The Interplay of Cultural Intolerance and Action-Assortativity for the Emergence of Cooperation and Homophily

Ennio Bilancini* Leonardo Boncinelli† Jiabin Wu‡

May 20, 2020

Abstract

This paper investigates the emergence of cooperation in a heterogeneous population that is divided into two cultural groups. Agents are randomly matched in pairs to engage in a prisoner dilemma. The matching process is *assortative in actions*, that is, cooperators are more likely to be matched with cooperators, defectors are more likely to be matched with defectors. Agents exhibit a form of cultural intolerance: when two agents of different cultures are matched, they suffer a cost due to their cultural differences. We find that when cultural intolerance is sufficiently strong, homophily emerges together with perfect correlation between culture and behavior: all agents from one cultural group cooperate, while all agents from the other cultural group defect, and interactions among agents within the same cultural group are more frequent. The relation between cultural intolerance and societal welfare is non-monotonic. In particular, stronger cultural intolerance can increase cooperation when action-assortativity is weak, while it can increase defection when action-assortativity is strong. Moreover, everyone cooperating does not necessarily maximize total welfare unless cultural intolerance can be made sufficiently weak.

JEL classification code: C72; C73; Z10.

Keywords: Cooperation; prisoner dilemma; cultural intolerance; action-assortativity; homophily.

*Dipartimento di Economia "Marco Biagi", Università degli Studi di Modena e Reggio Emilia, Viale Berengario 51, 43 ovest, 41121 Modena, Italia. Tel.: +39 059 205 6843, fax: +39 059 205 6947, email: ennio.bilancini@unimore.it.

†Dipartimento di Scienze per l'Economia e l'Impresa, Università degli Studi di Firenze, Via delle Pandette 9, 50127 Firenze, Italia. Tel.: +39 055 2759578, fax: +39 055 2759910, email: leonardo.boncinelli@unifi.it.

‡Department of Economics, University of Oregon, 1285 University of Oregon, Eugene, OR, USA. Tel: +1 (541) 346-5778, email: jwu5@uoregon.edu.

1 Introduction

It is a matter of fact that societies, in almost every age and place, are comprised of groups which differ between each other on a cultural basis. Often these differences – think, for instance, of language and religion – entail a cost that individuals suffer when interactions occur between members of different groups. The literature studying under which conditions cooperation can emerge in societies has so far given little consideration to the role of costly interactions between different cultural groups.[1] This paper attempts to address this issue.

1.1 Modeling background

We consider a heterogeneous population of agents who are randomly matched in pairs to engage in a prisoner dilemma. Each agent carries one of two different cultural types and they are divided into two cultural groups correspondingly. Each cultural type is substantiated by a set of identity traits, like language and religion, which are essential to the identity of a culture (Akerlof and Kranton, 2000, 2005). At the same time, these traits originate a cost when members of different cultural groups interact with each other: think of communication costs due to different languages, or coordination costs due to differences in work times for religious habits (e.g., Muslims observe Ramadan, Jews rest on Saturday, Christians rest on Sunday). Such a cost can also be psychological.[2] The cost of cross-cultural interactions can be considered as a measure of *cultural intolerance*.

Besides their identity traits, agents also carry auxiliary traits that characterize their actions, i.e., cooperate or defect in the prisoner dilemma. These auxiliary traits may have a cultural nature as well, but they do not convey an identity and, hence, they do not trigger cultural intolerance.

The matching process is not uniformly random, but exhibits *action-assortativity*, which describes people's general tendency to interact with those who act like them more often than with those who behave differently.[3] In our model, action-assortativity implies that

[1] The importance of cultural factors for the evolution of cooperation has been studied, among others, by Henrich and Boyd (2001), Boyd et al. (2003), Henrich (2004), Boyd and Richerson (2009), and Boyd et al. (2011). None of these or similar studies, however, consider the cost of interactions due to cultural mismatch.

[2] Starting with the seminal work of Becker (1957) on taste-based discrimination, numerous researches have found that mistrust, animosity and negative attitude across cultural groups have significant impact on trade and investment (Guiso et al., 2009, Michaels and Zhi, 2010, Fisman et al., 2014), on labor market outcomes (Becker, 1993, Bertrand and Mullainathan, 2004, Bandiera et al., 2009), on financial activities such as mergers and bank loans (Giannetti and Yafeh, 2012, Ahern et al., 2015).

[3] In social psychology, social matching theory (Walster et al., 1966) argues that people tend to form

cooperators are more likely to be matched with cooperators, defectors are more likely to be matched with defectors.

Action-assortativity can emerge as a result of repeated interactions in which agents can voluntarily form and break partnerships conditioned on their opponents' actions but not cultural types. Such a context is particular relevant in today's world where many countries enforce anti-discrimination laws to prohibit the offer of business/education/housing/employment opportunities based on culture, ethnicity, religion and race.[4,5]

Action-assortativity can also be the outcome of a meritocratic institution, which targets to promote good behavior independent of participants' backgrounds. Meritocratic institutions have been adopted since early human civilizations. Selection of officials and councilmen (the Chinese civil service exam), access to education (school-admission systems), honorary circles, reward and promotion schemes within a firm can all be considered as meritocratic institutions.[6]

Changes in traits such as preferences, customs and faiths usually take place across generations over time, while actions evolve over a much shorter time horizon. Since the focus of this paper is on the evolution of cooperation but not on the evolution of culture, we focus on a time horizon in which agents' identity traits are fixed while their auxiliary traits evolve. At the end of the paper, we briefly discuss what is likely to happen in the longer time horizon in which identity traits also evolve.

1.2 Main contribution

The central question we investigate is whether cultural intolerance between groups is relevant for cooperation. At a first glance, one may think that cultural intolerance does not affect, *per se*, the relative advantage of cooperation over defection, and hence it cannot have any effect on the evolution of cooperation. If this is the case, cultural intolerance simply reduces individual payoffs depending on the frequency of type-mismatches (i.e., the frequency of

successful partnerships with those who share similar levels of social desirability.

[4]Intuitively, one may think that observability of identity traits would bias the matching process. However, given a context with anti-discrimination laws, the fact that identity traits are *more* easily observable actually makes them *less* exploitable as reasons for terminating partnerships.

[5]Rivas (2013) provides a convincing mechanism that gives rise to action-assortativity. In his model, agents will tend to stay matched longer when the match provides a higher payoff with the result that, even if the match is initially random, cooperators will be more likely to interact with cooperators and defectors with defectors. When applied to our framework, such mechanism would lead also to type-assortativity, but only in the absence of anti-discrimination laws.

[6]See the discussions in Nax et al. (2014) and Nax et al. (2015).

matches between agents belonging to different cultural groups) due to the costs associated with cross-cultural interactions; therefore, interventions aimed at reducing such costs are clearly good for societal interests. However, it turns out that it is not necessarily the case if we allow for action-assortativity in the matching process.

We find that action-assortativity plays a unique role in the situation we consider: given action-assortativity, cooperation and defection can work as instruments to avoid type-mismatches, and costly cross-cultural interactions can be, to some extent, beneficial to a society. More specifically, we show that when cultural intolerance is sufficiently strong, *homophily* emerges together with perfect correlation between culture and behavior: all agents from one cultural group cooperate, while all agents from the other cultural group defect, and interactions among agents within the same cultural group are more frequent. We call such states *type-monomorphic*.[7]

We stress that the result crucially depends on the joint presence of action-assortativity and cultural intolerance. Because of action-assortativity, cultural intolerance provides agents an incentive to conform to the action that is mostly played by their own group members – as it helps them to avoid costly type-mismatches. The magnitude of such an incentive depends on both the degree of action-assortativity and the degree of segregation of the two cultural groups in actions. When the two cultural groups are sufficiently segregated in actions, that is, the majority of one group cooperates, while the majority of the other group defects, action-assortativity increases the cost of behaving differently from the majority of one's own group; this effect lessens with a smaller degree of assortativity and a less pronounced segregation, but it vanishes only if the degree of assortativity goes to zero or if segregation is nil.

Our result not only offers a novel mechanism that accounts for homophily without assuming a priori that the matching process is biased towards cultural types; more importantly, it provides an evolutionary explanation for the phenomenon of segregation and stratification between cultural groups even in countries that have exercised anti-discrimination laws for decades. In Section 6.2, we discuss the empirical relevance of our result in more details.

[7]The tendency to interact with people sharing the same cultural trait is often called homophily. The term homophily is typically used to indicate, rather than an underlying preference trait, the behavior that results from it. Homophily is commonly observed in human societies along many cultural or ethnic dimensions including gender, language, religion, dress and origin, and it has been intensively studied in sociology and economics (see, among others, McPherson et al., 2001, Ruef et al., 2003, Currarini et al., 2009, 2010, Bramoullé et al., 2012, Currarini et al., 2016).

1.3 Welfare implications

Our analysis can have important implications in terms of welfare, which are possibly relevant for policy. Obviously, cooperation without any cultural intolerance is the first best outcome from a societal point of view. Hence, when action-assortativity is sufficiently strong such that cooperation by everybody can be achieved in the absence of cultural intolerance (see Bergstrom, 2003), any degree of cultural intolerance is unambiguously detrimental because it reduces mutual benefits in case of cross-cultural interactions. In addition, sufficiently strong cultural intolerance may further reduce cooperation by making the state where everybody cooperates unsustainable, leading to a type-monomorphic state where only part of the population cooperates.

However, in reality, cultural intolerance cannot be reduced to zero. Under such a circumstance, it is not necessarily optimal to reduce cultural intolerance as much as possible. In fact, in the presence of cultural intolerance, a type-monomorphic state where only the larger cultural group cooperates can entail a larger total surplus than the monomorphic state where everybody cooperates. Intuitively, the reduction of benefits from cooperation (due to the fact that part of the population defects to avoid type-mismatches) can be more than compensated by the reduction in the frequency of costly type-mismatches (thanks to the perfect correlation between action and culture).

Furthermore, when action-assortativity is sufficiently weak such that all cultures defect in the absence of cultural intolerance, a positive degree of cultural intolerance may benefit the society by increasing cooperation.

Summing up, the first best policy is to eradicate cultural intolerance whenever it is possible and as long as this does not compromise full cooperation. If such policy cannot be implemented because cultural intolerance cannot be eradicated or because action-assortativity remains too weak, then the second best policy can take quite different forms. When there is little cooperation in all cultures, cultural intolerance can be exploited to induce cooperation in the majoritarian culture, either by favoring coordination on cooperation for the majority only (if cultural intolerance is strong enough to sustain a type-monomorphic equilibrium) or by actively strengthening action-assortativity (which magnifies the benefits of cultural segregation in terms of actions). Surprisingly, the second best policy can also entail strengthening cultural intolerance in an attempt to move a society from a monomorphic state to a type-monomorphic one. To be clear, this latter case does not fit any real situation in which cultural intolerance takes the form of hatred towards other cultures, as such kind of cultural intolerance evidently would induce huge social losses. It may fit, however, cases in which cultural intolerance is mostly a coordination cost, like in the case of language or customs.

1.4 Beyond the baseline model

With the aim of investigating if, and to what extent, our findings about the interplay between cultural intolerance and action-assortativity are affected by other elements that can reasonably play a role, we develop three extensions of the baseline model.

First, we consider the existence of asymmetries in the degrees of cultural intolerance (as suggested by Bisin et al., 2004), with agents of one type suffering more than the others when type-mismatches occur. We obtain that our results are qualitatively maintained, with the additional insight that an increase in the degree of action-assortativity can make the system switch from a state where the majority cooperates and the minority defects to a state where the majority defects and the minority cooperates. In terms of welfare, this generates a non-monotonic pattern.

