Cooperation, Punishment and Immigration*

Paolo Pin† Brian W. Rogers‡

December 2013

Abstract

We study the incentive to cooperate in a nation 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 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 how policies aimed at immigrants can influence the outcomes of existing citizens.

While there is much discussion about the ramifications of immigration, and especially illegal immigration, there is relatively little research measuring the net economic impacts along various dimensions. Taking the United States as our leading example, it is clear that illegal immigration has numerous costs and benefits, but the overall net impact on the American economy is not at all clear.

One of the benefits of immigration is that the increase in labor supply allows domestic resources that are complementary to labor to be used more efficiently. In turn this increases profits.\(^1\) It also lowers prices, raising real income.\(^2\) On the other hand, the increased labor supply pushes wages down, at least in some sectors. Borjas (2003) estimates that between 1980-2000, immigration (both legal and illegal) led to a decrease in average U.S. wages of 3 percent. Although, as Hanson (2007) points out, this calculation ignores some long-run effects whereby the presence of additional labor should affect investments in capital accumulation.\(^3\) Of particular interest to the current study is the estimate of Borjas (2003) that, among high school drop outs, the decrease in wage attributable to immigration was 9 percent.\(^4\)

We develop a framework that allows us to analyze immigration and behavior in an equilibrium context. In order to focus the analysis, 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 undesirable.

\(^1\)This increase is less than one percent. See Hanson (2007). Borjas (2003) estimates this effect at 0.2 percent.
\(^2\)Cortes (2008) reports that a “ten percent increase in the share of low-skilled immigrants in the labor force decreases the price of immigrant-intensive services […] by 2 percent.”
\(^3\)Winegarden and Khor (1991) conclude that illegal immigrants have minimal effect on the unemployment rate of young and minority workers.
\(^4\)Immigrants also contribute to the tax base and demand costly services. According to Hanson (2007), the net fiscal effect 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.
There are two channels through which individuals enter the economy: through birth within the country and through immigration from abroad. Both kinds of agents face 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 set of 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 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. Our results may be summarized as follows.

First, in the absence of expulsion and when individuals have no offspring, we provide a complete characterization of steady state equilibria. If there exists a non-trivial equilibrium, it is unique and such that either all agents cooperate, or agents are indifferent and mix in such a way that many, but not all, agents cooperate. One potentially interesting observation about this characterization is that the state of the economy will exhibit hysterises, in that, starting from an equilibrium with a high level of cooperation, a shock that pushes the economy into the all-defect equilibrium will, even if temporary, leave the economy in the autarky state.

This characterization is descriptive of many large economies, at least in a stylized sense. For instance, a robust empirical regularity seems to be that society consists of high, but less than universal, levels of cooperation, which is a natural outcome in the equilibrium of our framework.

Second, we characterize the impact of expelling immigrants who defect. Increasing the intensity of this policy decreases the defection rate, which is natural, since immigrants pay a higher expected cost for defecting in terms of shortened lifetimes. However, there is an equilibrium effect of this policy whereby the incentives for citizens to defect increase at the institution of the policy. This is so because the indifference of citizens is generated by having a precisely determined proportion of defectors in society according to the payoff parameters of the environment. When immigrants shift towards cooperation, it becomes optimal for some citizens to change their behavior and

\[5\text{Consistently with law in the United States, we assume that all individuals born in the country, whether to citizens or to immigrants, become citizens.}\]
defect instead, returning the economy to equilibrium. 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 develop a rich set of valuable relationships during the course of their lives. When they die, any offspring inherit this social capital at their birth 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. It is in this very simple and natural sense that one can think of our model as deriving a form of cultural transmission. It does not rely on any form of paternalism, or on parents actively asserting influence over their children’s preferences or behaviors.

Third, we enrich the model by allowing for heterogeneity in the preferences of agents. 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 overall behavior in the economy. However, this effect is smaller than would be predicted by a naive analysis that failed to account for equilibrium effects. In particular the net effect on the defection rate is diminished by a substitution effect whereby more citizens find it optimal to defect as immigrants shift to cooperating. We identify these citizens as the class of agents who are adversely affected by the stronger immigration policy. In the model, they are the offspring of agents who are already defecting, which, under one interpretation, is our analogue of the low-skilled labor force. Importantly, under such a policy, upward mobility of these agents is reduced. In fact, it is completely eliminated in a model without heterogeneity in preferences. In the context of our model, 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.

The rest of the paper proceeds as follows. Section 2 discusses the academic literature we build on. Section 3 presents the framework, including our model of population dynamics and our specification of utility functions for agents. Section 4 characterizes equilibria when there is no inheritance and when the government makes no attempt to expel immigrants. Section 5 presents our results on the impact of immigration policy. Section 6 concludes and offers thoughts on how our model and results might be
extended. 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 brings in touch with 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. In particular, 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 for the same discount factors under random re-matching compared to fixed matching, 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 the same result. There are also studies that demonstrate equilibria with highly efficient outcomes when players have discretion over if and when to change partners over time (see Kranton 1996, Ghosh and Ray 1996, Datta 1996, Watson 1999, 2002, and Fujiwara-Greve and Okuno-Fujiwara 2009).

