Culture Clash: Incompatible Reputation Systems and Intergroup Conflict*

Vasiliki Fouka† Alain Schläpfer‡

December 2019

Abstract

Under which conditions does intergroup contact lead to conflict? We provide a novel answer to this question by highlighting the role of reputation mechanisms in sustaining cooperation. Punishment-based reputation (a "culture of honor") and reputation based on image scoring (indirect reciprocity) can both deter defection in one-time interactions within groups. Yet these reputation mechanisms can be incompatible in intergroup interactions. Using a game theoretic model, we show that injecting pools of individuals from a punishment-based culture into a culture of image scoring can lead to widespread intergroup conflict. Cooperation is a more likely outcome if the cultures that interact use a similar reputation mechanism. The theoretical framework helps us explain a variety of phenomena, such as variation in immigrant crime rates and patterns of outgroup discrimination.

Keywords: cooperation, reputation, indirect reciprocity

*For this early draft, we have greatly benefited from helpful discussions with Jim Fearon and David Laitin.

†Stanford University, Department of Political Science, 616 Serra Street, Stanford, CA 94305, USA. Email: vfouka@stanford.edu.

‡Santa Clara University, Department of Economics, 500 El Camino Real, Santa Clara, CA 95053, USA. Email: aschlapfer@scu.edu.

Introduction

The study of cooperation and conflict is central in the social sciences. Political scientists have strived to explain why and when conflict occurs, identifying factors such as differential access to resources (Bates 1983; Gellner 1983), the role of formal and informal institutions and political entrepreneurs (Fearon and Laitin 1996; Gagnon Jr 1994; Laitin 1995) or group identity (Sambanis and Shayo 2013). Economists and evolutionary biologists have focused on a different side of the same puzzle, trying to understand how unrelated individuals manage to sustain large-scale cooperation in one-time interactions (Nowak and Sigmund 1998b; Fehr and Gächter 2000; Fehr and Fischbacher 2003; Dal Bó 2005; Henrich et al. 2006). A central finding in these literatures is that the emergence of cooperation can be explained by the role of reputation. Here, we advance the hypothesis that reputational considerations that can sustain cooperation within groups can lead to the breakdown of cooperation across groups. In particular, incompatibility in terms of the role that reputation plays in different societies can be one of the drivers of intergroup conflict.

The informational content of reputations varies across cultures. In punitive reputation systems, individuals punish defectors in order to gain a reputation as being "tough", ensuring that future co-players will cooperate (Sigmund, Hauert, and Nowak 2001; Brandt, Hauert, and Sigmund 2003; dos Santos, Rankin, and Wedekind 2011). In cultures with image scoring, or cooperation-based reputation, defectors gain a negative image, triggering other group members to avoid them in future interactions (Nowak and Sigmund 1998a; Milinski, Semmann, and Krambeck 2002; Sylwester and Roberts 2013). Previous research has shown that both these systems can independently support cooperation within a group (Schläpfer 2018) and evidence from cross-cultural research suggests variation across societies in terms of the relative importance of each reputation system (Leung and Cohen 2011).

What happens when individuals from a society reliant on punishment-based reputation find themselves in a society of image scoring? With globalization, the lowering of transportation costs and rising immigration flows, intergroup contact is becoming more frequent. Unlike theories suggesting positive effects of contact on intergroup relations (Allport 1954), we propose that cultural incompatibility in terms of reputation systems can hinder intergroup cooperation and promote de-

viant behavior, avoidance and conflict. The intuition behind this is simple. In punishment-based cultures, individual punishment of deviant behavior is encouraged. Individuals who don't punish are considered "weak" and can be exploited. In societies of image scoring cooperative behavior is prescribed directly, with no immediate role for individual punishment. Members of a society with punishment-based reputation then lack incentives to cooperate in interactions with members of a society of image scoring, triggering the latter to avoid such interactions altogether.

We build a game theory model based on a social matching game to formalize these ideas. In the model, players are matched in pairs each period to play a prisoner's dilemma followed by a punishment phase. The number of players in a group is large so that the same pair rarely interacts twice, implying that direct reciprocity is unlikely to sustain cooperation. We first derive conditions under which either type of reputation system can support a cooperative equilibrium within a group. We then consider interactions across group boundaries, with each group relying on their distinct reputational mechanism. We show that, as long as both groups use the same type of reputation, cooperation across groups can be sustained under conditions that are somewhat tighter than in the within group case. When both groups rely on cooperation-based reputation, players cooperate in intergroup interactions to maintain the reputation of being "nice" within their own group. Similarly, when reputation is punishment-based in both groups, players punish in intergroup interactions in order to appear "tough" to their own peers, supporting an equilibrium of mutual cooperation. Such a cooperative equilibrium does not exist if groups differ in terms of their reputation systems, since members of the punishment-based group have no incentives to cooperate in intergroup interactions.