Second, we consider the possibility of assortativity in identity traits (type-assortativity). In this case, agents from the same cultural group have a higher probability to be matched with their own group members. Type-assortativity can also be regarded as some form of homophily. However, it is not attributed to the cost of cross-cultural interactions, but rather to geographic, cultural, linguistic and socioeconomic characteristics that often constraint agents' interactions within their vicinities (as suggested by Alger and Weibull, 2016). We find that allowing for some degree of type-assortativity does not change our main results qualitatively. Moreover, type-assortativity alone cannot account for the phenomenon of perfect correlation between cultures and behavior even when cultural intolerance is present, because cooperation and defection can no longer serve as instruments to help the agents to avoid type-mismatches. This demonstrates the importance of action-assortativity in our model.

Finally, we consider the case where action-assortativity is state-dependent, that is, the likelihood of assortative matching in actions depends on the fraction of the population that cooperates. We show that the introduction of state-dependent assortativity does not affect the substance of our results, although the analysis becomes more complex inducing us to focus on sufficient conditions for evolutionary stability. As an example, we also provide specific results regarding the stranger-in-the-night matching process (Bergstrom, 2013) which gives rise to state-dependent assortativity.

1.5 Related literature

The works most closely related to our study are perhaps Bergstrom (2003) and Bergstrom (2013), who consider matching assortativity and show that, when assortativity is in actions,

it can crucially allow for the evolution of cooperation. More precisely, we follow the idea developed in Bergstrom (2003) who defines the *index of assortativity* of a matching process between two cooperators (and similarly for two defectors) as the difference between the probability that a cooperator is matched with another cooperator and the probability that a defector is matched with a cooperator. Importantly, Bergstrom (2013) also considers a state-dependent index of assortativity, exploring in particular the so called stranger-in-the-night matching process. Basically, we add on Bergstrom's models by introducing two exogenous cultures (in the form of two distinct types) and allowing for cultural intolerance.[8]

Alger (2010) and Alger and Weibull (2010, 2012) apply the index of assortativity not to actions but to types, in order to study the evolution of preferences for altruism.[9] The work of Alger and Weibull (2013) is also related to the current paper. They consider a heterogeneous population in which agents with different cultural types carry different preference traits and they are matched assortatively according to their types. They show that moral preferences (i.e., agents whose preferences attach some extra value to the act of cooperating), and hence cooperation, can spread in the society. This is so because sufficiently strong type-assortativity allows agents of the moral type to internalize part of the benefits of cooperation, so that they obtain a higher payoff than the selfish type agents. One important difference between our approach and theirs is that we introduce the cost of cross-cultural interactions. Another important difference is that we do not allow evolution on types, as we focus on a shorter time horizon – i.e., type-related preferences are kept fixed in our analysis.[10]

The coexistence of cooperators and defectors together with some form of separation[11] between agents that adopt different actions has been already obtained theoretically in Bilancini and Boncinelli (2009), where the possibility to refuse to interact with those who defected in

[8]The effects of cultural intolerance have been studied also in the context of social coordination between risk-dominant and payoff-dominant conventions (Bilancini and Boncinelli, 2017). The main finding is that, if cultural intolerance is strong enough, then both conventions survive in the long run, with perfect correlation between culture and convention.

[9]Lehmann et al. (2015) study these and related issues in a biological model where the population of agents is structured into an infinite number of patches (or islands), and where assortativity emerges as the result of the locality of both social (strategic) interactions and gene transmission.

[10]The literature on indirect evolutionary approach considers a time horizon that is long enough for selection to be active on types. See, among others, Güth and Yaari (1992), Güth (1995), Bester and Güth (1998), McNamara et al. (1999), Sethi and Somanathan (2001), Ok and Vega-Redondo (2001), Van Veelen (2006), Dekel et al. (2007), Heifetz et al. (2007b,a), Kuran and Sandholm (2008), Akçay et al. (2009) and Wu (2016).

[11]In the study of the evolution of cooperation through group selection (see, among others, Traulsen and Nowak, 2006, Bowles, 2006, Van Veelen, 2009, Choi and Bowles, 2007) a kind of separation between agents is considered, which is however quite different from action-assortativity: individuals interact mostly within their group but their actions affect the likelihood that the group survives against other groups.

the past leads to an interaction structure where there is a segregated group of cooperators who leaves out all defectors. Pin and Rogers (2015) also obtain coexistence, in a setup developed to analyze the effects of monitoring migrants' behavior – where the population of agents is comprised of migrants (coming from abroad) and citizens (who are born locally). Wang et al. (2012) provide supporting experimental evidence. Rezaei and Kirley (2012) find similar results in a network setup where links are automatically severed upon defection.

Finally, the literature on cultural transmission, which focuses on the socialization of culture from one generation to another one, is also partly related to our paper. See Cavalli-Sforza and Feldman (1981), Boyd and Richerson (1988), Bisin and Verdier (2001) among many others, for explicit modellings of inter-generational cultural transmission processes.[12] With respect to these models, we consider a shorter time-horizon: long enough to have auxiliary traits (actions) evolve endogenously, but not too long as to have cultural traits that convey an identity (types) and cultural intolerance (cost of type-mismatch) exogenously given.

1.6 Structure of the paper

Section 2 describes the baseline model and introduces the adopted notion of evolutionary stability. Section 3 identifies the evolutionarily stable states. Section 4 conducts welfare analysis. Section 5 extends the baseline model. Section 6 provides a discussion and concludes.

2 The baseline model

The model and the following analysis are built upon Bergstrom (2003). As a distinctive feature of our model, agents are heterogeneous in cultural types, and interactions between agents of different cultural types are costly.

2.1 The Prisoner Dilemma game

Consider a large population of agents who are repeatedly and randomly matched in pairs to engage in some pairwise interaction. There are two actions available to the agents, labelled by C and D. Action C stands for cooperation, it costs c to the agent who adopts it, and gives a benefit $b > c$ to the partner interacting with the agent. Action D stands for defection, it

[12] See Richerson and Boyd (2008) for a comprehensive review of the role of culture in the evolution of human behavior. See also Bowles (1998) for a discussion of the impact of institutions on the evolution of preferences. See Bisin and Verdier (2010) for a survey on recent economic applications of cultural transmission models.

costs nothing to the agent who adopts it, and gives no benefit to the partner. The resulting payoff matrix is given by

	C	D
C	$b-c, b-c$	$-c, b$
D	$b, -c$	$0, 0$

which is known as the *Prisoner Dilemma with additive payoffs*.

2.2 Heterogeneous types and cultural intolerance

Each agent in the population carries one of two cultural types, labelled by x and y. Hence, the population is divided into two cultural groups. An agent suffers a cost of cross-cultural interactions d in case of type-mismatch, i.e., for interacting with an agent of a cultural type different from her own. The cost d measures the intensity of cultural intolerance in the population. The proportion of x-type agents over the whole population is β. Without loss of generality suppose that $\beta \in (0.5, 1)$. A population state is characterized by $s = (s_x, s_y) \in [0,1]^2$, where s_x denotes the fraction of x-type agents that are cooperators, and s_y is interpreted analogously.[13] The cultural type can be initially observable or not. Whatever the case may be, we assume that agents' actions cannot be conditioned on the cultural type of the partner.[14]

Let us define:

$$\begin{aligned}
\eta_{x|C}(s) &= \frac{\beta s_x}{\beta s_x + (1-\beta)s_y}; \\
\eta_{y|C}(s) &= \frac{(1-\beta)s_y}{\beta s_x + (1-\beta)s_y}; \\
\eta_{x|D}(s) &= \frac{\beta(1-s_x)}{\beta(1-s_x) + (1-\beta)(1-s_y)}; \\
\eta_{y|D}(s) &= \frac{(1-\beta)(1-s_y)}{\beta(1-s_x) + (1-\beta)(1-s_y)};
\end{aligned}$$

[13] Strictly speaking, we limit ourselves to consider pure strategies only, so that agents can be either cooperators or defectors.

[14] See the beginning of Subsection 6.1 for a discussion of this assumption.

where $\eta_{x|C}(s)$ is the fraction of cooperators that are x-types in state s; $\eta_{x|D}(s)$, $\eta_{y|C}(s)$ and $\eta_{y|D}(s)$ are analogously interpreted.

2.3 Two-pool assortative matching process

In our daily lives, we tend to interact with those who act like us and avoid those who behave differently. To capture such a tendency, we adopt the two-pool assortative matching process with uniform assortativity by (Cavalli-Sforza and Feldman, 1981). Two-pool assortative matching process is a random matching process such that every agent in the population is matched, with probability $p \in [0, 1]$, with an agent taking the same strategy as she does (i.e., she draws from an assortative pool), and with probability $1 - p$ with a randomly selected agent (i.e., she draws from a random pool consisting of all individuals who did not match from an assortative pool).

Given a population state s, the probabilities of matching among the agents are given as

$$\begin{aligned} Pr(C|C) &= p + (1-p)(\beta s_x + (1-\beta)s_y); \\ Pr(D|C) &= (1-p)(\beta(1-s_x) + (1-\beta)(1-s_y)); \\ Pr(D|D) &= p + (1-p)(\beta(1-s_x) + (1-\beta)(1-s_y)); \\ Pr(C|D) &= (1-p)(\beta(s_x) + (1-\beta)(s_y)); \end{aligned}$$

where $Pr(C|C)$ denotes the probability that a cooperator is matched with another cooperator, $Pr(D|C)$ denotes the probability that a cooperator is matched with a defector. $Pr(D|D)$ and $Pr(C|D)$ are analogously interpreted.

One can observe that $Pr(C|C)$ and $Pr(D|D)$ are increasing in p, while $Pr(D|C)$ and $Pr(C|D)$ are decreasing in p. Hence, probability p captures how unlikely that a cooperator is matched with a defector. Probability p is referred to as the index of assortativity in Bergstrom (2003).

Note that when $p = 0$, the matching process is reduced to a uniformly random matching process. When $p = 1$, the cooperators and the defectors are completely segregated from each other.

For now, we assume that p is a constant across population state as in Bergstrom (2003). In Subsection 5.3, we generalize the analysis to the case in which the index of assortativity is not uniform across states.

2.4 Notion of evolutionary stability

As justified in the Introduction, we assume that auxiliary traits (actions) evolve faster than identity traits (cultural types). Therefore, we focus on the evolution of actions taking types as given. We do so under the standard assumption of payoff monotonicity (Weibull, 1995), simply comparing the expected payoffs of cooperators and defectors.

In a setting like ours, where the distribution of types is held fixed while the distribution of actions is not, there is no agreement in the literature on the appropriate notion of evolutionary stability.[15] Therefore, we take a conservative position opting for a quite demanding definition: one that ensures that a state is stable under any reasonable dynamics which satisfies payoff monotonicity.

We will say that a state s is ***evolutionarily stable*** if there exists an invasion barrier $\bar{\epsilon} > 0$ such that, for every pair $\epsilon_x, \epsilon_y \geq 0$, with $0 < \epsilon_x + \epsilon_y < \bar{\epsilon}$, describing the fraction of x-type mutants and y-type mutants, respectively, and for every pair (σ_x, σ_y) describing the fraction of cooperators among x-type mutants and y-type mutants, respectively, we have that mutants perform worse than the incumbents of the same type; in particular, if $\epsilon_x > 0$ and $\sigma_x \neq s_x$ then the average payoff of x-type mutants must be strictly lower than the average payoff of x-type incumbents, and if $\epsilon_y > 0$ and $\sigma_y \neq s_y$ then the average payoff of y-type mutants must be strictly lower than the average payoff of y-type incumbents. If a state is evolutionarily stable, the fraction of mutants of each type will decrease over time in any payoff monotone dynamics.