Our setting has the features of endogenous termination of relationships and anonymity. However, our work bears little in common with any of 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.

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

\footnote{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.}
immigration policy, on equilibrium outcomes.

We also contribute to the economic literature investigating the impact of immigration. On this, see the survey of Hanson (2010) 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. 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 (2012) estimate a very different model on behavior and immigration in which citizenship and immigration status is linked to the criminal behavior.

In this context, we study equilibrium effects of a policy: the intervention fails because it does not take into account the effects of its disclosure, i.e., the variation in optimal responses by the agents in the society, due to the acknowledgment that the rules of the game have changed. This is just an example of the classical Lucas (1976) critique against myopic policy interventions.

Finally our paper is related to the economic literature on identity. See Akerlof and Kranton (2000) for a review of this work. There is a connection between our model and the idea of cultural transmission, 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. In our paper there is a similar outcome whereby children take similar actions as their parents. But 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 in order to maximize lifetime payoffs. We look for an equilibrium under a steady state of the population dynamics. This allows us to study the potential consequences of policies aimed at influencing the incentives of immigrants.
3.1 Population dynamics

There is a single society, modeled as a continuum mass of people, that evolves in discrete time. At each period there is a fixed inflow, of mass $\eta$, of agents, which we refer to as *immigrants*. There is also an endogenous mass of offspring of citizens that enter at each period. Every entering agent chooses to be a *cooperator* or a *defector*.7

Agents exit the system through death and by expulsion of immigrant defectors. A proportion $(1 - \delta)$ of agents die at every period, independent of their status. Dying agents have a single offspring with probability $\mu$. A proportion $\nu$ of immigrant defectors are punished by expulsion at every period.

At each period every agent has out-degree $k$. After the interactions of the period take place, any agent can choose to sever any subset of its (in- and out-links). Links are also severed when an agent is expelled, or when it dies and has no offspring. Otherwise the link survives to the next period. Any agent who, through the severance of links at the previous period, has fewer than $k$ out-links, re-matches with new partners who are chosen uniformly at random from the entire population.

We proceed by formalizing this system. At each period $t$ we have a directed network $g_t = (S_t, K_t)$, homogeneous in out–degree $k \in \mathbb{N}$, partitioned, as will be clear below, into cooperators, immigrant defectors and citizens defectors. Denote by $q_t \in [0, 1]$ the fraction of cooperators. The population evolves in the following way during each period:

(a) The set of offspring (defined from the previous period) enter and are added to $S_{t-1}$. Each offspring born to a defector decides with probability $p_{D,t} \in [0, 1]$ whether to be a cooperator or a defector. Each offspring born to a cooperator decides with probability $p_{C,t} \in [0, 1]$ whether to be a cooperator or a defector.

(b) A set of mass $\eta$ of new immigrants enter and are added to $S_{t-1}$. Each of them decides with probability $p_{I,t} \in [0, 1]$ whether to be a cooperator or a defector.

(c) Every entering agent casts $k$ links to agents in the society. Every new agent casts links until it has a total of $k$ out-links. All partners are chosen uniformly at random.

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

(e) A proportion $(1 - \delta) \in (0, 1)$ of $S_t$ is randomly selected to die, independent of status.

---

7The decision is taken only once at birth. It is possible to extend our arguments to situations in which the decision is taken at every period. On this see Appendix A.
(f) 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 its parent, i.e., the same set of links in $K_t$. Every offspring is a citizen, whether or not the parent was an immigrant.

(g) 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).

3.2 Steady state

We analyze a steady state where $S_t = S$, $I_t = I$, $q_t = q$, $p_{I,t} = p_I$, $p_{C,t} = p_C$ and $p_{D,t} = p_D$. We shall also take $p_C = 1$, i.e., offspring of cooperators choose to cooperate, for the main analysis. In Appendix A we show that this is a weak assumption and discuss robustness. Figure 1 represents the flows of individuals through the system.

The first steady-state condition is that of fixed population size, i.e.,

$$\eta = (1 - \delta)(1 - \mu)|S| + |I|\delta\nu.$$ 

Writing the size of the immigrant defector population as

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

we can define $|S|$ in steady state:

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

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

We can also find the size of $|C|$ and $|D|$, which are given by the steady condition

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

\[8\]In principle, we could consider a different model in which the probability $\nu$ of being expelled is independent from the probability $\delta$ of dying. This alternative assumption would change the probability of passing from $I$ 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 $\nu$ into $(\delta + \mu - \delta\mu)\nu$. In this sense, the alternative assumption simply results in a rescaling of $\nu$ that depends on $\delta$ and $\mu$. 

