the distribution Φ. Without punishment, interiority requires two conditions. First, b must be small enough to permit cooperation, as in Proposition 1. Second, the support of Φ must, in a sense, be wide enough. This latter condition is required because an interior equilibrium obtains when the indifferent type a_i falls in the interior of the support of Φ.

Finally, as is shown in the proof, when Φ puts more mass on the indifferent type a_D, so that there are more citizens who are nearly indifferent between cooperating and defecting, the effect of increasing punishment on citizens' cooperation rate is higher. In Appendix C we run simulations that provide evidence for the quantitative effects of this result.

6 Conclusion

We have developed a model to study specific aspects of how the flow of immigration influences a nation's economy. We have paid particular attention to the incentive effects of punishing immigrants who defect on their partners. Our main result is that a policy of expelling such immigrants is not likely to be an optimal, or even desirable, policy instrument. This conclusion derives from an intuition brought out in our analysis: the fact that behaviors in society balance the incentives between cooperation and defection. That is, there is a sense in which there is a natural (equilibrium) level of cooperation. Thus, if the number of defectors is reduced by expelling some of them, then there arises a tendency for others to shift behaviors and recover a balance near the original level of cooperation.

The suggestion of this effect is perhaps more general: if a policy changes the incentives of only a subset of the population, then its efficacy may be mitigated by a substitution effect through which other individuals change behaviors so as to restore the natural level of cooperation as dictated by the payoffs and parameters of society. Instead, a more efficient approach may involve a policy that changes the incentives of all individuals, regardless of their status (as immigrants or otherwise). In the specific context of our model, this could take the form, e.g., of reducing the penalty of meeting defectors (b in the analysis) by providing insurance against being transgressed upon.

We stress that, even though the model suggests the suboptimality of certain classes of policies, it is not possible to conduct an optimal policy analysis – or even a welfare analysis – with this approach. First, the model is perhaps too stylized to make such a question directly policy-relevant. Second, it would necessarily require one to take a stance on the relative costs of different policies, which brings the analysis outside the scope of what we consider. Fruitful extensions of this work may involve (i) explicitly modeling the costs of an expulsion policy, as well as the costs of other policies, (ii) extending the modeling of behavior to account for other stage games or the possibility of non-stationary behavior, and (iii) understanding incentives outside of steady states.

Acknowledgements

We thank for their comments Pierpaolo Battigalli, James Dana, Osborne Jackson, Andrea Moro, Fernando Vega Redondo, Myrna Wooders, and audiences at Bocconi University, University of Copenhagen, University of Missouri, Northeastern University, Sussex University, Vanderbilt University, University of Venice, at the 2013 Dynamic Interactions for Economic Theory Conference and at the 2013 WIN Workshop. P.P. received support from the Italian Ministry of University under FIRB project

RBFR1269HZ “Social and spatial interactions in the accumulation of civic and human capital”.

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Ali, S. N. and D. A. Miller (2013). Enforcing cooperation in networked societies. *Unpublished manuscript, University of California at San Diego*.

Bisin, A. and T. Verdier (2012). The economics of cultural transmission and socialization. *Handbook of Social Economics,* J. Benhabib, A. Bisin and M. O. Jackson (eds.), Amsterdam: North Holland.

Blonski, M. (1999). Anonymous games with binary actions. *Games and Economic Behavior 28*(2), 171–180.

Borjas, G. J. (1994). The economics of immigration. *Journal of economic literature*, 1667–1717.

Borjas, G. J. (2003). The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. *The quarterly journal of economics 118*(4), 1335–1374.

Borjas, G. J. (2008). *Issues in the Economics of Immigration.* University of Chicago Press.

Borjas, G. J. (2014). *Immigration economics.* Harvard University Press.

Cortes, P. (2008). The effect of low-skilled immigration on us prices: evidence from cpi data. *Journal of political Economy 116*(3), 381–422.

Dall’Asta, L., M. Marsili, and P. Pin (2012). Collaboration in social networks. *Proceedings of the National Academy of Sciences 109*(12), 4395–4400.

Datta, S. (1996). Building trust. Technical report, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Djajić, S. (1987). Illegal aliens, unemployment and immigration policy. *Journal of Development Economics 25*(1), 235–249.

Dustmann, C. and T. Frattini (2014). The fiscal effects of immigration to the uk. *The Economic Journal 124*(580), F593–F643.

Dustmann, C., T. Frattini, and I. P. Preston (2013). The effect of immigration along the distribution of wages. *The Review of Economic Studies 80*(1), 145–173.

Ellison, G. (1994). Cooperation in the prisoner's dilemma with anonymous random matching. *The Review of Economic Studies 61*(3), 567–588.

Fosco, C. and F. Mengel (2011). Cooperation through imitation and exclusion in networks. *Journal of Economic Dynamics and Control 35*(5), 641–658.

Fudenberg, D. and E. Maskin (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica: Journal of the Econometric Society*, 533–554.

Fujiwara-Greve, T. and M. Okuno-Fujiwara (2009). Voluntarily separable repeated prisoner's dilemma. *The Review of Economic Studies 76*(3), 993–1021.