We show that the model's predictions are compatible with a number of empirical patterns. Variation in imprisonment rates of second generation immigrants in the US can be explained by origin country averages in the importance of being respected, which proxies for reliance on a punishment-based reputation. This holds even after controlling for GDP per capita and crime rates in the country of origin. The distance between countries in terms of the importance placed on respect is a predictor of bilateral discrimination. Consistent with the model, discrimination is more pronounced among members of relatively more cooperation-based societies towards members

of relatively more punishment-based societies, rather than vice versa. These patterns cannot be explained by bilateral economic differences or differences between countries in other dimensions of culture.

Cross-cultural differences in the role of punishment have been studied in the context of behavioral experiments, most notably by Henrich et al. (2006). Herrmann, Thöni, and Gächter (2008) demonstrate substantial variation in the role of anti-social punishment across societies in the context of a public goods game. Since punishment is meted anonymously, this study does not allow a role for reputational considerations in shaping punishment decisions. Closer to our setup, Brooks, Hoff, and Pandey (2018) show that reputational considerations can impede coordination on socially optimal conventions across, but also within cultural groups.

This study contributes to at least four additional strands of literature. First, studies on reputation-based cooperation have focused on showing that indirect reciprocity (Nowak and Sigmund 1998a; Milinski, Semmann, and Krambeck 2002; Sylwester and Roberts 2013) and punishment (Sigmund, Hauert, and Nowak 2001; Brandt, Hauert, and Sigmund 2003; dos Santos, Rankin, and Wedekind 2011) can sustain cooperation in one-shot interactions. The possibility that both these types of reputation can co-exist and evolve endogenously has been established theoretically by Schläpfer (2018). To our knowledge, no study considers the theoretical implications of "mixing" of reputation systems in the context of intergroup contact, and specifically the pernicious effects that mixing can have for intergroup cooperation.

Second, several theories of conflict consider the role of differences across groups that can lead to clashes. These can range from access to power (Horowitz 1985), economic opportunities (Bates 1983; Gellner 1983) or ethnic identification (Sambanis and Shayo 2013). To these explanations we add the role of cultural differences, focusing specifically on the mechanisms through which each group solves the problem of ingroup cooperation.

Third, our conclusions have implications for the study of intergroup conflict in the context of immigration. Various studies on attitudes towards immigrants and refugees show that citizens of western countries tend to discriminate against Middle Eastern, Balkan and Eastern European immigrants (Hainmueller and Hangartner 2013; Bansak, Hainmueller, and Hangartner 2016). Along-

side religion, our paper identifies cultural incompatibility (Huntington 1996) as an additional factor that can drive conflict between immigrants and natives.

Finally, the paper relates to a growing literature in economics and political science on the formation and implications of culture. Several studies have tied cross-societal differences in patterns of cooperation and conflict to differences in social structure (Moscona, Nunn, and Robinson 2017; Moscona, Nunn, and Robinson 2019; Enke 2019) or underlying payoff structures that vary across societies (Bednar and Page 2007; Bednar and Page 2018). We provide a complementary explanation for such differences based on the role of reputation. We also extend the scope of the research question, from identifying variation in culture, to studying the implications of this variation for cross-cultural contact.

The rest of the paper proceeds as follows. We start by presenting qualitative evidence on cross-cultural variation in reputation systems. We then discuss cases in which the interaction of punishment-based and cooperation-based societies leads to deviant behavior on the part of the former and avoidance on the part of the latter, often resulting in intergroup conflict. Next, we present the theoretical model and empirical patterns consistent with its predictions. The final section concludes and discusses paths to future research.

Punishment-based and cooperation-based reputation

Reputational considerations have been identified as a fundamental driver of human cooperation by a large literature in the social sciences (Fearon and Laitin 1996; Milinski, Semmann, and Krambeck 2002; Dal Bó 2005). Cooperation can be efficiently sustained through different types of reputation systems. On the one hand, building the reputation of someone who punishes transgressions against them can effectively deter cheating in future interactions. On the other, cooperation can be achieved if individuals build a reputation of someone who cooperates, and avoid interactions with non-cooperators. Schläpfer (2018) shows that these are two distinct reputation systems that can sustain different cooperative equilibria.