8
The solution is

\[
\begin{align*}
|C| &= \frac{1}{1-\mu} \cdot \frac{1}{1-\mu(pD)} \left( \mu \cdot pD|I| + pI \frac{1-\mu(1-pD)}{1-\delta} \eta \right) \\
|D| &= \frac{1}{1-\mu} \cdot \frac{1}{1-\mu(pD)} \left( \mu(1-\mu(1-pD)-pD)|I| \right)
\end{align*}
\]

A main quantity of interest for our analysis is the proportion \( q = \frac{|C|}{|S|} \) of cooperators in society, given by

\[
q = 1 - \frac{(1-\delta)(1-\mu)(1-pI)}{(1-\mu + \mu pD)(1-\delta + \delta \nu pI)}.
\]  

Equation (1) is one of the main building blocks of our analysis. As a first check it may be noted that, as expected, when \( \mu = 1 \) and/or \( \delta = 1 \) we obtain that \( q = 1 \), and that the partial derivatives of \( q \) with respect to \( pI \), \( pD \) and \( \nu \) are all positive. Also, when \( \mu = 0 \) there is no effect of \( pD \), and conversely. When \( \nu = 0 \) the effect of \( \delta \) disappears.
3.3 Payoffs

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

Each link in $K_t$ represents the play, at period $t$, of the prisoner’s dilemma game given by

<table>
<thead>
<tr>
<th></th>
<th>$C$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>$1,b$</td>
<td>$1-a$</td>
</tr>
<tr>
<td>$D$</td>
<td>$1+a$</td>
<td>$b$</td>
</tr>
</tbody>
</table>

with $a, b > 0$ and $a - b < 1$.

Clearly it is dominant (for both cooperators and defectors) to sever links with defectors and maintain those with (surviving) cooperators.\(^9\) We want to analyze non-trivial steady states where there is not necessarily a clear dominance in the choice to cooperate or defect.

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, we have that $\delta_N = \delta$. Let $n^{out}_{XY}(t)$ 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 birth of the agent.

The number of out-links from cooperators to cooperators is:

\[
\begin{align*}
n^{out}_{CC}(t) &= \delta_N n^{out}_{CC}(t-1) + q(k - \delta_N n^{out}_{CC}(t-1)) \\
&= k \left( q + (1-q)\delta_N n^{out}_{CC}(t-1) \right) \\
&= k q \frac{1 - (\delta_N(1-q))^{t+1}}{1 - \delta_N(1-q)},
\end{align*}
\]

where the last equality is solved recursively setting $n^{out}_{CC}(0) = k q$.

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

The proportion of cooperators that have occupied their position in the network for $t$ periods is $s(t) = (1-\delta_N)\delta_N$.\(^{10}\) We denote the age distribution of defectors by $s_D(t)$.

---

\(^9\)Defectors are indifferent about maintaining in-links from other defectors. But those links are out-links for another defector, and so will be severed. Note that the fact that the $(D,D)$ payoff is zero is not without loss.

\(^{10}\)By this we mean the length of time a given cooperator can trace back his lineage in the pool of cooperators.
We then have that the per-period inflow of links from cooperators and defectors are

\[ r_C = \sum_{t=0}^{\infty} q_s(t) (k - \delta_N n_{CC}^{out}(t - 1)) = k \frac{q (1 - \delta_N^2)}{1 - \delta_N^2 (1 - q)} , \]

\[ r_D = k \sum_{t=0}^{\infty} (1 - q) s_D(t) = k(1 - q) . \]

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

\[ n_{CC}^{in}(t) = k \delta_N n_{CC}^{in}(t - 1) + r_C \]

\[ = r_C \frac{1 - \delta_N^{t+1}}{1 - \delta_N^t} \]

\[ = k \frac{q (1 + \delta_N)}{1 - \delta_N^t} (1 - \delta_N^{t+1}) \]

where the second line is solved explicitly setting \(n_{CC}^{in}(0) = r_C\).

Clearly \(n_{CD}^{in}(t) = n_{DD}^{in}(t) = r_D\) and \(n_{DC}^{in}(t) = r_C\).

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

\[ \pi_C(t) = 1 \cdot (n_{CC}^{out}(t) + n_{CC}^{in}(t)) - b \cdot (n_{CD}^{out}(t) + n_{CD}^{in}(t)) , \]

\[ \pi_D(t) = (1 + a) \cdot (n_{DC}^{out}(t) + n_{DC}^{in}(t)) . \]

There are three situations an agent can be born into: cooperating, defecting as a citizen, and defecting as an immigrant.\(^{11}\) We can write the associated expected lifetime payoffs as

\[ u_C = \sum_{t=0}^{\infty} \delta^t \pi_C(t) \quad (2) \]

\[ u_D = \sum_{t=0}^{\infty} \delta^t \pi_D(t) \quad (3) \]

\[ u_I = \sum_{t=0}^{\infty} (\delta (1 - \nu))^t \pi_D(t) , \quad (4) \]

where the distinction in the two roles of defecting derives from the fact that immigrants, and only immigrants, face the possibility of being expelled.