Ghosh, P. and D. Ray (1996). Cooperation in community interaction without information flows. *The Review of Economic Studies 63*(3), 491–519.

Hanson, G. H. (2007). *The economic logic of illegal immigration.* Council Special Report No. 26, Council on Foreign Relations.

Hanson, G. H. (2010). *The Economics and Policy of Illegal Immigration in the United States.* Migration Policy Institute.

Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory 2*(1), 1–23.

Immorlica, N., B. Lucier, and B. Rogers (2010). Emergence of cooperation in anonymous social networks through social capital. In *Proceedings of the 11th ACM Conference on Electronic Commerce.*

Immorlica, N., B. Lucier, and B. Rogers (2013). Cooperation in anonymous dynamic social networks. Mimeo.

Izquierdo, L., S. Izquierdo, and F. Vega-Redondo (2010). The option to leave: Conditional dissociation in the evolution of cooperation. *Journal of Theoretical Biology 267*(1), 76–84.

Izquierdo, L., S. Izquierdo, and F. Vega-Redondo (2014). Leave and let leave: A sufficient condition to explain the evolutionary emergence of cooperation. *Journal of Economic Dynamics and Control 46*, 91–113.

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (2012). Social capital and social quilts: Network patterns of favor exchange. *The American Economic Review 102*(5), 1857–1897.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of economic theory 71*(1), 44–74.

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies 59*(1), 63–80.

Kemnitz, A. and K. Mayr (2012). Return migration and illegal immigration control. Technical report, Norface Research Programme on Migration, Department of Economics, University College London.

Kranton, R. E. (1996). The formation of cooperative relationships. *Journal of Law, Economics, and Organization 12*(1), 214–233.

Levine, P. (1999). The welfare economics of immigration control. *Journal of Population Economics 12*(1), 23–43.

Mas-Colell, A., M. D. Whinston, and J. Green (1995). Microeconomic theory.

Mastrobuoni, G. and P. Pinotti (2014). Legal status and the criminal activity of immigrants. *Fortcoming on American Economic Journal: Applied Economics*.

Shenk, M. K., M. B. Mulder, J. Beise, G. Clark, W. Irons, D. Leonetti, B. S. Low, S. Bowles, T. Hertz, A. Bell, and P. Piraino (2010). Intergenerational wealth transmission among agriculturalists. *Current Anthropology 51*(1), 65–83.

Vega-Redondo, F. (2006). Building up social capital in a changing world. *Journal of Economic Dynamics and Control 30*(11), 2305–2338.

Watson, J. (1999). Starting small and renegotiation. *Journal of economic Theory 85*(1), 52–90.

Watson, J. (2002). Starting small and commitment. *Games and Economic Behavior 38*(1), 176–199.

Appendix A Commitment to behavior and inheritance of cooperation

We have assumed in the analysis above that the offspring of a cooperator chooses to cooperate, i.e., that $p_C = 1$. The intuition is that the original incentive for the parent to cooperate derives specifically from the expectation of accumulating relationships with other cooperators. As this valuable network of relationships – the impetus to cooperate in the first place – is inherited by the offspring, it is natural that the offspring cooperates as well. However, there exist special situations under which the offspring may instead prefer to defect. We are concerned in this section with arguing that such situations can be safely ruled out without affecting the force of the argument in the main text.

To this end we show that under a certain mildly restrictive condition, every offspring of a cooperator prefers to cooperate.

We introduce the following condition, borrowed from Immorlica et al. (2013), in the form appropriate for our analysis.

Definition **A.** *We say that the* value of social ties is positive *at a steady state level of cooperation q if*

$$\frac{1+b}{1+a} \geq 1 - (1-q)\delta^2.$$

Definition A characterizes the situations for which the accumulation of links with cooperators tilts incentives in favor of cooperation, with the implication that if an agent is willing to cooperate at birth, then it is also willing to cooperate at any later period. To see this, notice that in order for cooperation to always be sequentially rational, it is sufficient (and actually necessary) that the marginal gain to meeting a cooperator instead of a defector is bigger when cooperating ($\frac{1+b}{1-\delta^2(1-q)}$) compared to when defecting $(1 + a)$.

Notice that the value of social ties depends both on the payoff parameters a and b and also on the endogenously determined level of cooperation q^*. Thus in general one has to compute the equilibrium outcome before determining whether or not the value of social ties is positive at the equilibrium corresponding to a given set of parameters. Notwithstanding this observation, notice that there is a simple sufficient condition on payoffs to ensure that that the value of social ties is positive. Namely, it is enough that $b \geq a$, which is simply to say that the stage game payoffs are supermodular. With this in mind we present:

Proposition **A.** *If the value of social ties is positive at a non-trivial equilibrium $q^* > 0$, then it is optimal for every offspring of a cooperator to cooperate.*

Proof: The result follows from Theorem 2 in Immorlica et al. (2013). □

They show that under the hypotheses of our proposition, the marginal value of a link with a cooperator is maximized by perpetual cooperation. It is therefore sequentially rational for a cooperating agent to continue cooperating at every history. Since our agents make a one-time decision, an offspring of a cooperator is in the same position of a corresponding agent in Immorlica et al. (2013) who has been cooperating for the same duration as the offspring's parent.