3 Evolutionarily stable states

We denote with $\pi(C, x|s)$ the expected payoff in population state s of a cooperator that is an x-type. We define $\pi(D, x|s)$, $\pi(C, y|s)$ and $\pi(D, y|s)$ analogously. For the ease of exposition, we define $\kappa = \beta s_x + (1 - \beta) s_y$.

For a x type cooperator, if she enters the assortative pool, she always gets a payoff of $b - c$ from the interaction. However, she may encounter a y type cooperator with probability $\eta_{y|C}(s)$, which costs her a penalty of d. On the other hand, if she instead enters the random pool, she only encounters another cooperator with probability κ. Moreover, among all the agents she can encounter in the random pool, $1 - \beta$ of them are y type agents. So she receives a penalty of d with probability $1 - \beta$. Therefore, the expected payoff of a x type agent is

[15] Evolutionary stability in incomplete information games with fixed distribution of types is studied in Ely and Sandholm (2005), who consider best-response dynamics, Cressman (2003, Section 4.7.2) and Amann and Possajennikov (2009), who apply replicator dynamics.

given by

$$\pi(C,x|s) \;=\; p(b-d\eta_{y|C}(s))+(1-p)[\kappa b-d(1-\beta)]-c.$$

Similarly, we can write down the expected payoffs of the other three types of agents:

$$\begin{aligned}\pi(D,x|s) &= -pd\eta_{y|D}(s)+(1-p)[\kappa b-d(1-\beta)],\\ \pi(C,y|s) &= p(b-d\eta_{x|C}(s))+(1-p)[\kappa b-d\beta]-c,\\ \pi(D,y|s) &= -pd\eta_{x|D}(s)+(1-p)[\kappa b-d\beta].\end{aligned}$$

We observe that

$$\begin{aligned}\pi(C,x|s)-\pi(D,x|s) &= p(b-d(\eta_{y|C}(s)-\eta_{y|D}(s)))-c,\\ \pi(C,y|s)-\pi(D,y|s) &= p(b-d(\eta_{x|C}(s)-\eta_{x|D}(s)))-c.\end{aligned}$$

Using the differences above, the following result can be proved:

PROPOSITION **1.** *Let $d>0$. Then:*

- $s=(1,1)$ *is an evolutionarily stable state if and only if* $p\geq\frac{c}{b-\beta d}$*;*
- $s=(0,0)$ *is an evolutionarily stable state if and only if* $p\leq\frac{c}{b+\beta d}$*;*
- $s=(1,0)$ *and* $s=(0,1)$ *are evolutionarily stable states if and only if* $\frac{c}{b+d}\leq p\leq\frac{c}{b-d}$.

There are no other states that can be evolutionarily stable.

Figure 1 provides a graphical illustration of Proposition 1. As one can see, the range of values that p can take in $[0,1]$ can be partitioned in five regions. If $p\in[0,c/(b+d))$, then $s=(0,0)$ is the unique evolutionarily stable state. If $p\in[c/(b+d),c/(b+\beta d)]$, then $s=(0,0)$, $s=(1,0)$ and $s=(0,1)$ are all and the only evolutionarily stable states. If $p\in(c/(b+\beta d,c/(b-\beta d))$, then $s=(1,0)$ and $s=(0,1)$ are all and the only the evolutionarily stable states. If $p\in[c/(b-\beta d),c/(b-d)]$, then $s=(1,1)$, $s=(1,0)$ and $s=(0,1)$ are all and the only evolutionarily stable states. Finally, if $p\in(c/(b-d),1]$, then $s=(1,1)$ is the unique evolutionarily stable state.

It is straightforward to show that when $d=0$, if $p>\frac{c}{b}$, $s=(1,1)$ is uniquely evolutionarily stable; if $p<\frac{c}{b}$, $s=(0,0)$ is uniquely evolutionarily stable. In other words, without cultural intolerance, people either all cooperate when the matching is highly assortative, or all defect when the matching is less assortative.[16]

[16] If $d=0$ and $p=\frac{c}{b}$ then no state is evolutionarily stable and both $(0,0)$ and $(1,1)$ are neutrally stable.

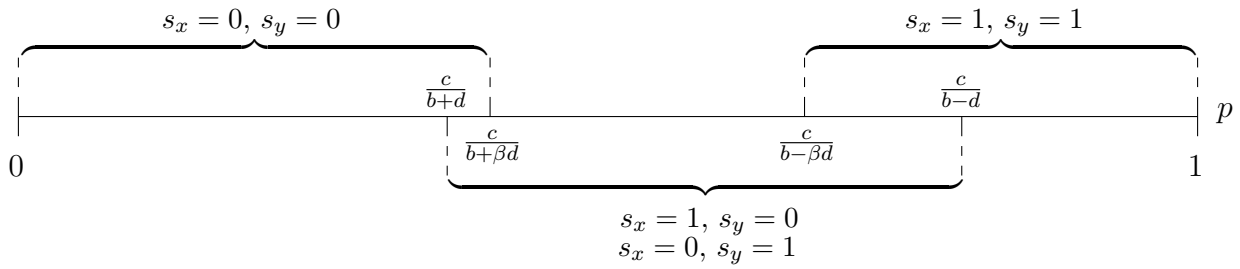


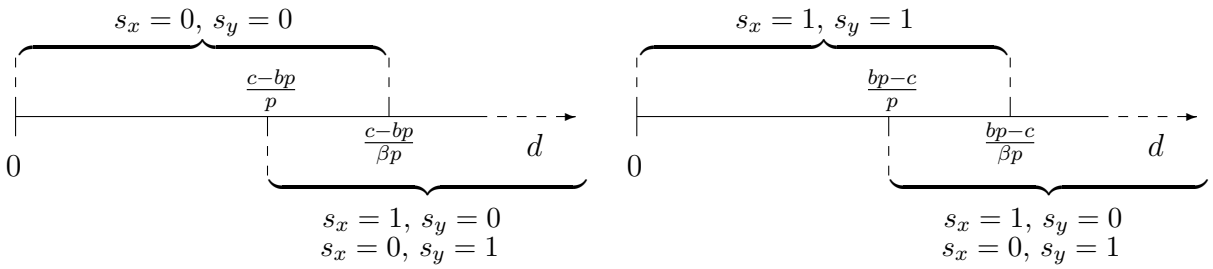
Figure 1: Evolutionarily stable states as a function of p. The picture is drawn assuming $b = 2$, $c = 1$, $d = 3/4$ and $\beta = 2/3$.

However, when cultural intolerance d is positive, type-monomorphic states $s = (0, 1)$ and $s = (1, 0)$, in which partial cooperation is sustained, emerge as the unique evolutionarily stable states in the region of $p \in (c/(b + \beta d), c/(b - \beta d))$.

The existence of type-monomorphic states hinges on the interplay between homophily and assortativity in actions. Cultural intolerance introduces an incentive for the agents from the same cultural group to coordinate on the action played by most of their group members, because the presence of assortativity in actions help them to avoid being matched with agents from the other group. Eventually, the agents with different cultural types are sorted to coordinate on different actions. This sorting result is the main novelty of this model compared to the literature on the evolution of cooperation in prisoner dilemmas.

Figure 2 depicts the evolutionarily stable states as a function of d. When assortativity level is low ($p < \frac{c}{b}$), increasing cultural intolerance helps to foster cooperation in a population consisting of only defectors. On the other hand, when assortativity is high ($p > \frac{c}{b}$), increasing cultural intolerance induces defection in a population consisting of only cooperators. In Section 4, we provide a detailed analysis of the welfare effect induced by the cost of cultural intolerance d.

To give an idea of the kind of payoff-monotone dynamics at play here, in Figure 3 we depict the phase diagram under best response dynamic, for a numerical example. In particular, we set $\beta = 0.6$, $b = 3$, $c = 1$, $d = 1$ and p takes value from $0.2, 0.35, 0.4$ and 0.5. The x-axis represents the proportion of x group agents cooperating (s_x). The y-axis represents the proportion of y group agents cooperating (s_y). When $p = 0.2$, best response dynamic converges to $s = (0, 0)$ from any initial state. When $p = 0.5$, best response dynamic converges to $s = (1, 1)$ from any initial state. When $p = 0.35$, The simplex of population states is divided into three regions, which define the basins of attractions of $s = (0, 0)$, $(0, 1)$, $(1, 0)$.



(a) The picture is drawn assuming $p < c/b$. (b) The picture is drawn assuming $p > c/b$.

Figure 2: Evolutionarily stable states as a function of d.

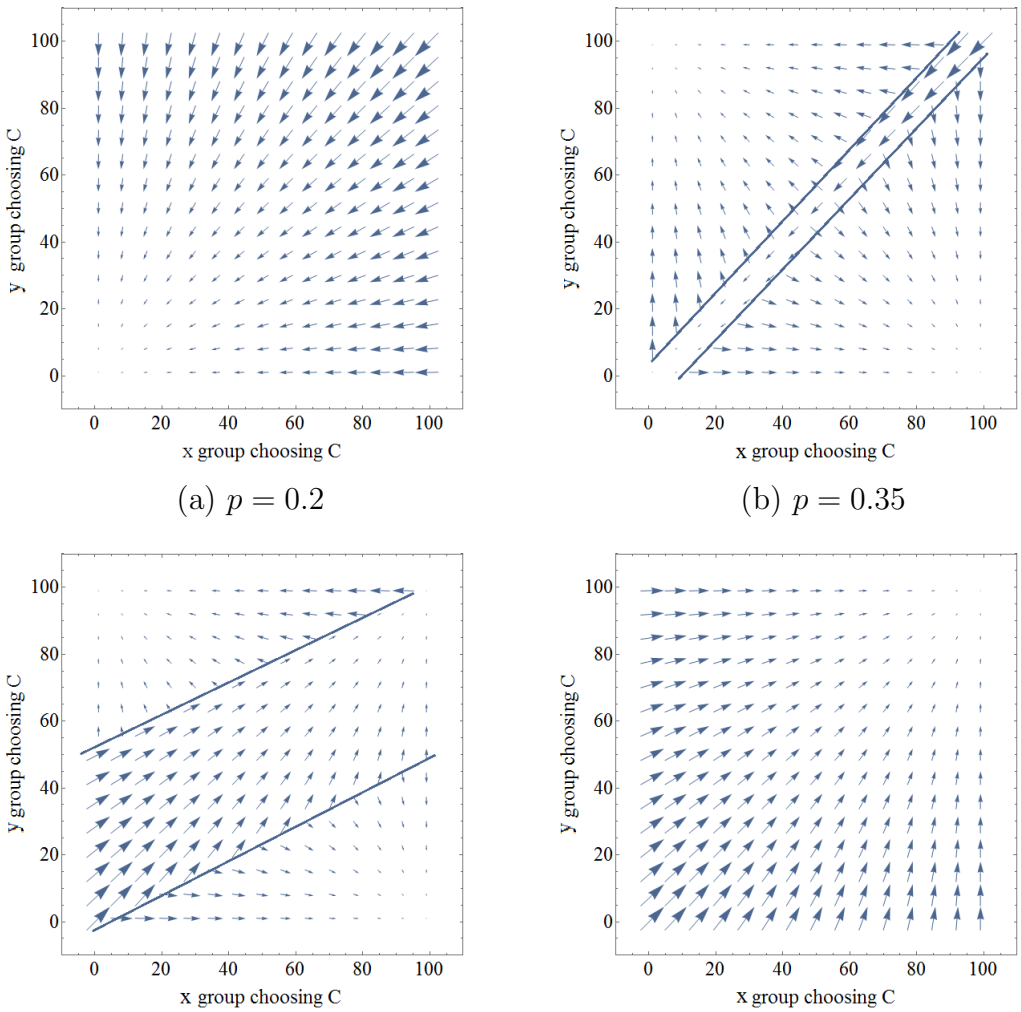
The dynamic would converge to one of these three states depending on the initial states. When $p = 0.4$, The simplex of population states is divided into three regions, which define the basins of attractions of $s = (1, 1)$, $(0, 1)$, $(1, 0)$. The dynamic would converge to one of these three states depending on the initial states.