The best known examples of punishment-based reputation systems are cultures of honor. In societies of honor, individuals (primarily men) are sensitive to insult, attacks on family (especially

female family members) or property, and are willing to use violence to defend their reputation against those who slight them. Failure to punish personal insult results in a loss of reputation that is shameful for the individual and implies loss of status in the community. Punishment can take the form of honor killings, and can escalate into vendettas, or tit-for-tat situations in which the punisher (or his family members) are punished by the punishee. Societies of honor have been studied by ethnographers in the Mediterranean (Campbell 1964; Boehm 1984), the Middle East (Stewart 1994) and Central Asia (Keiser 1991), and are commonly associated with subsistence systems reliant on pastoralism. Cultural psychologists Nisbett and Cohen have identified a culture of honor descending from Scotch-Irish farmers as one of the roots of violence in the US South (Nisbett and Cohen 1996; Grosjean 2014). While violence and blood killings have been given much emphasis in the anthropological literature, they correspond to deviations rather than to the everyday experience of honor-based societies. The threat of punishment is what sustains cooperation and deters transgressors. The sociological study of Gould (2000) in Corsica, an honor society known for vendettas, demonstrated that the frequency of honor killings was very low in practice.

Leung and Cohen (2011) identify a different reputation system in *cultures of face*. In such cultures, often thought to be characteristic of Asian societies, a person's dignity is defined not by their ability to defend their honor, but by behaving appropriately within the system in which they are embedded. "Losing face" (Hamamura et al. 2009) and the associated shame is in such cases the threat that sustains proper behavior. Unlike in honor cultures, it is not incumbent on victims of aggression to redress grievances themselves (Leung and Cohen 2011). Abstinence from punishment of transgressors is also a characteristic of *cultures of dignity*. In response to offense, an aggrieved member of a culture of dignity may "exercise covert avoidance, quietly cutting off relations with the offender without any confrontation" (Campbell and Manning 2014, 2018). Avoidance and toleration, for example of accidental personal injuries, distinguish cultures of face and dignity from cultures of honor. Unlike societies of face, societies of dignity resort to these behaviors not out of reputational considerations, but out of internalized moral norms (Kim and Cohen 2010). Reputational considerations and avoidance of transgressors may, however, coexist as strategies

sustaining cooperation.

Cross-cultural interactions and cultural incompatibility

What happens when societies reliant on different reputation systems interact? A characteristic example is provided by conflicts between pastoralists and agriculturalists, the former relatively more reliant on punishment to sanction offenders and maintain honor (Figueredo et al. 2004). Farmer-herder conflicts have been documented since antiquity (Turchin 2005) and are still a perennial problem in parts of the world, such as the region of the Sahel in Africa. As examples of low-intensity conflicts (LIC), they have been extensively analyzed by geographers, anthropologists and political scientists (Little 1992; Thébaud and Batterbury 2001; Turner 2004). While climate, uneven access to state institutions and fights over resources have been documented as triggers of these conflicts, participants themselves have identified cultural differences as additional causal factors (Ofuoku and Isife 2009). Surveys of farmers and herders often reveal damage of crops initiated by pastoral groups as the point of onset of the conflict (Turner et al. 2006). To such actions, farmers often reply with fencing and cutting access of herders to their crops and water resources. And while, sometimes extreme, violence of herders has often been documented (Benjaminsen, Maganga, and Abdallah 2009), equivalent retaliating actions on the part of farmers are rarer. In certain cases farmers have migrated out of a region in order to avoid herders' violence against them (Tonah 2006). This example illustrates a pattern that we analyze more systematically in the rest of the paper: when societies reliant on punishment interact with societies reliant on image scoring, the former have an incentive to commit offenses against the latter, since punishment does not anymore function as deterrent. Image scoring societies respond with avoidance and shunning. The ensuing conflict is not due to the higher propensity of pastoral societies to commit violence. As Ellickson (1991) shows in the case of rural Shasta county in California, herding societies are able to enforce norms of cooperation and face instances of within group conflict only when the group is infiltrated by outsiders, unfamiliar with those norms.

Aggression, avoidance, and the breakdown of cooperation ensue in other situations of cross-cultural contact, such as immigration. Often, that is the case when immigrants from punishment-