The result also shows that, when the value of social ties is positive, at no history would an agent ever prefer to change behavior from cooperation to defection, or vice versa. To see why, notice that since a cooperator could have an offspring at any period with positive probability, if such an offspring always prefers to cooperate, then in the event that the offspring was not born, but the parent was given an opportunity to revise his strategy, his continuation payoffs are identical to the liftetime payoffs of the hypothetical offspring.

Appendix B Proofs

Proof of Proposition 1 (page 18): The structure of equilibria can be derived from the properties of the following function

$$V(q) \equiv \left(\frac{1-\delta}{k}\right)(u_C - u_D) \quad = \quad \frac{2q-(1-\delta^2)f(q)}{g(q)} - f(q),$$

obtained from (5) and (6) by setting $f(q) \equiv (1+a-b)q + b$ and $g(q) \equiv 1-\delta^2(1-q)$.

We make the following observations. To summarize the argument, points (A) and (B) characterize boundary equilibria, point (C) shows concavity of $V(q)$, and the remainder of the proof characterizes stable interior equilibria by noting that, given the preceeding points, their existence is equivalent to a global maximum of $V(q)$ that occurs for $0 < q^* < 1$ and has $V(q^*) > 0$.

(A) $V(0) = -2b$ is always negative.

(B) $V(1) = \delta^2 - a(2-\delta^2)$ is nonnegative if and only if

$$a \leq \frac{\delta^2}{2-\delta^2} \quad .$$

(C) Since $g(q) > 0$ for all $\delta < 1$, the second derivative of $V(q)$ has the same sign as

$$\begin{aligned} V''(q)\cdot(g(q))^3 \quad &= \quad 2\Big(2q-(1-\delta^2)f(q)\Big)(g')^2 - 2g(q)g'(q)\Big(2-(1-\delta^2)f'(q)\Big) \\ &= \quad -2\delta^2(1-\delta^2)\Big(1+b+\delta^2-a(1-\delta^2)\Big), \qquad \text{(a)} \end{aligned}$$

which is always negative given that $a < 1 + b$, as we assume throughout.

(D) The first order condition, $V'(q) = 0$, yields

$$q^* \quad = \quad -\frac{(1-\delta^2)}{\delta^2} \pm \frac{\sqrt{(1-\delta^2)\Delta}}{\delta^2(1+a-b)} \quad , \tag{b}$$

using $\Delta = (1+a-b)\Big(1+b+\delta^2 - a(1-\delta^2)\Big)$. Four things should be noted:

(i) q^* is defined in the real numbers if and only if $\Delta \geq 0$.

(ii) Since $-\frac{(1-\delta^2)}{\delta^2} < 0$, there is at most one $q > 0$ that satisfies the first order condition. We will call q^* just this positive-valued solution from (b), if it exists.

(iii) Let $\Delta \geq 0$. There exists $q^* \geq 0$ if and only if

$$\frac{\sqrt{(1-\delta^2)\Delta}}{|1+a-b|} \geq 1 - \delta^2 \quad ,$$

which is equivalent to

$$|1 - a + b + \delta^2(1+a)| \geq (1-\delta^2)|1+a-b| \quad .$$

If $1 + a - b > 0$, this is equivalent to

$$b \geq 2\left(1 + a - \frac{2+a}{2-\delta^2}\right) \equiv \underline{b}(a,\delta) \quad ; \tag{c}$$

otherwise we need the opposite inequality.

(iv) Let $\Delta \geq 0$. There exists $q^* \leq 1$ if and only if

$$\frac{\sqrt{(1-\delta^2)\Delta}}{|1+a-b|} \leq 1 \quad ,$$

which is equivalent to

$$(1-\delta^2)|1 - a + b + \delta^2(1+a)| \leq |1+a-b| \quad .$$

If $1 + a - b > 0$, this is equivalent to

$$b \leq -\delta^2(1+a) + \frac{2(a+\delta^2)}{2-\delta^2} \equiv \bar{b}(a,\delta) \quad ; \tag{d}$$

otherwise we need the opposite inequality.

(E) Let $\Delta \geq 0$. We check the conditions under which both inequalities (c) and (d) can be satisfied. We have that $\underline{b}(0,\delta) = \frac{-2\delta^2}{2-\delta^2} < 0 < \frac{\delta^4}{2-\delta^2} = \bar{b}(0,\delta)$. Moreover

$$0 < \frac{\partial}{\partial a}\underline{b}(a,\delta) = \frac{2-2\delta^2}{2-\delta^2} < \frac{2-2\delta^2+\delta^4}{2-\delta^2} = \frac{\partial}{\partial a}\bar{b}(a,\delta) \quad , \tag{e}$$

which implies that for all $a \geq 0$, $\underline{b}(a,\delta) < \bar{b}(a,\delta)$. Thus if if $1+a-b<0$, there is no b for which $q^* \in [0,1]$, while if $1+a-b>0$, then there exists $q^* \in [0,1]$ if and only if $b \in \big[\max\{\underline{b}(a,\delta),0\}, \bar{b}(a,\delta)\big]$, where this interval is always non–empty.[25]