We note that the basins of attraction of different equilibria can be potentially used to apply a stochastic stability analysis (Kandori et al., 1993, Young, 1993) in order to obtain equilibrium selection. However, the stochastic stability analysis relies on a very long time horizon. As we have discussed in the introduction, when time horizon is sufficient long, agents' identity traits are expected to evolve as well, and this would require a different model. We provide a brief discussion on what to expect in terms of long-run dynamic in Subsection 6.3.

One final comment regards the role of the relative size of cultural groups. Note that, in the presence of cultural intolerance, β also affects how action-assortativity determines the emergence of cooperation. In particular, as β gets closer to 1, i.e., as the two cultural groups have more and more unequal sizes, a greater action-assortativity is necessary for $s = (1, 1)$ to be evolutionarily stable and a smaller one for $s = (0, 0)$, while the stability of type-monomorphic states is unaffected. So, in a sense, the asymmetry in the size of cultural groups makes the emergence of type-monomorphic states more likely, at least for intermediate values of action-assortativity.

4 Welfare

We have shown how a larger cultural intolerance can promote more or less cooperation in a society, depending on which state – either cooperation or defection by everybody –



(a) $p = 0.2$ (b) $p = 0.35$

(c) $p = 0.4$ (d) $p = 0.5$

Figure 3: Phase diagrams under best response dynamic.

emerges in the absence of cultural intolerance, which in turn depends on the degree of action-assortativity. The effects of cultural intolerance, however, are not limited to changes in the extent of cooperation; they also involve modifications to the cost and frequency of type-mismatches. In order to evaluate the overall impact of cultural intolerance, we focus our attention on societal welfare, which is simply the sum of individual payoffs over the whole population.

We denote the societal welfare of a state $s = (s_x, s_y)$ with $W(s_x, s_y)$. In a monomorphic state societal welfare is $W(1,1) = b - c - 2\beta(1-\beta)d$ if everybody cooperates, while it is $W(0,0) = -2\beta(1-\beta)d$ if everybody defects. Instead, in a type-monomorphic state we have either that $W(1,0) = \beta(b-c) - 2\beta(1-\beta)(1-p)d$ or that $W(0,1) = (1-\beta)(b-c) - 2\beta(1-\beta)(1-p)d$ depending on whether, respectively, the majority cooperates or the minority cooperates.

In any range of parameters where only one state is evolutionarily stable, the minimal cultural intolerance is clearly optimal for any $p > 0$, since type-mismatches are costly, ceteris paribus, from a societal point of view. However, $d = 0$ is not necessarily the most desirable situation. Indeed, besides the cost of type mismatches, there are two other effects of cultural intolerance on welfare that are not necessarily negative. These effects exist when cultural intolerance is strong enough that also type-monomorphic states are evolutionarily stable. One effect is always beneficial and is the incentive to coordinate on the same action that is adopted by one's own type, which can greatly reduce the number of type-mismatches; the strength of such effect crucially, and positively, depends on the degree of action-assortativity. The second effect is the result of the first: in order to avoid type mismatches an agent can be induced to adopt a different action; evidently, this effect is beneficial when the change is from defection to cooperation and detrimental when the change goes in the opposite direction.

When $p < c/b$, the monomorphic state that is evolutionarily stable is defection by everybody, so a type-monomorphic state can be preferred as it allows more cooperation (see Figure 4, left panel, for an example). In particular, when both kinds of states are evolutionarily stable we have that $W(1,0) > W(0,1) > W(0,0)$ for all feasible values of d. So, if d is not too large, the type-monomorphic state with some cultural intolerance is preferable to the monomorphic state where everybody defects but there is no cultural intolerance at all – and, of course, the type-monomorphic state where the majority cooperates is the most preferred. The gains from cooperation, although only a part of the population is involved, more than offset the increased cost of type mismatches, which however are reduced in number.

Instead, when $p > c/b$, it is obviously optimal to have $d = 0$, because the monomorphic state where everybody cooperates is evolutionarily stable – and cooperation by everybody

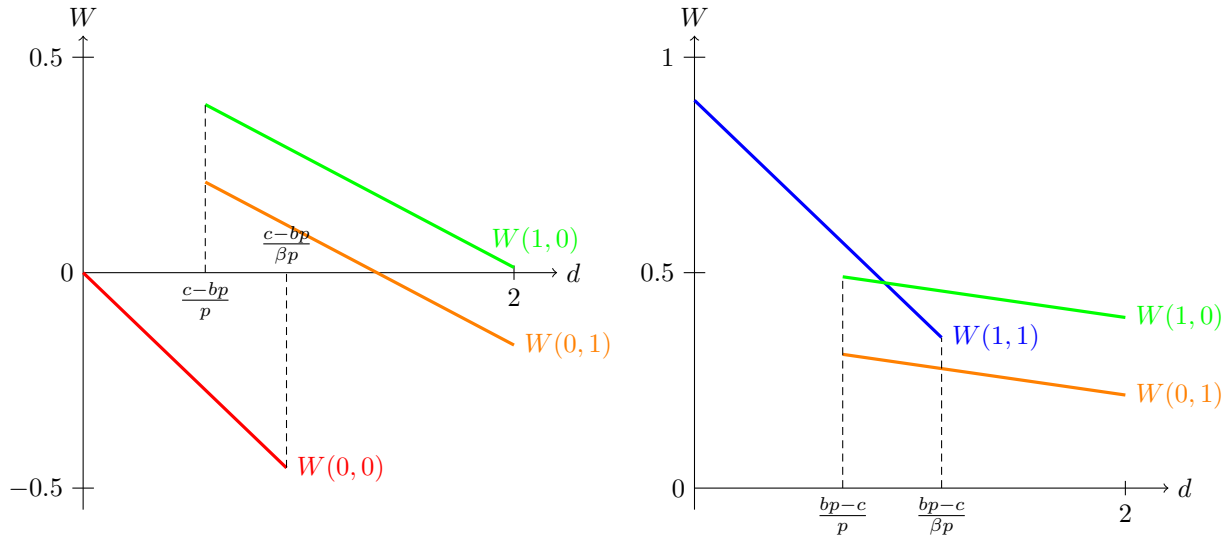


Figure 4: Welfare of evolutionarily stable states as a function of d. The pictures are drawn assuming $b = 2.1$, $c = 1.2$, $\beta = 0.6$, and $p = 0.45$ on the left, while $p = 0.85$ on the right.

in the absence of cultural intolerance is, by construction, the first best. However, if some cultural intolerance is present, and it is sufficiently strong to make also type-monomorphic states evolutionarily stable, it may happen that $W(1,0) > W(0,1) > W(1,1)$ (see Figure 4, right panel, for an example). If d is not too large, the lost benefits of cooperation (due to the fact that part of the population now defects to avoid type mismatches) can be more than offset by the reduced number of type-mismatches, even if the cost of a single type mismatch can be greater.

The following proposition summarizes:

PROPOSITION **2.** *Suppose that* $(s_x = 1, s_y = 0)$ *and* $(s_x = 0, s_y = 1)$ *are evolutionarily stable states. It follows that* $W(1,0) > W(0,1)$*. In addition, if they are not the only evolutionarily stable states, then we have that either:*

- $(s_x = 0, s_y = 0)$ *is also evolutionarily stable, with* $W(1,0) > W(0,1) > W(0,0)$*;*
- $(s_x = 1, s_y = 1)$ *is also evolutionarily stable, with* $W(1,0) > W(1,1)$ *if and only if* $dp > \frac{(b-c)}{2\beta}$*, and* $W(0,1) > W(1,1)$ *if and only if* $dp > \frac{(b-c)}{2(1-\beta)}$*.*

Proposition 2 indicates that homophily and action-assortativity have a non-trivial interplay for what concerns their effect on welfare. In particular, when $d > 0$ it is possible that p has a non-monotonic impact on welfare. Figure 5 illustrates an example of this case. We stress that for $d = 0$ this can not happen.

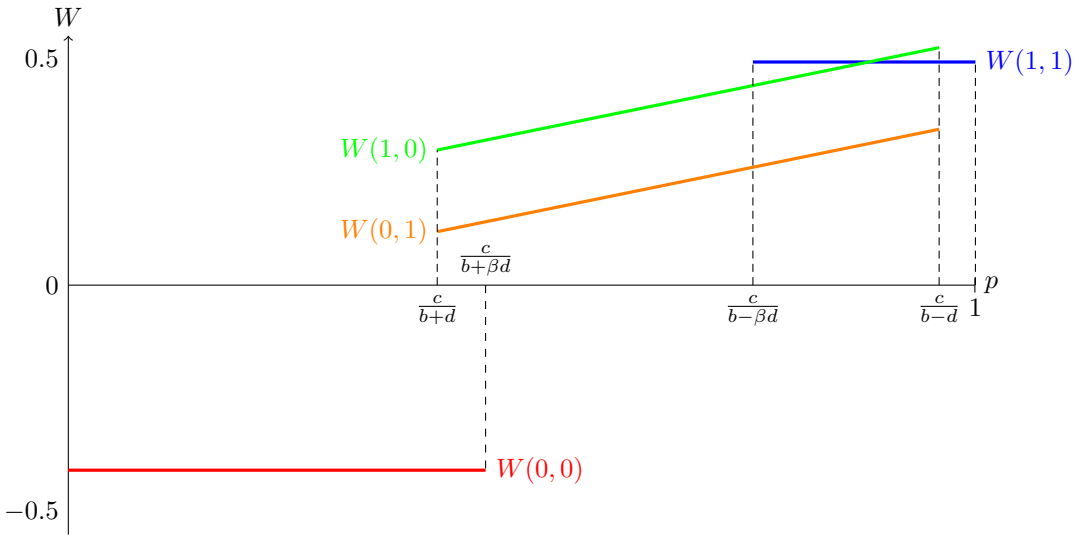


Figure 5: Welfare of evolutionarily stable states as a function of p. The picture is drawn assuming $b = 2.1$, $c = 1.2$, $d = 0.85$ and $\beta = 0.6$.

5 Model extensions

5.1 Asymmetric cultural intolerance

Cultural intolerance may be asymmetric. For example, as measured in Bisin et al. (2004), in the United States, the cultural intolerance of Protestants towards Catholics and that of Catholics towards Protestants are similar. However, the cultural intolerance of Jews towards Catholics is much stronger than that of Catholics towards Jews. In this section, we study the consequence of asymmetric cultural intolerance.

Let d_x denote the cost of cross-cultural interactions suffered by group x members who are matched with group y members – and d_y denote the cost of cross-cultural interactions suffered by group y members who are matched with group x members – and suppose that $d_x \neq d_y$. Intuitively, $d_x > d_y$ captures the situation in which a conservative majority is hostile against a minority who is not so unhappy to work with the majority. Conversely, $d_x < d_y$ captures the situation in which a relatively open-minded majority faces an inward-looking minority.