\(^{11}\)A new cooperator could either be entering the pool of cooperators from scratch, or instead inheriting the position of its parent. The utility \(u_C\) below corresponds to the former, as the latter decision is taken care of from the requirement that \(p_C = 1\).
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

\[
\begin{align*}
u_C &= \left( \frac{k}{1 - \delta} \right) \left( \frac{2q - b(1 - q)(1 - \delta^2)}{1 - \delta^2(1 - q)} - b(1 - q) \right), \\
u_D &= \nu_I = \left( \frac{k}{1 - \delta} \right) \left( \frac{(1 + a)q(1 - \delta^2)}{1 - \delta^2(1 - q)} + (1 + a)q \right).
\end{align*}
\]

An equilibrium 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, and the steady state is consistent with these optimal choices over time. A stable equilibrium 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 with decrease (increase) over time as new agents enter. We now fully characterize the set of stable equilibria.

**Proposition 1.** 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 a stable equilibrium with \( 0 < q < 1 \) if and only if \( a - b < 1 \) and the following condition is satisfied:

\[
b \in \left[ \max \{ a - 1, 0 \}, \frac{2(\delta^4 + a(2 - 3\delta^2 + \delta^4))}{(2 - \delta^2)^2} - \frac{4\sqrt{\delta^2(1 - \delta^2)(a(2 - \delta^2) - \delta^2)}}{(2 - \delta^2)^2} \right],
\]

whenever this interval is defined.

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. We show that \( V(q) \) is concave and thus has at most two zeros. When the intersection is unique and non-degenerate, then \( V(1) > 0 \) and \( q = 1 \) is a stable equilibrium in which all the agents in the population cooperate. This turns out to happen for \( a \) sufficiently small. When the intersections are two, 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. 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
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 $\delta = 0.9$ (yellow for stable ones with $q = 1$ – and this region extends to infinity for any positive $b$ – and light-blue 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.
the other three panels are depicted in this graph, with the corresponding implications for equilibrium properties.

The conditions on payoffs that we introduced in the beginning \((a, b > 0 \text{ and } a - b < 1)\) can, to some extent, be relaxed. That is, 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 inlink 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 are more efficient equilibria where agents alternate between \(C\) and \(D\), maintaining only the profitable links with agents they are synchronized with.

Figure 2 reports a graphical representation of all the inequalities discussed in this section, for the case of \(\delta = 0.9\).

We remark that the analysis of equilibrium, as depicted in Figure 3, displays a hysteresis 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.\(^{12}\) 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 autarky state, as it is a stable equilibrium. In general, whether the economy is in the autarky state or in the non-trivial equilibrium, when the latter exists, is, in part, determined by the historical path of the economy.

\(^{12}\)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 3: Equilibrium correspondence as a function of $a$. The red curve corresponds to autarky. The Solid blue curve corresponds to the stable non-trivial equilibrium, while the dashed blue 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. 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 accommodates inheritance and expulsion.

**Lemma 2.** In the general case, with $\nu > 0$ and $\mu > 0$:
- $u_C$ is independent of $a$ and $\nu$, and linearly decreasing in $b$;
- $u_D$ is independent of $b$ and $\nu$, and linearly increasing in $a$;
- and finally $u_I$ is independent of $b$, linearly increasing in $a$ and decreasing in $\nu$.

This 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 dependences are easy to check.

We now show that the effects of changes to the payoff parameters produce the intuitive results in equilibrium.
Proposition 3. Consider a set of parameters for which there is a stable equilibrium with \(0 < q^* < 1\). The marginal effects of an increase in parameters \(a\) and \(b\) on \(q^*\) are negative. The marginal effect of an increase in \(\nu\) on \(q^*\) is positive.

For the proof, as with Proposition 1, we consider a function \(V(q)\) that is proportional to \(u_C(q) - u_D(q)\). 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.\(^{13}\)

The comparative statics with respect to \(\delta\) 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 \(\delta\) 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 \(0 < p_I < 1\).\(^{14}\) Simplifying equation (1) accordingly, we have

\[
q = 1 - \frac{(1 - \delta)(1 - p_I)}{1 - \delta + \delta p_I}
\]

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

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

\(^{13}\)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.

\(^{14}\)This follows from equation (1). If \(p_I = 1\) then \(q = 1\). If \(p_I = 0\) then \(q = 0\).
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. 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 \( \nu \).

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

\[ u_D(a) = u_C, \quad (7) \]
\[ u_I(a) = u_C. \quad (8) \]

Equations (7) and (8) can be solved for a unique value of \( a \) each, to yield thresholds \( a_D \) and \( a_I \), below which an immigrant and a citizen, respectively, prefer to cooperate.\(^{15}\) These, in turn, generate the probabilities of interest via \( p_D = \Phi(a_D) \) and, similarly, \( p_I = \Phi(a_I) \).