(F) We have that, if $\Delta \geq 0$, for a valid $q^* \in [0,1]$,

$$V(q^*) \quad = \quad -b + \frac{2}{\delta^2} - \frac{2\sqrt{(1-\delta^2)\Delta}}{\delta^2} \quad .$$

(G) If $\Delta \geq 0$, then $V(q^*) > 0$ if and only if

$$\sqrt{\Delta} \leq \frac{2-b\delta^2}{2\sqrt{1-\delta^2}} \quad ,$$

which is equivalent to requiring both that $b < 2/\delta^2$ and (taking squares it becomes a second order polynomial in b)

$$b \quad \notin \quad \left(\tilde{b}(a,\delta) - \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2}\ , \tilde{b}(a,\delta) + \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2}\right) \text{(f)}$$

where we have defined $\tilde{b}(a,\delta) \equiv \frac{2(\delta^4 + a(2-3\delta^2+\delta^4))}{(2-\delta^2)^2}$.

Note that:

(i) This interval is defined if and only if $a(2-\delta^2)-\delta^2 \geq 0$, which is to say $a \geq \frac{\delta^2}{2-\delta^2}$.

(ii) When $a = \frac{\delta^2}{2-\delta^2}$, then the two endpoints of this interval reduce to $\tilde{b}(\frac{\delta^2}{2-\delta^2},\delta) = \bar{b}(\frac{\delta^2}{2-\delta^2},\delta)$, where the latter is defined in (d).

(iii) $\frac{\partial}{\partial a}\tilde{b}(a,\delta) < \frac{\partial}{\partial a}\bar{b}(a,\delta)$, where the latter is computed in (e).

(iv) We always have $1+a > \tilde{b}(a,\delta)$ because $\tilde{b}(0,\delta) < 1$ and $\frac{\partial}{\partial a}\tilde{b}(a,\delta) < 1$.

This means that, for $a \geq \frac{\delta^2}{2-\delta^2}$, condition $1+a-b \geq 0$, together with both conditions (d) and (f), are all satisfied whenever

$$b \leq \tilde{b}(a,\delta) - \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2} \quad . \tag{g}$$

(H) There is no relevant solution for b above the upper bound in (f), which would have the form $\tilde{b}(a,\delta) + \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2} < b < \bar{b}(a,\delta)$. To see this, note that setting $\bar{b}(a,\delta) = \tilde{b}(a,\delta) + \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2}$ produces $b = \frac{2\left(-2+d^2\right)^2}{d^6} > 2/\delta^2$, and so is excluded by the requirement above that $b < 2/\delta^2$.

[25] The case $1+a-b<0$ could have been excluded immediately by noting that $a-b<1$ implies $1-a+b+\delta^2(1+a)>0$, and then $1+a-b<0$ would imply $\Delta<0$. The advantage of our proof is that we prove that the condition $1+a-b<0$ must be excluded even without assuming that $a-b<1$. We also prove that when $1+a-b>0$ and $a-b<1$, then $\Delta>0$.

(I) Both $\tilde{b}(a,\delta) - \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2}$ and $\underline{b}(a,\delta)$ equal 0 if and only if $a = \frac{\delta^2}{1-\delta^2}$. Moreover, for every $a > \frac{\delta^2}{2-\delta^2}$, and so in particular for $a > \frac{\delta^2}{1-\delta^2}$ we have that

$$\frac{\partial}{\partial a}\left(\tilde{b}(a,\delta) - \frac{4\sqrt{\delta^2(1-\delta^2)(a(2-\delta^2)-\delta^2)}}{(2-\delta^2)^2}\right) < \frac{\partial}{\partial a}\tilde{b}(a,\delta) < \frac{\partial}{\partial a}\underline{b}(a,\delta) \quad .$$

This means three things:

(i) Condition (g) can never be satisfied together with condition (c), if $a > \frac{\delta^2}{1-\delta^2}$.

(ii) Condition (c) is never binding, because whenever $\underline{b}(a,\delta)$ is positive, it falls inside the interval in (f).

(iii) So, the only lower bound that matters for b is that $b \geq a-1$.