Under asymmetric cultural intolerance, the following result holds:

PROPOSITION **3.** *Let $d_x > 0$ and $d_y > 0$. Then:*

- *$s = (1, 1)$ is an evolutionarily stable state if and only if $p \geq \frac{c}{b - \max\{\beta d_y, (1-\beta) d_x\}}$;*

- $s = (0, 0)$ *is an evolutionarily stable state if and only if* $p \leq \frac{c}{b+\max\{\beta d_y,(1-\beta)d_x\}}$;
- $s = (1, 0)$ *is an evolutionarily stable state if and only if* $\frac{c}{b+d_x} \leq p \leq \frac{c}{b-d_y}$;
- $s = (0, 1)$ *is an evolutionarily stable state if and only if* $\frac{c}{b+d_y} \leq p \leq \frac{c}{b-d_x}$.

There are no other states that can be evolutionarily stable.

Compared to Proposition 1, Proposition 3 provides an additional insight to the problem. That is, asymmetry in cultural intolerance serves as an equilibrium selection mechanism between the two type-monomorphic states. For instance, when $d_x > d_y$ we have that for $c/(b + d_x) < p < c/(b + d_y)$ the type-monomorphic state $s = (1, 0)$ is evolutionarily stable, while $s = (0, 1)$ is not.

Also, Proposition 3 indicates that when action-assortativity increases, the group with stronger cultural intolerance is likely to cooperate first. However, as action-assortativity keeps increasing, the group with weaker cultural intolerance finds cooperation beneficial. This leads to defection in the group with stronger cultural intolerance because of its members' strong desire to differentiate themselves from the other group.

5.2 Assortativity in cultural types

Besides the tendency to interact more with people who act similarly to us, we tend to interact more with those who are similar to ourselves along the cultural dimensions, such as language, religion, dress and origin. To capture the assortativity in both actions and cultural types, we develop the analysis in a variant of the two-pool assortative process, that we call *dual two-pool assortative process*. Every agent in the population enters with probability p a pool where all agents play her same action, and with probability $1 - p$ she enters a random pool consisting of all individuals who did not enter the action-assortative pool. After that, every agent, whatever pool has entered, enters with probability q a sub-pool where all agents have her same type, and with probability $1 - q$ she enters a random pool consisting of all individuals who did not enter the type-assortative pool.

We observe that p can still be understood as the index of assortativity in actions, while q can analogously be understood as the index of assortativity in cultural types.

We remind that $\kappa = \beta s_x + (1 - \beta)s_y$. Expected payoffs at population state s are given

by:

$$\begin{aligned}
\pi(C,x|s) &= pqb + p(1-q)(b-\eta_{y|C}(s)d) + (1-p)q\kappa b + (1-p)(1-q)(\kappa b - (1-\beta)d) - c,\\
\pi(D,x|s) &= pq0 + p(1-q)(-\eta_{y|D}(s)d) + (1-p)q\kappa b + (1-p)(1-q)(\kappa b - (1-\beta)d),\\
\pi(C,y|s) &= pqb + p(1-q)(b-\eta_{x|C}(s)d) + (1-p)q\kappa b + (1-p)(1-q)(\kappa b - \beta d) - c,\\
\pi(D,y|s) &= pq0 + p(1-q)(-\eta_{x|D}(s)d) + (1-p)q\kappa b + (1-p)(1-q)(\kappa b - \beta d).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\pi(C,x|s) - \pi(D,x|s) &= p(b - d(\eta_{y|C}(s) - \eta_{y|D}(s))(1-q)) - c,\\
\pi(C,y|s) - \pi(D,y|s) &= p(b - d(\eta_{x|C}(s) - \eta_{x|D}(s))(1-q)) - c.
\end{aligned}$$

Using the differences above, the following result can be proved:

PROPOSITION **4.** *Let $d > 0$ and $q > 0$. Then:*

- $s = (1,1)$ *is an evolutionarily stable state if and only if* $p \geq \frac{c}{b-(1-q)\beta d}$;
- $s = (0,0)$ *is an evolutionarily stable state if and only if* $p \leq \frac{c}{b+(1-q)\beta d}$;
- $s = (1,0)$ *and* $s = (0,1)$ *are evolutionarily stable states if and only if* $p \geq \frac{c}{b+(1-q)d}$ *and* $p < \frac{c}{b-(1-q)d}$.

There are no other states that can be evolutionarily stable.

Proposition 4 is qualitatively similar to Proposition 1. From this we can conclude that our results, which are based on assortativity in actions, are robust to the presence of assortativity in types as well. However, we observe that the higher is q, the smaller is the interval of values of p for which type-monomorphic states are evolutionarily stable (and, also, the larger are the interval of values of p for which monomorphic states are evolutionarily stable). Indeed, a higher q reduces the net effect of actions as instruments to avoid type-mismatches.

Note that Proposition 4 implies that when $p = 0$, $s = (0,0)$ is the only evolutionarily stable state. When $d = 0$, $s = (0,0)$ and $s = (1,1)$ are the only possible evolutionarily stable states. Therefore, assortativity in cultural types alone, without either cultural intolerance or assortativity in actions, cannot produce type-monomorphic stable states.

Let us conclude this discussion of the role of type-assortativity by pointing out a relationship that we have not considered here, but is potentially interesting. One might argue that cultural intolerance and the degree of type-assortativity are not independent of one another. In particular, q might be an increasing function of d: the higher the cost of type-mismatches,

the more effective will be the mechanisms developed by each culture to avoid interactions with members of other cultures. In such a case, which states are evolutionarily stable, as d varies, depends on how sensitive q is to changes in d. In principle, all our results remain valid, provided that q is sufficiently insensitive. We consider the study of specific mechanisms linking q to d as a promising direction for future research.

5.3 State-dependent assortativity

So far we have considered a particular matching process, the two-pool matching process, that entails a state-independent index of assortativity. However, the index of assortativity can be, in general, state-dependent. In this subsection we show that the introduction of state-dependent assortativity does not affect the gist of our results.

We denote with $p(C|D,s)$ the probability that a defector will encounter a cooperator in state s. We define $p(C|C,s)$, $p(D|C,s)$ and $p(D|D,s)$ analogously. So, we have that:

$$\begin{aligned}
\pi(C,x|s) &= p(C|C,s)(b-d\eta_{y|C}(s))+p(D|C,s)(-d\eta_{y|D}(s))-c,\\
\pi(D,x|s) &= p(C|D,s)(b-d\eta_{y|C}(s))+p(D|D,s)(-d\eta_{y|D}(s)),\\
\pi(C,y|s) &= p(C|C,s)(b-d\eta_{x|C}(s))+p(D|C,s)(-d\eta_{x|D}(s))-c,\\
\pi(D,y|s) &= p(C|D,s)(b-d\eta_{x|C}(s))+p(D|D,s)(-d\eta_{x|D}(s)).
\end{aligned}$$

In this setup the index of assortativity in actions is given by:

$$a(s)=p(C|C,s)-p(C|D,s).$$

We observe that $p(C|C,s)-p(C|D,s)=p(D|D,s)-p(D|C,s)$. Therefore,

$$\begin{aligned}
\pi(C,x|s)-\pi(D,x|s) &= a(s)(b-d(\eta_{y|C}(s)-\eta_{y|D}(s)))-c,\\
\pi(C,y|s)-\pi(D,y|s) &= a(s)(b-d(\eta_{x|C}(s)-\eta_{x|D}(s)))-c.
\end{aligned}$$

Using the differences above, the following result can be proved:

PROPOSITION **5.** *Let $d>0$. Then:*

- $(s_x=1,s_y=1)$ *is an evolutionarily stable state if* $a(s)>\frac{c}{b-\beta d}$;
- $(s_x=0,s_y=0)$ *is an evolutionarily stable state if* $a(s)<\frac{c}{b+\beta d}$;
- $(s_x=1,s_y=0)$ *and* $(s_x=0,s_y=1)$ *are evolutionarily stable states if* $\frac{c}{b+d}<a(s)<\frac{c}{b-d}$.

There are no other states that can be evolutionarily stable.

The results in Proposition 5 involve inequalities where the index of assortativity is asked to be larger or smaller than some thresholds. Since the index $a(\cdot)$ is now state-dependent, the possibility for a state to be evolutionarily stable hinges on the actual specification of $a(\cdot)$. In particular, specific values of the index of assortativity should be derived from the underlying matching process. In the following we explore a specific matching process that is alternative to the two-pool matching process and which entails a state-dependent index of assortativity: the strangers-in-the-night matching process with uniform assortativity (Bergstrom, 2013).

The strangers-in-the-night matching process is such that every agent in the population is randomly matched with another agent, and the pair is actually formed to play the game with probability n if the two agents are *alike* in actions, and with probability m if the two agents are *different* in actions. We assume, as typical, that $n > m$.

To adapt a strangers-in-the-night matching process with uniform assortativity in actions to our setup, we particularize eq. 20 in Bergstrom (2013) to the following, for every state s:

$$a(s) = \frac{[\beta s_x + (1-\beta)s_y][\beta(1-s_x) + (1-\beta)(1-s_y)](n^2-m^2)}{[\beta s_x + (1-\beta)s_y][\beta(1-s_x) + (1-\beta)(1-s_y)](n-m)^2 + nm}. \tag{1}$$

The next result follows directly from Proposition 5 in the light of (1).

PROPOSITION **6.** *Let $d > 0$. Then:*

- $(s_x = 0, s_y = 0)$ *is always an evolutionarily stable state;*
- $(s_x = 1, s_y = 0)$ *and* $(s_x = 0, s_y = 1)$ *are evolutionarily stable states if* $\frac{c}{b+d} < \frac{\beta(1-\beta)(n^2-m^2)}{\beta(1-\beta)(n-m)^2+nm} < \frac{c}{b-d}$.

There are no other states that can be evolutionarily stable.

By looking at (1) one can see why the state where everybody cooperates is unstable while the state where everybody defects is stable: assortativity tends to zero as the state tends to become monomorphic. Moreover, it can be useful to contrast our Proposition 6 with Bergstrom (2013, Theorem 6), where similar results are obtained. We stress, however, that in our setting, differently from Bergstrom (2013), the coexistence of cooperation and defection necessarily comes with the separation of cooperators and defectors on the basis of their cultural types.

6 Discussion

In this paper we have explored the interplay between assortativity in actions and homophily as induced by cultural intolerance. We have found that, in the presence of action-assortativity, cultural intolerance works in favor of states where there is perfect correlation between culture and behavior. This happens because, thanks to action-assortativity, cooperation and defection can work as instruments to avoid interaction with individuals of another culture.

6.1 Action-assortativity vs. strategy-assortativity

In the model, we have considered assortativity in actions that are independent of the opponent's type. An alternative would be to rely on a broader notion of assortativity in actions, which might be referred to as strategy-assortativity: a generic agent chooses a strategy (a, a'), where a is the action he plays against a partner of the majority culture and a' the action he plays against a partner of the minority culture; strategy-assortativity would imply that agents are more likely to be matched together if they have chosen the same strategy, i.e., the same pair (a, a').

Replacing action-assortativity with strategy-assortativity would serve as an interesting direction for future research. However it would probably better fit real situations that differ from those we have in mind in this paper; in particular, as we have discussed earlier, the most relevant context related to our model involves anti-discrimination laws which naturally prohibit people from choosing cooperation or defection based on their opponents' cultural types. Moreover, we remark that an auxiliary trait that induces conditional behavior – depending on the opponent's type – may fail to be evolutionary successful if it is more costly to be supported than the simpler auxiliary traits that always induce the same behavior towards any opponent.