The following result generalizes Proposition 4 in this new context.

**Proposition 5.** Fix parameters such that there is an interior equilibrium for \( \nu = 0 \). This characterizes a unique value of \( p_D \). If this \( p_D \) is strictly positive, then for any \( \nu \geq 0 \) we have that \( \frac{dp_D}{d\nu} < 0 \).

The proof relies on the implicit function theorem, which is applied to an appropriately defined system of three implicit equations characterizing equilibrium conditions. As in the proof of Proposition 3, we make use of the assumption of being in a stable equilibrium. From the formulas in the proof we obtain that the larger is the probability mass on \( a_D \) – i.e. the derivative of the cumulative distribution \( \Phi' \) computed at \( a_D \) – the larger the effect of \( \nu \) on \( p_D \). 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 how the flow of immigrants 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

\(^{15}\)There is always a unique solution for \( q > 0 \), which is the case of interest.
immigrants is not an optimal, or even desirable, policy instrument. This conclusion derives from the fact that behaviors in society balance the incentives between cooperation and defection. In equilibrium, there is a sense in which there is a natural 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 situation close to the original equilibrium. On the other hand, if the payoffs to defection are reduced by some means that affects all agents similarly, the economy can more efficiently be moved in the desired direction.

Modeling the strategic interactions between individuals through a prisoners’ dilemma has the benefit of capturing important elements of social relationships. However, the analysis is necessarily abstract as a result. One interpretation of our model is of a choice between participation in the formal economy versus black market activities. In this setting, our conclusion would be that, as immigrants are driven from the black market by the threat of expulsion, citizens shift into such behaviors. Further work could explore this direction through a more explicit modeling of labor market choices including wages.

Our framework has the important advantage of being tractable. As always, this comes at some cost of richness. One direction that seems particularly interesting to explore is to understand the differences between different forms of immigration. Immigrants enter both legally and illegally (and there are much finer divisions one may wish to consider) and have different choices, incentives, and opportunities as a result. Any meaningful policy would have to take such heterogeneity seriously.

A related final observation is that there are some important ingredients of our analysis that are treated exogenously. Perhaps the most prominent example is $\eta$, the rate of immigration. Clearly, this should generally be sensitive to the outcomes that can be expected upon immigration and, in particular, to the policy instrument we study, $\nu$, the expulsion rate for immigrant defectors. Were this to be endogenized, clearly the effect of $\nu$ on $\eta$ would be decreasing. In turn, this would magnify the effect of $\nu$ on the level of cooperation, but would also come at the cost of reducing the rate of immigrant cooperators.

Acknowledgements

We thank James Dana, Osborne Jackson and Myrna Wooders, and seminar participants at the 2013 WIN Workshop, Northeastern University and Vanderbilt University for comments. 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


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 from the expectation of accumulating social capital in the form of relationships with other cooperators. As this valuable network of relationships 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.

To this end we provide the following result. Under a certain mildly restrictive condition, we prove that, in fact, every offspring of a cooperator prefers to cooperate.

We introduce the following condition.

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

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

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 provide the result of this section.

**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, it is 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.

\[\square\]
Appendix B  Proofs

Proof of Proposition 1 (page 12): The structure of equilibria can be derived from the properties of the following function

\[ V(q) \equiv \left( \frac{1 - \delta}{k} \right) u_C - u_D = \frac{2q - (1 - \delta^2)(1 + a - b)q}{1 - \delta^2(1 - q)} - (1 + a - b)q. \]

That is obtained from (5) and (6). If we call \( f(q) \equiv (1 + a - b)q \) and \( g(q) \equiv 1 - \delta^2(1 - q) \), we can write

\[ V(q) \equiv \frac{2q - (1 - \delta^2)f(q)}{g(q)} - f(q). \]

The following statements hold.

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

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

\[ a \leq \frac{\delta^2}{2 - \delta^2}; \]

(C) The second derivative of \( V(q) \) has the same sign as (given that \( g(q) \) is always positive)

\[
V''(q) \cdot (g(q))^3 = 2 \left( 2q - (1 - \delta^2)f(q) \right) (g'(q))^2 - 2g(q)g'(q) \left( 2 - (1 - \delta^2)f'(q) \right)
\]

\[ = -2\delta^2(1 - \delta^2) \left( 1 + b + \delta^2 - a(1 - \delta^2) \right) \quad \text{(a)} \]

which is always negative as long as \( a < 1 + b \), so \( V(q) \) is always concave in the region of parameters that we are interested in;

(D) We satisfy first order conditions (i.e. \( V'(q) = 0 \)) when

\[ q^* = 1 - \frac{1}{\delta^2} \pm \sqrt{(1 - \delta^2)(1 + a - b) \left( 1 + b + \delta^2 - a(1 - \delta^2) \right)} \div \delta^2(1 + a - b). \quad \text{(b)} \]