This concludes the proof. A graphical representation of the curves defined in the proof is given in Figure 4. □

Proof of Lemma 2 (page 22): We compute (2)–(4) with $\delta_N = \delta + (1-\delta)\mu$, to yield

$$u_C = \frac{\frac{(b+1)q}{1-\delta(1-q)(\delta+\mu-\delta\mu)^2} - b(2-q) + \frac{q^2\left(\delta^2(\mu-1)-\delta\mu-1\right)}{(q+\delta^2-1)((q-1)(\delta+\mu-\delta\mu)^2+1)} + \frac{q(\delta+1)}{(q+\delta^2-1)(\delta(\mu-1)-1)}}{1-\delta} , \tag{h}$$

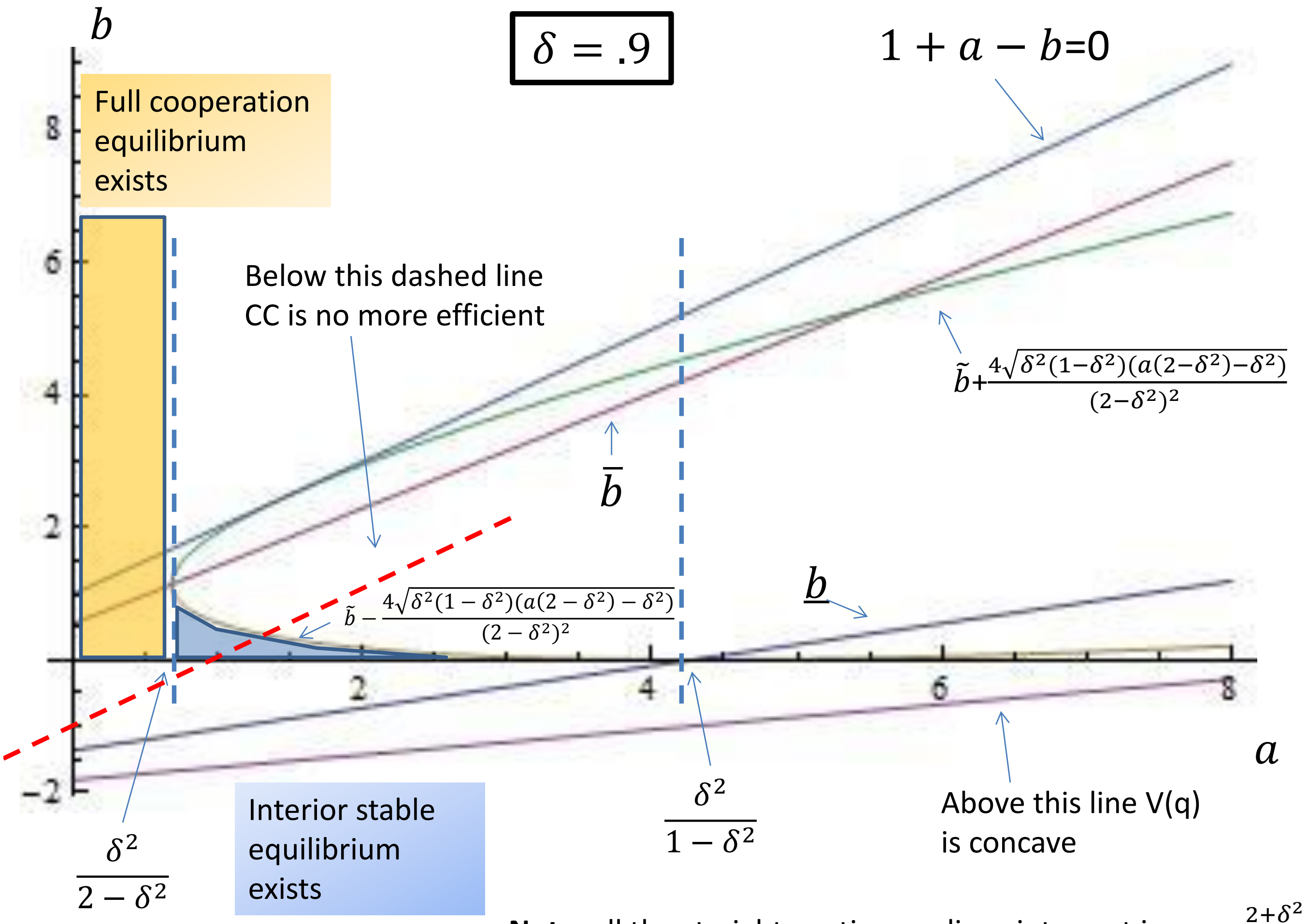
$$u_D = \frac{(a+1)q\left(1+\frac{1-(\delta+\mu-\delta\mu)^2}{1-(1-q)(\delta+\mu-\delta\mu)^2}\right)}{1-\delta} , \tag{i}$$

$$u_I = \frac{(a+1)q\left(1+\frac{1-(\delta+\mu-\delta\mu)^2}{1-(1-q)(\delta+\mu-\delta\mu)^2}\right)}{1-\delta(1-\nu)} . \tag{j}$$

The dependence of these three functions with respect to a, b and ν are evident. To see that $\frac{\partial u_C}{\partial b} < 0$, note that $\frac{q}{1-\delta(1-q)(\delta+\mu-\delta\mu)^2} < 2-q$ for any $0 \leq q < 1$, $0 \leq \delta \leq 1$ and $0 \leq \mu \leq 1$, because $\frac{q}{1-\delta(1-q)(\delta+\mu-\delta\mu)^2}$ is increasing in q, $2-q$ is decreasing in q, and they are equal when $q = 1$. □