6.2 Empirical relevance

We believe that our results are empirically relevant in at least two respects. First, our findings can account for the coexistence of a group of cooperators and a group of defectors, as well as the segregation in actions of the two cultural groups, which is a pattern that can be observed in the real life. For example, social scientists have long documented and analyzed the hyper-segregation phenomenon in the United States (Massey and Denton, 1993, Cutler et al., 1999). Certain cultural groups experienced high levels of segregation and developed a

persistent "ghetto culture", which is associated with a poor work ethic (Hannerz, 1969, Lewis, 1969, Wilson, 1987, 1997, Lemman, 1991, Bonney, 1975, Sáez-Martı and Zenou, 2012).[17]

One supporting theory for such a phenomenon is based on the controversial hypothesis of "acting white" by Fordham and Ogbu (1986), who argue that some minorities are discouraged from achieving in school (or in life in general) by the negative prejudices of ethnic peers. Although the hypothesis gains public supports, several rigorous empirical studies including Cook and Ludwig (1997) and Tyson et al. (2005) find no statistical support for it.

Our model provides an alternative explanation for the phenomenon as a consequence of strategic choice: because of action-assortativity, shirking in effort (defect) prevents a member of these cultural groups from being estranged from their own cultural group members. However, at the same time, defecting separates them from good jobs or educational opportunities (matching with cooperators), which induces low academic achievement, high unemployment and poverty.[18]

In this regard, we stress that our model can explain why, even in the presence of strong anti-discrimination laws, we may observe segregation and stratification between cultural groups.

Second, the experimental evidence provided by Currarini and Mengel (2016) shows that agents are willing to pay to increase the probability to match with their own group members. In particular, Currarini and Mengel (2016) find that when such an homophilous behavior is allowed, in-group biases (in terms of favorable behavior towards own-group members) are reduced or disappear. Our analysis can help to explain the experimental data: if it is cultural intolerance that leads to homophily but not in-group altruism, we should indeed observe that agents are ready to pay for interacting with their own group members and cooperation should not be conditional on group types.

[17]Such a phenomenon is no longer unique in the United states, but also becomes more prevalent in the Western European countries after decades of immigration waves from the Northern Africa and the Western Asia.

[18]Note that Akerlof and Kranton (2000) offers an explanation for persistent poverty of certain cultural groups. Their model premises on two assumptions: 1) a minority group is socially excluded by the dominant group in the society, 2) there exists an alternative identity for the minority group to choose which is prescribed with bad behavior. Hence, given high social exclusion, choosing the alternative identity may help a minority group member to avoid suffering from being not accepted by the dominant group. Our model, though shares similar spirits, does not assume a priori the existence of social exclusion and an alternative identity, but still leads to a result of segregation and stratification.

6.3 Future research

In this paper, we have studied the case in which the preference for interacting with people of the same group stems from cultural intolerance: a cost when interacting with individuals of different cultures. The same preference can well stem from cultural love: a benefit when interacting with individuals of similar culture.[19] In terms of evolutionary stability of equilibria, nothing would change with respect to the results that we have provided in this paper. However, welfare implications, and therefore policy implications, would change. Indeed, while cultural intolerance imposes direct cost in all cases of cultural mismatch, cultural love provide a direct benefit that incurs no indirect cost as long as it does not compromise cooperation (e.g., by moving the system from a state where everybody cooperates to a state where only one culture cooperates).

Another interesting research direction involves considering the case where the cost of cross-cultural interaction is not independent of the actions taken. In particular, such a cost might be larger for a cooperator than for a defector because cooperation may require an agent to engage in active collaboration with his partner. The higher cost associated with cooperation naturally makes cooperation less attractive. Nevertheless, one can show that a type-monomorphic state would still be evolutionarily stable even in the extreme situation where the cost of cross-cultural interaction is paid only by a cooperator but not by a defector for the following reason. In a type-monomorphic state, the threshold value for the degree of action-assortativity that makes cooperation better than defection for the population of cooperators is higher than in the case considered in this paper, but still lower than that for the population of defectors; this is so because, even if defection has the same cost of cross-cultural interaction for both populations (namely, zero), cooperation entails a larger cost for the population of defectors than for the population of cooperators. Hence, for degrees of action-assortativity in between the two thresholds, a type-monomorphic state would be evolutionarily stable.

Finally, the most natural continuation of this work is to consider the evolution over a longer time horizon, where identity traits can also vary, and therefore the size of cultural groups is endogenous. Intuitively, the identity trait that will spread in the longer run is the one whose carriers earn the highest payoff. So, starting from the evolutionarily stable equilibria of the model studied in this paper, we can speculate that if there is homophily

[19] Both cost and benefit have been considered in the literature of identity economics. Akerlof and Kranton (2000) assume there are negative identity externalities between two agents associated with different identities. Akerlof and Kranton (2005) on the other hand assume that an agent gains an identity utility by associating herself with a certain identity.

resulting from cultural intolerance, then the cooperating culture will spread at the expenses of the defecting culture. On the other hand, if homophily does not emerge because cultural intolerance is too weak, both cultures survive in the long run (as they earn the same payoff). In other words, it is cultural intolerance itself that may lead cultural selection. A last observation is worth exploring in this regard: since, as noted at the end of Section 3, a greater inequality in the size of the two cultures makes type-monomorphic states more likely, it also makes the selection of the cooperating culture more likely in the longer run.

References

Ahern, K. R., D. Daminelli, and C. Fracassi (2015). Lost in translation? The effect of cultural values on mergers around the world. *Journal of Financial Economics 117*(1), 165–189.

Akçay, E., J. Van Cleve, M. W. Feldman, and J. Roughgarden (2009). A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *Proceedings of the National Academy of Sciences 106*(45), 19061–19066.

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

Akerlof, G. A. and R. E. Kranton (2005). Identity and the economics of organizations. *Journal of Economic Perspectives 19*(1), 9–32.

Alger, I. (2010). Public goods games, altruism, and evolution. *Journal of Public Economic Theory 12*(4), 789–813.

Alger, I. and J. W. Weibull (2010). Kinship, incentives, and evolution. *American Economic Review 100*(4), 1725–1758.

Alger, I. and J. W. Weibull (2012). A generalization of hamilton's rule: Love others how much? *Journal of Theoretical Biology 299*, 42–54.

Alger, I. and J. W. Weibull (2013). Homo moralis: Preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2302.

Alger, I. and J. W. Weibull (2016). Evolution and kantian morality. *Games and Economic Behavior 98*, 56–67.

Amann, E. and A. Possajennikov (2009). On the stability of evolutionary dynamics in games with incomplete information. *Mathematical Social Sciences 58*(3), 310–321.

Bandiera, O., I. Barankay, and I. Rasul (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica 77*(4), 1047–1094.

Becker, G. S. (1957). The economics of discrimination.

Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of Political Economy 101*(3), 385–409.

Bergstrom, T. C. (2003). The algebra of assortative encounters and the evolution of cooperation. *International Game Theory Review 5*(03), 211–228.

Bergstrom, T. C. (2013). Measures of assortativity. *Biological Theory 8*(2), 133–141.

Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg more employable than Jamal? a field experiment on labor market discrimination. *American Economic Review 94*(4), 991–1013.

Bester, H. and W. Güth (1998). Is altruism evolutionarily stable? *Journal of Economic Behavior & Organization 34*(2), 193–209.

Bilancini, E. and L. Boncinelli (2009). The co-evolution of cooperation and defection under local interaction and endogenous network formation. *Journal of Economic Behavior & Organization 70*(1), 186–195.

Bilancini, E. and L. Boncinelli (2017). Social coordination with locally observable types. *Economic Theory*. doi: https://doi.org/10.1007/s00199-017-1047-y.

Bisin, A., G. Topa, and T. Verdier (2004). Religious intermarriage and socialization in the united states. *Journal of Political Economy 112*(3), 615–664.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory 97*(2), 298–319.

Bisin, A. and T. Verdier (2010). The economics of cultural transmission and socialization. In M. J. Jess Benhabib, Alberto Bisin (Ed.), *Handbook of social economics*, Volume 1, Chapter 9, pp. 339–416.

Bonney, N. (1975). Work and ghetto culture. *British Journal of Sociology 26*, 435–447.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic literature 36*(1), 75–111.

Bowles, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. *Science 314*(5805), 1569–1572.

Boyd, R., H. Gintis, S. Bowles, and P. J. Richerson (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences 100*(6), 3531–3535.

Boyd, R. and P. J. Richerson (1988). *Culture and the evolutionary process*. University of Chicago Press.

Boyd, R. and P. J. Richerson (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*(1533), 3281–3288.

Boyd, R., P. J. Richerson, and J. Henrich (2011). Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology 65*(3), 431–444.

Bramoullé, Y., S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory 147*(5), 1754–1786.

Cavalli-Sforza, L. L. and M. W. Feldman (1981). *Cultural transmission and evolution: a quantitative approach*. Princeton University Press.

Choi, J.-K. and S. Bowles (2007). The coevolution of parochial altruism and war. *Science 318*(5850), 636–640.

Cook, P. J. and J. Ludwig (1997). Weighing the "burden of'acting white'": Are there race differences in attitudes toward education? *Journal of policy analysis and management*, 256–278.

Cressman, R. (2003). *Evolutionary dynamics and extensive form games*. MIT Press.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Currarini, S., M. O. Jackson, and P. Pin (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences 107*(11), 4857–4861.

Currarini, S., J. Matheson, and F. Vega-Redondo (2016). A simple model of homophily in social networks. *European Economic Review 90*, 18–39.

Currarini, S. and F. Mengel (2016). Identity, homophily and in-group bias. *European Economic Review 90*, 40–55.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999). The rise and decline of the american ghetto. *Journal of Political Economy 107*(3), 455–506.

Dekel, E., J. C. Ely, and O. Yilankaya (2007). Evolution of preferences. *Review of Economic Studies 74*(3), 685–704.

Ely, J. C. and W. H. Sandholm (2005). Evolution in bayesian games i: theory. *Games and Economic Behavior 53*(1), 83–109.

Fisman, R., Y. Hamao, and Y. Wang (2014). Nationalism and economic exchange: Evidence from shocks to sino-japanese relations. *Review of Financial Studies 27*(9), 2626–2660.

Fordham, S. and J. U. Ogbu (1986). Black students' school success: Coping with the burden of acting white. *The urban review 18*(3), 176–206.

Giannetti, M. and Y. Yafeh (2012). Do cultural differences between contracting parties matter? evidence from syndicated bank loans. *Management Science 58*(2), 365–383.

Guiso, L., P. Sapienza, L. Zingales, et al. (2009). Cultural biases in economic exchange? *Quarterly Journal of Economics 124*(3), 1095–1131.

Güth, W. (1995). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory 24*(4), 323–344.

Güth, W. and M. Yaari (1992). An evolutionary approach to explain reciprocal behavior in a simple strategic game. *U. Witt. Explaining Process and Change–Approaches to Evolutionary Economics. Ann Arbor*, 23–34.

Hannerz, U. (1969). *Soulside: Inquiries into ghetto culture and community*. Columbia University Press.

Heifetz, A., C. Shannon, and Y. Spiegel (2007a). The dynamic evolution of preferences. *Economic Theory 32*(2), 251–286.

Heifetz, A., C. Shannon, and Y. Spiegel (2007b). What to maximize if you must. *Journal of Economic Theory 133*(1), 31–57.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization 53*(1), 3–35.

Henrich, J. and R. Boyd (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology 208*(1), 79–89.

Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 29–56.

Kuran, T. and W. H. Sandholm (2008). Cultural integration and its discontents. *Review of Economic Studies 75*(1), 201–228.

Lehmann, L., I. Alger, and J. Weibull (2015). Does evolution lead to maximizing behavior? *Evolution 69*(7), 1858–1873.

Lemman, N. (1991). *The promised land: The great Black migration and how it changed America*. Vintage.

Lewis, O. (1969). Culture of poverty. In D. P. Moynihan (Ed.), *On Understanding Poverty*, pp. 201–213. London., Basic Books, Inc.

Massey, D. S. and N. A. Denton (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.

McNamara, J. M., C. E. Gasson, and A. I. Houston (1999). Incorporating rules for responding into evolutionary games. *Nature 401*(6751), 368–371.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*, 415.

Michaels, G. and X. Zhi (2010). Freedom fries. *American Economic Journal: Applied Economics 2*(3), 256–281.

Nax, H. H., S. Balietti, R. O. Murphy, and D. Helbing (2015). Meritocratic matching can dissolve the efficiency-equality tradeoff: The case of voluntary contributions games.

Nax, H. H., R. O. Murphy, and D. Helbing (2014). Stability and welfare of 'merit-based' group-matching mechanisms in voluntary contribution game. *Available at SSRN 2404280*.

Ok, E. A. and F. Vega-Redondo (2001). On the evolution of individualistic preferences: An incomplete information scenario. *Journal of Economic Theory 97*(2), 231–254.

Pin, P. and B. W. Rogers (2015). Cooperation, punishment and immigration. *Journal of Economic Theory 160*, 72–101.

Rezaei, G. and M. Kirley (2012). Dynamic social networks facilitate cooperation in the n-player prisoners dilemma. *Physica A: Statistical Mechanics and its Applications 391*(23), 6199–6211.

Richerson, P. J. and R. Boyd (2008). *Not by genes alone: How culture transformed human evolution*. University of Chicago Press.

Rivas, J. (2013). Cooperation, imitation and partial rematching. *Games and Economic Behavior 79*, 148–162.

Ruef, M., H. E. Aldrich, and N. M. Carter (2003). The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American Sociological Review 68*(2), 195–222.

Sáez-Martı, M. and Y. Zenou (2012). Cultural transmission and discrimination. *Journal of Urban Economics 72*(2), 137–146.

Sethi, R. and E. Somanathan (2001). Preference evolution and reciprocity. *Journal of Economic Theory 97*(2), 273–297.

Traulsen, A. and M. A. Nowak (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences 103*(29), 10952–10955.

Tyson, K., W. Darity Jr, and D. R. Castellino (2005). It's not a black thing: Understanding the burden of acting white and other dilemmas of high achievement. *American Sociological Review 70*(4), 582–605.

Van Veelen, M. (2006). Why kin and group selection models may not be enough to explain human other-regarding behaviour. *Journal of Theoretical Biology 242*(3), 790–797.

Van Veelen, M. (2009). Group selection, kin selection, altruism and cooperation: when inclusive fitness is right and when it can be wrong. *Journal of Theoretical Biology 259*(3), 589–600.

Walster, E., V. Aronson, D. Abrahams, and L. Rottman (1966). Importance of physical attractiveness in dating behavior. *Journal of Personality and Social Psychology 4*(5), 508.

Wang, J., S. Suri, and D. J. Watts (2012). Cooperation and assortativity with dynamic partner updating. *Proceedings of the National Academy of Sciences 109*(36), 14363–14368.

Weibull, J. (1995). *Evolutionary game theory*. MIT press.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press.

Wilson, W. J. (1997). *When Work Disappears: The world of the new urban poor*. Vintage.

Wu, J. (2016). Social connections and cultural heterogeneity. *Mimeo*.

Young, H. P. (1993). The evolution of conventions. *Econometrica*, 57–84.

A Appendix - Proofs

In the following we collect the proofs of all Propositions in the paper. For the sake of brevity, wherever possible we avoid repeating similar arguments developed in other proofs, limiting ourselves to highlight what adjustments have be done to prove the desired results (see, e.g., the proof of Proposition 1).

A.1 Proof of Proposition 1

The validity of Proposition 1 follows from the proof of Proposition 3 if we set $d_x = d_y = d$.

A.2 Proof of Proposition 2

Suppose that $(s_x = 1, s_y = 0)$ and $(s_x = 0, s_y = 1)$ are evolutionarily stable states. We simply note that $W(1,0) - W(0,1) = (2\beta - 1)(b - c) > 0$ for $\beta > 0.5$, which shows the first claim of the proposition.

Suppose that $(s_x, s_y) = (0,0)$ is also evolutionarily stable. To show the validity of the first bullet, it is enough to observe that $W(0,1) - W(0,0) = (1-\beta)(b-c) + 2\beta(1-\beta)pd > 0$, since $b > c$.

Suppose instead that $(s_x, s_y) = (1,1)$ is also evolutionarily stable. We note that $W(1,0) - W(1,1) = -(1-\beta)(b-c) + 2\beta(1-\beta)dp > 0$ requires $dp > \frac{(b-c)}{2\beta}$, while $W(0,1) - W(1,1) = -\beta(b-c) + 2\beta(1-\beta)dp > 0$ requires $dp > \frac{(b-c)}{2(1-\beta)}$, i.e., the two conditions in the second bullet.

A.3 Proof of Proposition 3

Preliminarily, we argue that to prove that a monomorphic or a type-monomorphic state is evolutionarily stable, it is sufficient to check for pure invasions, i.e., $\sigma_x, \sigma_y \in \{0,1\}$, that is for invasions such that all x-type mutants are either only cooperators or only defectors and similarly for y-type mutants.

To see why, consider a generic state $s = (s_x, s_y)$ and an invasion of $\tilde{\epsilon}_x$ and $\tilde{\epsilon}_y$ mutants such that $0 < \tilde{\sigma}_x, \tilde{\sigma}_y < 1$. Hence, the new state $s' = (s'_x, s'_y)$ is such that $s'_x = s_x(1 - \tilde{\epsilon}_x/\beta) + \tilde{\sigma}_x\tilde{\epsilon}_x/\beta$ and $s'_y = s_x(1 - \tilde{\epsilon}_y/(1-\beta)) + \tilde{\sigma}_y\tilde{\epsilon}_y/(1-\beta)$. Consider (without loss of generality) the case where $\tilde{\sigma}_x > s_x$ and $\tilde{\sigma}_y < s_y$. Since the $\tilde{\epsilon}_x$ mutants are $\tilde{\sigma}_x\tilde{\epsilon}_x$ cooperators and $(1-\tilde{\sigma}_x)\tilde{\epsilon}_x$ defectors, and similarly for the $\tilde{\epsilon}_y$ mutants, we can rewrite the new state as $s'_x = s_x + \epsilon_x/\beta$ and $s'_y = s_y + \epsilon_y/(1-\beta)$, where $\epsilon_x = (\tilde{\sigma}_x - s_x)\tilde{\epsilon}_x$ and $\epsilon_y = (\tilde{\sigma}_y - s_y)\tilde{\epsilon}_y$. We observe that the expected payoff of $\tilde{\epsilon}_x - \epsilon_x$ mutants, whose fraction of cooperators is s_x, is the same as the incumbents of type x. Similarly, the expected payoff of $\tilde{\epsilon}_y - \epsilon_y$ mutants, whose fraction of cooperators is s_y, is the same as the incumbents of type y. Therefore, the success of the invasion crucially depends on whether ϵ_x cooperators and ϵ_y defectors earn a greater expected payoff than their respective incumbent types. From this observation follows that the original invasion is successful if and only if a smaller pure invasion of ϵ_x and ϵ_y mutants, with $\sigma_x = 1$ and $\sigma_y = 0$, is successful.

We consider the state $s = (s_x, s_y) = (1,1)$, and we suppose that a small fraction $\epsilon = \epsilon_x + \epsilon_y$ of mutants invades and a state $(s_x - \frac{\epsilon_x}{\beta}, s_y - \frac{\epsilon_y}{1-\beta})$ is reached. The following expressions (2) and (3) are, respectively, the relative gain that y-type cooperators have over y-type defectors, and the relative gain that x-type cooperators have over x-type defectors:

$$\pi(C, y|(s_x - \frac{\epsilon_x}{\beta}, s_y - \frac{\epsilon_y}{1-\beta})) - \pi(D, y|(s_x - \frac{\epsilon_x}{\beta}, s_y - \frac{\epsilon_y}{1-\beta})) = p\left(b - \frac{\beta - \epsilon_x}{1 - \epsilon_x - \epsilon_y}d_y + \frac{\epsilon_x}{\epsilon_x + \epsilon_y}d_y\right) - c; \quad (2)$$

$$\pi(C,x|(s_x-\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))-\pi(D,x|(s_x-\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))=p\left(b-\frac{1-\beta-\epsilon_y}{1-\epsilon_x-\epsilon_y}d_x+\frac{\epsilon_y}{\epsilon_x+\epsilon_y}d_x\right)-c. \quad (3)$$

The worst case for expression (2) to be positive is when $\epsilon_x = 0$ and $\epsilon_y = \epsilon$. Analogously, the worst case for expression (3) to be positive is when $\epsilon_y = 0$ and $\epsilon_x = \epsilon$. Hence, it is easy to check that there exists an invasion barrier $\bar{\epsilon} > 0$ such that for any (ϵ_x, ϵ_y), with $\epsilon_x \geq 0$, $\epsilon_y \geq 0$, and $0 < \epsilon_x + \epsilon_y < \bar{\epsilon}$, expressions (2) and (3) are both positive if and only if $p \geq \frac{c}{b-\max\{\beta d_y,(1-\beta)d_x\}}$. This shows the validity of the first bullet in the statement of the proposition.

We now consider the state $s = (s_x, s_y) = (0, 0)$, and we suppose that a small fraction $\epsilon = \epsilon_x + \epsilon_y$ of mutants invades and a state $(s_x + \frac{\epsilon_x}{\beta}, s_y + \frac{\epsilon_y}{1-\beta})$ is reached. The following expressions (4) and (5) are, respectively, the relative gain that y-type defectors have over y-type cooperators, and the relative gain that x-type defectors have over x-type cooperators:

$$\pi(D,y|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(C,y|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left(-b-\frac{\beta-\epsilon_x}{1-\epsilon_x-\epsilon_y}d_y+\frac{\epsilon_x}{\epsilon_x+\epsilon_y}d_y\right)+c; \quad (4)$$

$$\pi(D,x|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(C,x|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left(-b-\frac{1-\beta-\epsilon_y}{1-\epsilon_x-\epsilon_y}d_x+\frac{\epsilon_y}{\epsilon_x+\epsilon_y}d_x\right)+c. \quad (5)$$

The worst case for expression (4) to be positive is when $\epsilon_x = 0$ and $\epsilon_y = \epsilon$. Analogously, the worst case for expression (5) to be positive is when $\epsilon_y = 0$ and $\epsilon_x = \epsilon$. Hence, it is easy to check that there exists an invasion barrier $\bar{\epsilon} > 0$ such that for any (ϵ_x, ϵ_y), with $\epsilon_x \geq 0$, $\epsilon_y \geq 0$, and $0 < \epsilon_x + \epsilon_y < \bar{\epsilon}$, expressions (4) and (5) are both positive if and only if $p \leq \frac{c}{b+\max\{\beta d_y,(1-\beta)d_x\}}$, which shows the validity of the second bullet in the statement of the proposition.