Three things should be noted:

(i) \( q^* \) is defined in the real numbers if and only if \((1 + a - b)(1 + b + \delta^2 - a(1 - \delta^2)) \equiv \Delta \geq 0.\)

(ii) As \( 1 - \frac{1}{\delta^2} < 0 \), then if there is a \( q > 0 \) that satisfies first order conditions, it is unique, and its form depends on the sign of \( 1 + a - b \). We will call \( q^* \) just this positive-valued solution from (b), if it exists.
(iii) If $\Delta \geq 0$, then $q^* \geq 0$ if and only if
\[
\frac{\sqrt{(1 - \delta^2)\Delta}}{|1 + a - b|} \geq 1 - \delta^2 ,
\]
which is equivalent to say that
\[
|1 - a + b + \delta^2(1 + a)| \geq (1 - \delta^2)|1 + a - b| .
\]
If $1 + a - b > 0$ and $\Delta > 0$, this is equivalent to say that
\[
b \geq 2 \left(1 + a - \frac{2 + a}{2 - \delta^2}\right) \equiv \bar{b}(a, \delta) ,
\]
otherwise we need the opposite inequality.

(iv) If $\Delta \geq 0$, then $q^* \leq 1$ if and only if
\[
\frac{\sqrt{(1 - \delta^2)\Delta}}{|1 + a - b|} \leq 1 ,
\]
which is equivalent to say that
\[
(1 - \delta^2)|1 - a + b + \delta^2(1 + a)| \leq |1 + a - b| .
\]
If $1 + a - b > 0$ and $\Delta > 0$, this is equivalent to say that
\[
b \leq -\delta^2(1 + a) + \frac{2(a + \delta^2)}{2 - \delta^2} \equiv \bar{b}(a, \delta) ,
\]
otherwise we need the opposite inequality.

(E) Let us check when it is that both inequalities (c) and (d) can be satisfied, when $\Delta > 0$. We have that $\bar{b}(0, \delta) = \frac{\delta^4}{2 - \delta^2} > 0$, moreover
\[
0 < \frac{\partial}{\partial a}b(a, \delta) = \frac{2 - 2\delta^2}{2 - \delta^2} < \frac{2 - 2\delta^2 + \delta^4}{2 - \delta^2} = \frac{\partial}{\partial a} \bar{b}(a, \delta) ,
\]
and then $\bar{b}(a, \delta) = \bar{b}(a, \delta)$ for $a = -1 - \frac{2}{\delta^2} < 0$.
This means that, if $1 + a - b > 0$ and $\Delta > 0$, then $q^* \in [0, 1]$ if and only if $b \in \left[\max\{\bar{b}(a, \delta), 0\}, \bar{b}(a, \delta)\right]$, where this interval is always non–empty.
If instead $1 + a - b < 0$, then there is no $b > 0$ for which $q^* \in [0, 1]$.

\[16\]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 $\Delta < 0$. With this other argument however we prove that the condition $1 + a - b < 0$ must be excluded also without considering $a - b < 1$. This argument also proves that when $1 + a - b > 0$ and $a - b < 1$, then $\Delta > 0$.
(F) We have that, if $\Delta \geq 0$, then
\[ V(q^*) = -b + \frac{2}{\delta^2} - \frac{2 \sqrt{(1 - \delta^2)\Delta}}{\delta^2}. \]

(G) If $\Delta \geq 0$, then $V(q^*) > 0$ if and only if
\[ \sqrt{\Delta} \leq \frac{1 - b\delta^2/2}{\sqrt{1 - \delta^2}}, \]
which is equivalent to requiring both that $b < 2/\delta^2$ and (taking squares it becomes a second order polynomial in $b$)
\[ b \not\in \left(\hat{b}(a, \delta) - \frac{4\sqrt{\delta^2(1 - \delta^2)(a(2 - \delta^2) - \delta^2)}}{(2 - \delta^2)^2}, \hat{b}(a, \delta) + \frac{4\sqrt{\delta^2(1 - \delta^2)(a(2 - \delta^2) - \delta^2)}}{(2 - \delta^2)^2}\right) \]
where we have defined $\hat{b}(a, \delta) \equiv \frac{2(\delta^3 + a(2 - 3\delta^2) + \delta^4)}{(2 - \delta^2)^2}$.