Proof of Proposition 3 (page 23): An interior equilibrium is given by the condition that $u_C - u_I = 0$ and by equation (1). However, equation (1) does not depend on a and b, and is only a relation between p_D and p_I given q. So, we can consider only the implicit function $u_C - u_I = 0$, from equations (h) and (j).
The requirement for stability is that $\frac{\partial u_C}{\partial q} < \frac{\partial u_I}{\partial q}$, or equivalently $\frac{\partial(u_C - u_I)}{\partial q} < 0$.
We have also from Lemma 2 that $\frac{\partial u_C}{\partial a} = 0$, $\frac{\partial u_C}{\partial b} < 0$, $\frac{\partial u_C}{\partial \nu} = 0$, $\frac{\partial u_I}{\partial a} > 0$, $\frac{\partial u_I}{\partial b} = 0$ and $\frac{\partial u_I}{\partial \nu} < 0$.
By the implicit function theorem:

$$\frac{dq^*}{da} = -\frac{\partial(u_C - u_I)/\partial a}{\partial(u_C - u_I)/\partial q} = \frac{\partial u_I/\partial a}{\partial(u_C - u_I)/\partial q} < 0 \quad ,$$



Note: all the straight continuous lines intersect in a= - $\frac{2+\delta^2}{\delta^2}$

Figure 4: This is an enlargement of the bottom–left part of Figure 2, adding plots of the curves used in the proof of Proposition 1

$$\frac{dq^*}{db} = -\frac{\partial(u_C - u_I)/\partial b}{\partial(u_C - u_I)/\partial q} = -\frac{\partial u_C/\partial b}{\partial(u_C - u_I)/\partial q} < 0 \quad ,$$

and

$$\frac{dq^*}{d\nu} = -\frac{\partial(u_C - u_I)/\partial \nu}{\partial(u_C - u_I)/\partial q} = \frac{\partial u_I/\partial \nu}{\partial(u_C - u_I)/\partial q} > 0 \quad . \quad \square$$

Proof of Proposition 4 (page 23): When $\nu = 0$ equation (1) becomes simply

$$q = 1 - \frac{(1-\mu)(1-p_I)}{1-\mu+\mu p_D} \quad ,$$

from which the explicit relation between p_I and p_D is linear:

$$p_I = q - (1-q)\frac{\mu}{1-\mu}p_D \quad .$$

Moreover, we have that offspring of defectors and immigrants face the same incentives. By definition of interior equilibrium both $0 \leq p_D \leq 1$ and $0 < p_I < 1$ are admissible, as long as $0 < q < 1$. So, we allow for any couple of values (p_D, p_I) that give the same value for q in equation (1), and p_D can be any value between 0 (so that $p_I = q$) and $\min\{\frac{q}{1-q}\frac{1-\mu}{\mu}, 1\}$ (so that $p_I = \max\{0, q - (1-q)\frac{\mu}{1-\mu}\}$).
When $\nu > 0$, still we need p_I to be interior for q also to be, and then $u_D = u_I$ from equations (h) and (j). Then, considering also equation (j) that determines incentives of the offspring of defectors, we must have $p_D = 0$. In this case, solving equation (1), we have that $p_I = \frac{(1-\delta)q}{(1-\delta)+\delta\nu(1-q)}$.
Summing up, in an interior equilibrium characterized by a q that solves $u_D = u_I$, p_D can attain any value in the interval $[0, \min\{\frac{1}{q} + \frac{1}{\mu} - 1, 1\}]$ when $\nu = 0$, but must be 0 for $\nu > 0$. This proves the statement. $\square$

Proof of Proposition 5 (page 25): From equations (h)–(j), it is possible to obtain explicit unique solutions for a_D and a_I that solve respectively (7) and (8). Call them, as functions of all the other parameters of the model, $a_D(b, q, \mu, \delta)$ and $a_I(b, q, \mu, \delta, \nu)$. We can then define as implicit functions the relations between p_I, p_D, and q:

$$p_I - \Phi\left(a_I(b, q, \mu, \delta, \nu)\right) = 0 \quad , \tag{k}$$

$$p_D - \Phi\left(a_D(b, q, \mu, \delta)\right) = 0 \quad , \tag{l}$$

$$q - F(\mu, \delta, \nu, p_I, p_D) = 0 \quad , \tag{m}$$

where the last equation is just a way of writing (1).
In this system the endogenous variables are q, p_I and p_D. If zero is outside the support of Φ we necessarily have the trivial solution $(0, 0, 0)$. Any other set of solutions characterizes an equilibrium of interest, and p_D is uniquely determined by equation (l).