We then consider the state $s = (s_x, s_y) = (1, 0)$, and we suppose that a small fraction $\epsilon = \epsilon_x + \epsilon_y$ of mutants invades and a state $(s_x - \frac{\epsilon_x}{\beta}, s_y + \frac{\epsilon_y}{1-\beta})$ is reached. The following expressions (6) and (7) are, respectively, the relative gain that y-type defectors have over y-type cooperators, and the relative gain that x-type cooperators have over x-type defectors:

$$\pi(D,y|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(C,y|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left(-b+\frac{\beta-\epsilon_x}{\beta-\epsilon_x+\epsilon_y}d_y-\frac{\epsilon_x}{1-\beta+\epsilon_x-\epsilon_y}d_y\right)+c; \quad (6)$$

$$\pi(C,x|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(D,x|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left(b+\frac{1-\beta-\epsilon_y}{1-\beta-\epsilon_y+\epsilon_x}d_x-\frac{\epsilon_y}{\beta-\epsilon_x+\epsilon_y}d_x\right)-c. \quad (7)$$

The worst case for expression (6) to be positive is when $\epsilon_y = 0$ and $\epsilon_x = \epsilon$. To see why, plug $\epsilon_y = \epsilon - \epsilon_x$ into RHS of expression (6) and differentiate it with respect to ϵ_x. One can check that when ϵ is sufficiently small, the derivative is negative. Analogously, the worst case for expression (7) to be positive is when $\epsilon_x = 0$ and $\epsilon_y = \epsilon$. Hence, it is easy to check that there exists an invasion barrier $\bar{\epsilon} > 0$ such that for any (ϵ_x, ϵ_y), with $\epsilon_x \geq 0$, $\epsilon_y \geq 0$, and $0 < \epsilon_x + \epsilon_y < \bar{\epsilon}$, expressions (6) and (7) are both positive for any $\epsilon > 0$ small enough if and only if $\frac{c}{b-d_y} \geq p \geq \frac{c}{b+d_x}$, which shows the validity of the third bullet in the statement of the proposition.

Lastly, we consider the state $s = (s_x, s_y) = (0, 1)$, and we suppose that a small fraction $\epsilon = \epsilon_x + \epsilon_y$ of mutants invades and a state $(s_x + \frac{\epsilon_x}{\beta}, s_y - \frac{\epsilon_y}{1-\beta})$ is reached. The following expressions (8) and (9) are, respectively, the relative gain that y-type cooperators have over y-type defectors, and the relative gain that x-type defectors have over x-type cooperators:

$$\pi(C,y|(s_x+\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))-\pi(D,y|(s_x+\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))=p\left(b+\frac{\beta-\epsilon_x}{\beta-\epsilon_x+\epsilon_y}d_y-\frac{\epsilon_x}{1-\beta+\epsilon_y-\epsilon_x}d_y\right)-c, \quad (8)$$

$$\pi(D,x|(s_x+\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))-\pi(C,x|s_x+\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))=p\left(-b+\frac{1-\beta-\epsilon_y}{1-\beta-\epsilon_y+\epsilon_x}d_x-\frac{\epsilon_y}{\beta-\epsilon_x+\epsilon_y}d_x\right)+c, \tag{9}$$

The worst case for expression (8) to be positive is when $\epsilon_y = 0$ and $\epsilon_x = \epsilon$. Analogously, the worst case for expression (9) to be positive is when $\epsilon_x = 0$ and $\epsilon_y = \epsilon$. Hence, it is easy to check that there exists an invasion barrier $\bar{\epsilon} > 0$ such that for any (ϵ_x, ϵ_y), with $\epsilon_x \geq 0$, $\epsilon_y \geq 0$, and $0 < \epsilon_x + \epsilon_y < \bar{\epsilon}$, expressions (8) and (9) are both positive if and only if $\frac{c}{b-d_x} \geq p \geq \frac{c}{b+d_y}$, which shows the validity of the third bullet in the statement of the proposition.

Finally, we show that no other state can ever be evolutionarily stable. Ad absurdum, suppose that a state (s_x, s_y) is evolutionarily stable and that $s_x \in (0,1)$ (this is without loss of generality). In such a state the expected payoff of x-type cooperators must be equal to the expected payoff of x-type defectors. Now consider an invasion of ϵ mutants, all being x-type cooperators, i.e., $\epsilon_x = \epsilon$ and $\epsilon_y = 0$. Denote with $(s_x+\frac{\epsilon}{\beta}, s_y)$ the resulting state. From $\pi(C,x|(s_x,s_y)) = \pi(D,x|(s_x,s_y))$, it follows that $p(b-d_x(\eta_{y|C}(s_x,s_y)-\eta_{y|D}(s_x,s_y)))-c=0$. Since $\eta_{y|C}(s_x+\frac{\epsilon}{\beta},s_y)-\eta_{y|D}(s_x+\frac{\epsilon}{\beta},s_y) < \eta_{y|C}(s_x,s_y)-\eta_{y|D}(s_x,s_y)$, we have that, for any $\epsilon > 0$, an x-type cooperator obtains a strictly greater payoff than a x-type defector, i.e.:

$$\pi(C,x|(s_x+\frac{\epsilon}{\beta},s_y))-\pi(D,x|(s_x+\frac{\epsilon}{\beta},s_y))=p(b-d_x(\eta_{y|C}(s_x+\frac{\epsilon}{\beta},s_y)-\eta_{y|D}(s_x+\frac{\epsilon}{\beta},s_y)))-c>0, \tag{10}$$

which implies that mutants obtain a higher payoff than incumbents, in contrast with (s_x, s_y) being evolutionarily stable.

A.4 Proof of Proposition 4

The validity of Proposition 4 can be shown along the lines of the proof of Proposition 3. In the following we limit ourselves to highlight the differences and provide brief comments.

In place of (2) and (3), we have:

$$\pi(C,y|(s_x-\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))-\pi(D,y|(s_x-\epsilon_x,s_y-\epsilon_y))=p\left[b+(1-q)\left(-\frac{\beta-\epsilon_x}{1-\epsilon_x-\epsilon_y}d+\frac{\epsilon_x}{\epsilon_x+\epsilon_y}d\right)\right]-c; \tag{11}$$

$$\pi(C,x|(s_x-\frac{\epsilon_x}{\beta},s_y-\frac{\epsilon_y}{1-\beta}))-\pi(D,x|(s_x-\epsilon_x,s_y-\epsilon_y))=p\left[b+(1-q)\left(-\frac{1-\beta-\epsilon_y}{1-\epsilon_x-\epsilon_y}d+\frac{\epsilon_y}{\epsilon_x+\epsilon_y}d\right)\right]-c; \tag{12}$$

from which, taking into account that $\beta > 1-\beta$ since $\beta \in (0.5, 1)$, we can derive the necessary and sufficient condition for evolutionarily stability, $p \geq \frac{c}{b-(1-q)\beta d}$, which shows the validity of the first bullet in the statement of the proposition.

In place of (4) and (5), we have:

$$\pi(D,y|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(C,y|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left[-b+(1-q)\left(-\frac{\beta-\epsilon_x}{1-\epsilon_x-\epsilon_y}d+\frac{\epsilon_x}{\epsilon_x+\epsilon_y}d\right)\right]+c; \tag{13}$$

$$\pi(D,x|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(C,x|(s_x+\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left[-b+(1-q)\left(-\frac{1-\beta-\epsilon_y}{1-\epsilon_x-\epsilon_y}d+\frac{\epsilon_y}{\epsilon_x+\epsilon_y}d\right)\right]+c; \tag{14}$$

from which, taking into account that $\beta > 1-\beta$ since $\beta \in (0,1)$, we can derive the necessary and sufficient condition for evolutionarily stability, $p \leq \frac{c}{b+(1-q)\beta d}$, which shows the validity of the second bullet in the statement of the proposition.

In place of (6) and (7), we have:

$$\pi(D,y|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(C,y|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left[-b+(1-q)\left(+\frac{1-\beta-\epsilon_x}{1-\beta-\epsilon_x+\epsilon_y}d-\frac{\epsilon_x}{\beta+\epsilon_x-\epsilon_y}d\right)\right]+c; \tag{15}$$

$$\pi(C,x|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))-\pi(D,x|(s_x-\frac{\epsilon_x}{\beta},s_y+\frac{\epsilon_y}{1-\beta}))=p\left[b+(1-q)\left(\frac{1-\beta-\epsilon_y}{1-\beta-\epsilon_y+\epsilon_x}d-\frac{\epsilon_y}{\beta-\epsilon_x+\epsilon_y}d\right)\right]-c; \tag{16}$$

from which we can derive the necessary and sufficient condition for evolutionarily stability, $\frac{c}{b-(1-q)d} > p > \frac{c}{b+(1-q)d}$. In place of (8) and (9), we can write analogous expressions, which however coincide with (15) and (16), since here we are considering the case of symmetric cultural intolerance. Hence, we have shown the validity of the third bullet in the statement of the proposition.

Finally, the same argument used at the end of the proof of Proposition 3 can be used to show that no other state can ever be evolutionarily stable, with the only difference that: $\pi(C,x|(s_x,s_y)) = \pi(D,x|(s_x,s_y))$ implies that $p[b-(1-q)d(\eta_{y|C}(s_x,s_y)-\eta_{y|D}(s_x,s_y))]-c=0$, and $\pi(C,x|(s_x+\frac{\epsilon}{\beta},s_y))-\pi(D,x|(s_x+\frac{\epsilon}{\beta},s_y)) = p[(b-(1-q)d(\eta_{y|C}(s_x+\frac{\epsilon}{\beta},s_y)-\eta_{y|D}(s_x+\frac{\epsilon}{\beta},s_y))]-c \geq 0$.

A.5 Proof of Proposition 5

The validity of Proposition 5 can be shown by considering expressions from (2) to (7) re-written for the special case $d_x = d_y = d$ and substituting p with $a(s')$, where s' is the new state resulting from the considered invasion made of ϵ_x mutants of type x and ϵ_y mutants of type y. In particular, to establish the truth of the three bullets of Proposition 5, it is enough to follow the same arguments applied in the proof of Proposition 3 with the only difference that the claims about the signs of (2)-(7) have to be read as proving the sufficiency of the conditions involved (instead of both necessity and sufficiency).

Moreover, to establish that no other state can ever be evolutionarily stable, we can adjust the argument used in the last paragraph of the proof of Proposition 3, for which the following observation allows to apply the same reasoning. Let s be the original state supposed, ad absurdum, to be evolutionarily stable and let s' be the state resulting from an invasion of $\epsilon_x+\epsilon_y$ mutants. We observe that one can always consider the case where x-types are cooperators and y-types are defectors (or vice versa) where ϵ_x/ϵ_y is such that $a(s) = a(s')$, i.e., the fraction of cooperators in the whole population does not change. This allows to treat the index of assortativity in actions as fixed for the purpose of establishing the success of the invasion.

A.6 Proof of Proposition 6

By using equation 1, we can compute:

$$a(s_x = 0, s_y = 0) = 0 \tag{17}$$

$$a(s_x = 1, s_y = 0) = \frac{\beta(1-\beta)(n^2 - m^2)}{\beta(1-\beta)(n-m)^2 + nm} \tag{18}$$

$$a(s_x = 0, s_y = 1) = \frac{\beta(1-\beta)(n^2 - m^2)}{\beta(1-\beta)(n-m)^2 + nm} \tag{19}$$

$$a(s_x = 1, s_y = 1) = 0 \tag{20}$$

We can then verify whether the conditions of Proposition 5 are satisfied, thus obtaining the statement of the proposition.