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 extrema of this interval reduces both to $\hat{b}(\frac{\delta^2}{2 - \delta^2}, \delta) = b(\frac{\delta^2}{2 - \delta^2}, \delta)$, where the latter is defined in (d).
(iii) $\frac{\partial}{\partial a} \hat{b}(a, \delta) < \frac{\partial}{\partial a} \hat{b}(a, \delta)$, where the latter is computed in (e).
(iv) We always have $1 + a > \hat{b}(a, \delta)$ because $\hat{b}(0, \delta) < 1$ and $\frac{\partial}{\partial a} \hat{b}(a, \delta) < 1$.
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
\[ b \leq \hat{b}(a, \delta) - \frac{4\sqrt{\delta^2(1 - \delta^2)(a(2 - \delta^2) - \delta^2)}}{(2 - \delta^2)^2}. \]

(H) Both $\hat{b}(a, \delta) - \frac{4\sqrt{\delta^2(1 - \delta^2)(a(2 - \delta^2) - \delta^2)}}{(2 - \delta^2)^2}$ and $\hat{b}(a, \delta)$ equal 0 if and only if $a = \frac{\delta^2}{1 - \delta^2}$.
Moreover, for every $a > \frac{\delta^2}{2 - \delta^2}$, and then also for $a > \frac{\delta^2}{1 - \delta^2}$ we have that
\[ \frac{\partial}{\partial a} \left( \hat{b}(a, \delta) - \frac{4\sqrt{\delta^2(1 - \delta^2)(a(2 - \delta^2) - \delta^2)}}{(2 - \delta^2)^2} \right) < \frac{\partial}{\partial a} \hat{b}(a, \delta) < \frac{\partial}{\partial a} b(a, \delta). \]
This means three things:
(i) Condition (g) can never be satisfied together with condition (d), if $a > \frac{\delta^2}{1 - \delta^2}$.
(ii) Condition (d) is never binding, because whenever $\hat{b}(a, \delta)$ is positive, then condition (g) is not satisfied.
(iii) So, the only lower bound that matters for $b$ is that $b \geq a - 1$.  

24
\[ \delta = 0.9 \]

\[ 1 + a - b = 0 \]

Full cooperation equilibrium exists

\[ \frac{\delta^2}{2 - \delta^2} \]

Interior stable equilibrium exists

Below this dashed line CC is no more efficient

Above this line \( V(q) \) is concave

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

Proof of Lemma 2 (page 15): We compute (2)–(4) with \( \delta_n = \delta + (1 - \delta)\mu \), to yield

\[
\begin{align*}
        u_C &= \frac{(b+1)q}{1-\delta(1-q)(\delta+\mu-\mu\delta)^2} - b(2-q) + \frac{q^2(\delta^2(\mu-1)-\delta\mu-1)}{(q+\delta^2-1)((q-1)(\delta+\mu-\mu\delta)^2+1)} + \frac{q(\delta+1)}{(q+\delta^2-1)(\delta(\mu-1)-1)} (h) \\
        u_D &= \frac{(a+1)q(1+\frac{1-\delta+\mu-\mu\delta}{2-q}}{1-\delta} , \quad (i) \\
        u_I &= \frac{(a+1)q(1+\frac{1-\delta+\mu-\mu\delta}{2-q}}{1-\delta(1-\nu)} . \quad (j)
\end{align*}
\]

The dependence of these three functions with respect to \( a, b \) and \( \nu \) are evident. To see that \( \frac{\partial u_C}{\partial b} < 0 \), note that \( \frac{q(\delta^2(\mu-1)-\delta\mu-1)}{(q+\delta^2-1)((q-1)(\delta+\mu-\mu\delta)^2+1)} \leq 1 \), 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 \( \frac{\partial u_C}{\partial a} = 0, \frac{\partial u_C}{\partial b} < 0, \frac{\partial u_C}{\partial \nu} = 0, \frac{\partial u_I}{\partial a} = 0, \frac{\partial u_I}{\partial b} = 0, \frac{\partial u_I}{\partial \nu} > 0 \). By the implicit function theorem:

\[
\begin{align*}
        dq^* &= -\frac{\partial (u_C - u_I)/\partial a}{\partial (u_C - u_I)/\partial q} = -\frac{\partial u_I/\partial a}{\partial (u_C - u_I)/\partial q} < 0 , \\
        dq^* &= \frac{\partial (u_C - u_I)/\partial b}{\partial (u_C - u_I)/\partial q} = \frac{\partial u_C/\partial b}{\partial (u_C - u_I)/\partial q} < 0 , \\
        dq^* &= \frac{\partial (u_C - u_I)/\partial \nu}{\partial (u_C - u_I)/\partial q} = \frac{\partial u_I/\partial \nu}{\partial (u_C - u_I)/\partial q} > 0 . \quad \square
\end{align*}
\]

Proof of Proposition 3 (page 15): 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 \( \frac{\partial u_C}{\partial a} = 0, \frac{\partial u_C}{\partial b} < 0, \frac{\partial u_C}{\partial \nu} = 0, \frac{\partial u_I}{\partial a} = 0, \frac{\partial u_I}{\partial b} = 0, \frac{\partial u_I}{\partial \nu} > 0 \). By the implicit function theorem:

\[
\begin{align*}
        dq^* &= \frac{\partial (u_C - u_I)/\partial a}{\partial (u_C - u_I)/\partial q} = -\frac{\partial u_I/\partial a}{\partial (u_C - u_I)/\partial q} < 0 , \\
        dq^* &= \frac{\partial (u_C - u_I)/\partial b}{\partial (u_C - u_I)/\partial q} = \frac{\partial u_C/\partial b}{\partial (u_C - u_I)/\partial q} < 0 , \\
        dq^* &= \frac{\partial (u_C - u_I)/\partial \nu}{\partial (u_C - u_I)/\partial q} = \frac{\partial u_I/\partial \nu}{\partial (u_C - u_I)/\partial q} > 0 . \quad \square
\end{align*}
\]