We are interested in the sign of the derivative $\frac{dp_D}{d\nu}$ in a stable equilibrium. We refer to the left-hand part of equation (k) as $Eq.(k)$, and so on for the other two. Consider first equation (j) and its implication on equation (k): From the assumption of being in a stable equilibrium, now that a_D is endogenous, if q rises it becomes more profitable to play D, and then we would need a lower a_D of indifference. In formulas, this implies that $\frac{\partial a_D(b,q,\mu,\delta,\nu)}{\partial q} \leq 0$, where this inequality is strict when a_D lies in the support of Φ. With the same reasoning we have $\frac{\partial a_I(b,q,\mu,\delta,\nu)}{\partial q} \leq 0$ and $\frac{\partial a_D(b,q,\mu,\delta,\nu)}{\partial \nu} \geq 0$. Then, we apply the implicit function theorem (as a reference see e.g. the mathematical appendix of Mas-Colell et al. 1995) to compute the marginal effects of ν on the endogenous variables:

$$D_\nu \begin{pmatrix} p_I \\ p_D \\ q \end{pmatrix} = - \begin{bmatrix} \frac{\partial Eq.(k)}{\partial p_I} & \frac{\partial Eq.(k)}{\partial p_D} & \frac{\partial Eq.(k)}{\partial q} \\ \frac{\partial Eq.(l)}{\partial p_I} & \frac{\partial Eq.(l)}{\partial p_D} & \frac{\partial Eq.(l)}{\partial q} \\ \frac{\partial Eq.(m)}{\partial p_I} & \frac{\partial Eq.(m)}{\partial p_D} & \frac{\partial Eq.(m)}{\partial q} \end{bmatrix}^{-1} D_\nu \begin{pmatrix} Eq.(k) \\ Eq.(l) \\ Eq.(m) \end{pmatrix} . \tag{n}$$

If we call $\Delta_{Iq} \equiv \Phi' \frac{\partial a_I(b,q,\mu,\delta,\nu)}{\partial q} \leq 0$, $\Delta_{Dq} \equiv \Phi' \frac{\partial a_D(b,q,\mu,\delta)}{\partial q} \leq 0$ (and this inequality is strict when a_D is in the support of Φ) and $\Delta_{I\nu} \equiv \Phi' \frac{\partial a_I(b,q,\mu,\delta,\nu)}{\partial \nu} \geq 0$, we obtain that (n) simplifies to

$$\begin{aligned} D_\nu \begin{pmatrix} p_I \\ p_D \\ q \end{pmatrix} &= - \begin{bmatrix} 1 & 0 & -\Delta_{Iq} \\ 0 & 1 & -\Delta_{Dq} \\ -\frac{\partial F(\mu,\delta,\nu,p_I,p_D)}{\partial p_I} & -\frac{\partial F(\mu,\delta,\nu,p_I,p_D)}{\partial p_D} & 1 \end{bmatrix}^{-1} D_\nu \begin{pmatrix} -\Delta_{I\nu} \\ 0 \\ -\frac{\partial F(\mu,\delta,\nu,p_I,p_D)}{\partial \nu} \end{pmatrix} \\ &= \frac{1}{1 - \Delta_{Iq} \cdot \frac{\partial F}{\partial p_I} - \Delta_{Dq} \cdot \frac{\partial F}{\partial p_D}} \begin{bmatrix} \dots & \dots & \dots \\ -\Delta_{Dq} \cdot \frac{\partial F}{\partial p_I} & \dots & -\Delta_{Dq} \\ \dots & \dots & \dots \end{bmatrix} D_\nu \begin{pmatrix} -\Delta_{I\nu} \\ 0 \\ -\frac{\partial F}{\partial \nu} \end{pmatrix} . \end{aligned}$$

In the last derivation we have used the fact that

$$\begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ \gamma & \delta & 1 \end{bmatrix}^{-1} = \frac{1}{1 - \alpha\gamma - \beta\delta} \begin{bmatrix} 1 - \beta\delta & \alpha\delta & -\alpha \\ \beta\gamma & 1 - \alpha\gamma & -\beta \\ -\gamma & -\delta & 1 \end{bmatrix} ,$$

placing dots for all the elements that are not relevant for our purposes. We then have

$$\frac{dp_D}{d\nu} = \frac{\Delta_{Dq}}{1 - \Delta_{Iq} \cdot \frac{\partial F}{\partial p_I} - \Delta_{Dq} \cdot \frac{\partial F}{\partial p_D}} \left(\Delta_{I\nu} \cdot \frac{\partial F}{\partial p_I} + \frac{\partial F}{\partial \nu} \right) . \tag{o}$$

This quantity is always non–positive as $1 - \Delta_{Iq} \cdot \frac{\partial F}{\partial p_I} - \Delta_{Dq} \cdot \frac{\partial F}{\partial p_D} \geq 1$, and $\frac{\partial F}{\partial \nu} > 0$. Also, it becomes strictly decreasing when $\Delta_{Dq} < 0$, which happens when $0 < p_D < 1$. Finally note that if $p_D = 1$, then also $p_I = 1$ and $q = 1$. So, for an interior equilibirum with $p_D > 0$, and $\nu \geq 0$, it is always the case that $\frac{dp_D^*}{d\nu} < 0$. □