Proof of Proposition 4 (page 16): When \( \nu = 0 \) equation (1) becomes simply

\[
q = 1 - \frac{(1 - \mu)(1 - p_I)}{1 - \mu + \mu p_D} ,
\]

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

\[
p_I = q - (1 - q)\frac{\mu}{1 - \mu} p_D .
\]
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 $p_I = \frac{(1-\delta)q}{(1-\delta)+\delta(1-\delta)}$. 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.

**Proof of Proposition 5 (page 17):** 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 $q$:

\begin{align*}
p_I - \Phi(a_I(b, q, \mu, \delta, \nu)) &= 0, \\
p_D - \Phi(a_D(b, q, \mu, \delta)) &= 0, \\
q - F(\mu, \delta, \nu, p_I, p_D) &= 0,
\end{align*}

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 $\Phi$ 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 $\Phi$. 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-Collel et al. 1995) to compute the marginal effects of $\nu$ on the
endogenous variables:

\[
D_{\nu} \begin{pmatrix} p_I \\ p_D \\ q \end{pmatrix} = - \begin{bmatrix}
\frac{\partial E_g(k)}{\partial p_I} & \frac{\partial E_g(k)}{\partial p_D} & \frac{\partial E_g(k)}{\partial q} \\
\frac{\partial E_g(l)}{\partial p_I} & \frac{\partial E_g(l)}{\partial p_D} & \frac{\partial E_g(l)}{\partial q} \\
\frac{\partial E_g(m)}{\partial p_I} & \frac{\partial E_g(m)}{\partial p_D} & \frac{\partial E_g(m)}{\partial q}
\end{bmatrix}^{-1} D_{\nu} \begin{pmatrix} E_g(k) \\ E_g(l) \\ E_g(m) \end{pmatrix}.
\]

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 \(\Phi\)) 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

\[
D_{\nu} \begin{pmatrix} p_I \\ p_D \\ q \end{pmatrix} = - \begin{bmatrix}
\frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_I} & \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_D} & \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial q} \\
\frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_I} & \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_D} & \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial q} \\
\frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_I} & \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_D} & \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial q}
\end{bmatrix}^{-1} D_{\nu} \begin{pmatrix} p_I \\ p_D \\ q \end{pmatrix}.
\]

In the last derivation we have used the fact that

\[
\begin{pmatrix}
1 & 0 & \alpha \\
0 & 1 & \beta \\
\gamma & \delta & 1
\end{pmatrix}^{-1} = \frac{1}{1 - \alpha \gamma - \beta \delta} \begin{pmatrix}
1 - \beta \delta & \alpha \delta - \alpha \\
\beta \gamma - 1 & \alpha \gamma - \beta \\
-\gamma & -\delta & 1
\end{pmatrix},
\]

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

\[
\frac{dp_D}{d\nu} = \frac{\Delta_{Dq} \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_I} + \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial q}}{1 - \Delta_{Iq} \cdot \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_I} - \Delta_{Dq} \cdot \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_D}}.
\]

This quantity is always non–positive as \(1 - \Delta_{Iq} \cdot \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_I} - \Delta_{Dq} \cdot \frac{\partial F(\mu,\delta,\nu,p_I,q)}{\partial p_D} \geq 1\), and \(\frac{\partial F}{\partial \nu} > 0\). Finally, 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 equilibrium with \(p_D > 0\), and \(\nu \geq 0\), it is always the case that \(\frac{dp_D}{d\nu} < 0\).
Figure 5: Simulations based on Proposition 5, depicting the effects of $\nu$ on $q$, $p_I$ and $p_D$, in the stable equilibrium, changing $\epsilon$ of the uniform distribution $U(a - \epsilon, a + \epsilon)$.

Figure 6: Simulations based on Proposition 5, with a normal distribution $\mathcal{N}(a, s)$, depicting the effects of $\nu$ on $q$, $p_I$ and $p_D$, in the stable equilibrium, changing standard deviation $s$.

of the cumulative distribution $\Phi'$ computed at $a_D$ – the larger the effects of $\nu$ 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 $q$. 

29
the total level of cooperation.\footnote{All the simulation codes are available at: http://www.econ-pol.unisi.it/paolopin/WP/matlab_codes_PinRogers.zip.} 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 $\epsilon$ we have a parameter $s$ governing 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.