Appendix C Simulations

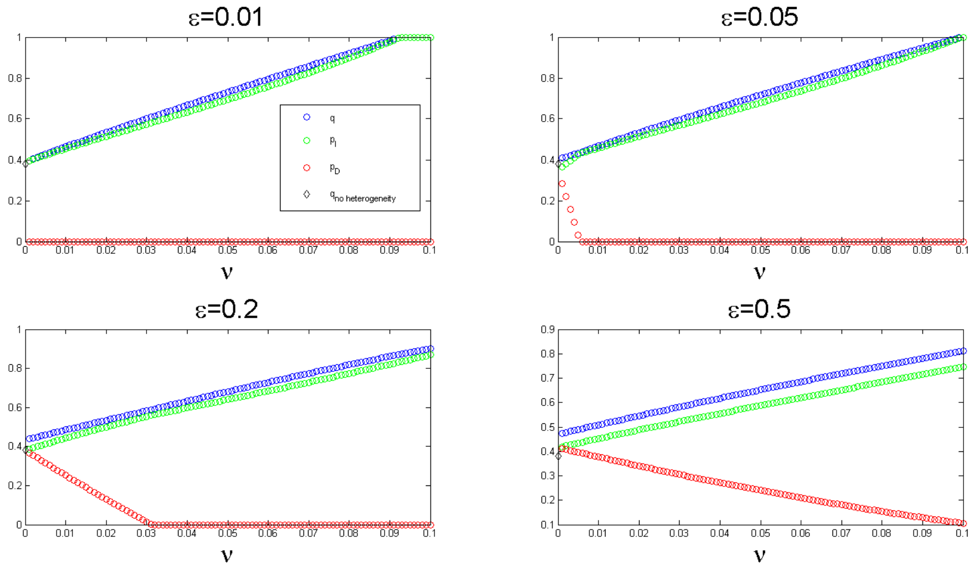


Figure 5: Simulations based on Proposition 5, depicting the effects of ν on q, p_I and p_D, in the stable equilibrium, changing ϵ of the uniform distribution $U(a-\epsilon, a+\epsilon)$. Parameters are $\delta = 0.9$, $a = 0.8$, $b = 0.6$ and $\mu = 0$.

In Section 5.3 we study how the equilibrium in an interior (mixed) equilibrium for a homogeneous (with respect to parameters) population, extends to a pure strategy equilibrium for a heterogeneous population. Proposition 5 provides qualitative answers for the corresponding comparative statics, but we can assess the quantitative effect directly. Looking at the proof of Proposition 5 (and in particular at equation (o) on page 37) we see that the larger is the probability mass of Φ on a_D – i.e. the derivative of the cumulative distribution Φ' computed at a_D – the larger the effects of ν on p_D. We run a set of simple simulations in Matlab to exhibit the quantitative effects on p_D, that is the average behavior of the agents that are offspring of defectors, and on q, the total level of cooperation.[26] Figure 5 depicts the results for the case of a uniform distribution $U(a-\epsilon, a+\epsilon)$ for parameter a. To show that the uniform distribution, which is the most natural to ensure that $a \in [0, b+1]$ (this is the requirements of the underlying game), is not a special case, we also run the same set of simulations under a normal distribution $\mathcal{N}(a, s)$. In this case instead of ϵ we have a parameter s governing

[26] All the simulation codes are available at:
https://github.com/paolopin/CooperationPunishmentImmigration.

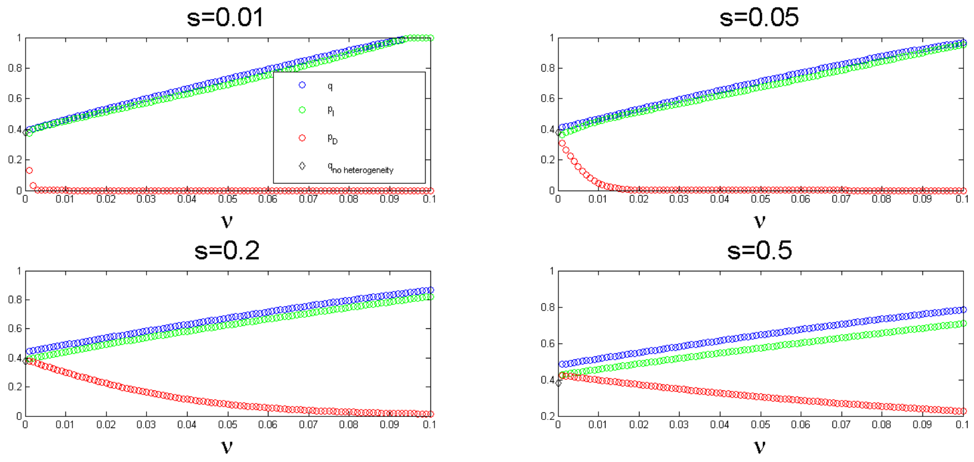


Figure 6: Simulations based on Proposition 5, with a normal distribution $\mathcal{N}(a, s)$, depicting the effects of ν on q, p_I and p_D, in the stable equilibrium, changing standard deviation s. Parameters are $\delta = 0.9$, $a = 0.8$, $b = 0.6$ and $\mu = 0$.

the standard deviation (the standard deviation of the uniform distribution is $\frac{\epsilon}{\sqrt{3}}$), and we obtain in Figure 6 a similar, but smoother, outcome.