based societies move to cooperation-based societies commonly found in the Western world. The lack of *acculturation* has been identified as a candidate reason behind immigrant delinquency in the sociological literature (Vega et al. 1993). While the concept can broadly refer to psychological distress from dealing with foreign norms of conduct or to transgressions that happen due to lack of familiarity with a new culture, it also applies to a more explicit clash between the cultures of natives and newcomers. One such example, the famous ethnography "Street Corner Society" describes this clash between the social structure of immigrants from rural Italy and that of the Anglo-saxon society of Boston (Whyte 2012). Sellin (1938) formulated the "culture-conflict" hypothesis that states that a group's crime rate is positively correlated with the extent of incongruity between the group's legal norms, folkways and values, and that of the host community. Shoham (1962) provides evidence for this hypothesis in the context of Israel. He observes higher crime rates, specifically homicide rates, among African immigrants and attributes these findings not only to cultural distance – which was highest between that group and Israeli society compared to immigrants from e.g. Europe – but also to "the existence of a cultural tradition among the African immigrants of settling disputes by violence, a method of "self-help" which may have been more or less accepted conduct in their countries of origin" (Shoham 1962). Similar patterns have been documented in other countries. Aronowitz (2002) finds higher delinquency rates among Turkish immigrants in Berlin who score higher in a *Turkish identity* index that measures identification with traditional Turkish values, such as the importance of honor. Junger-Tas (2001) and Bovenkerk and Fokkema (2016) note the disproportional criminality rates among Morrocans, another culture of honor, in the Netherlands (a cooperation-based society). Importantly for our argument, delinquency appears to be curtailed whenever the group manages to maintain tight social control over its members by forming a form of effective ghetto (Aronowitz 2002). That is, conditional on the characteristics of the immigrant group, delinquency increases the more opportunities group members have to interact with the broader society.

Model

We consider a group of people that are playing a social matching game. Each period, members of the group are randomly paired to play an extended prisoner's dilemma game that consists of two rounds. In the first round, players in each pair simultaneously decide whether or not to contribute 1 to a joint project. Combined contributions are then multiplied with R, $2 > R > 1$, and the result equally split among the two players. Besides contributing (c) and non-contributing (d) players can also unilaterally call off the first round (a), in which case both players earn zero profits. The resulting first round playoff matrix is then

Table 1: First stage game

	c	d	a
c	R-1,R-1	R/2-1,R/2	0,0
d	R/2,R/2-1	0,0	0,0
a	0,0	0,0	0,0

In the first round, jointly contributing maximizes the total payoffs that the game creates, but players have incentives to opt for the opportunistic choice d since $R/2 > R - 1$ and $0 > R/2 - 1$. We study how a reputation mechanism can help groups to achieve the preferred outcome of mutual cooperation. The option of avoiding first round interactions a gives players the possibility to avoid the worst case outcome $R/2 - 1$ without appearing opportunistic, and will play an important role in a reputational equilibrium.

After the first round game is played and outcomes are observed, each pair immediately plays a second round, where players can choose to punish their opponent. Punishment bears a cost $k > 0$ to punisher and inflicts damage $d > k$ to the punishee. Punishment by itself is undesirable, as it reduces the payoffs of both participants, but as will become clear, the threat of punishing non-contributors can help to incentivize players to cooperate in the first round.

We consider two types of rare distortions among players that alter payoff functions. Firstly, there are *bad* types, that have larger costs of contributing during the first stage than *nice* types of $\xi > 1$. We think of ξ as a number that is large enough so that these players do not contribute to projects in any equilibrium. Secondly, players of the *weak* type have higher cost of punishing than *tough* types, given by μk, $\mu > 1$. Again, we think of μ being a large number. We denote the

share of *bad* types in a group with p and the share of *weak* types with q. These two characteristics are independent, so that the share of *nice* and *tough* players, for which the first stage payoffs are shown in Table 1, is given by $(1-p)(1-q)$. We further assume that players are infinitely lived and maximize expected payoffs valued with a time discount factor $\delta \leq 1$.

We assume that groups are large so that the same individuals are rarely matched twice. This ensures that direct reciprocity – that is, cooperating with an opponent if this opponent has cooperated in the same pairing in the past – can only play a minor role, and we fully abstract from it. Cooperation can instead be sustained through indirect reciprocity. Groups can use a reputation mechanism to induce *nice* players to cooperate and contribute. We view reputation as a collective institution, which requires members of the group to observe, record, and transmit information about past actions of all group members. A reputation mechanism is thus a group-specific characteristic. We consider two types of reputation mechanisms: cooperation-based reputation that relies on information about players' actions during the first round of games, and punishment-based reputation that draws from information about second round actions. We now consider each mechanism in turn, assuming that a group does not have sufficient informational capacity to use both types of reputation at the same time, but instead has to rely on one of the two.

Sustaining cooperative equilibria with reputation

We first consider how a cooperation-based reputation (coop-rep) can help a group to sustain cooperation among *nice* players. Such a reputation can be formed if a group keeps track of first round actions of its members. We consider a simple binary form of image scoring, where a player's reputation turns positive (*coop+*) if they play c and contribute, and it turns negative (*coop-*) if they play d. If any of the two players chooses to avoid a, then the first round is called off and neither player's reputation is updated. A player's reputation is publicly available, so that every player can condition their actions on the reputation of their opponent.

Given a sufficiently high value of μ, *bad* players will never contribute, but we can derive conditions under which an equilibrium exists where *nice* players contribute whenever they are paired with each other. Consider a strategy profile where *nice* players play c with any opponent

with *coop+*, and they avoid first round interactions with *coop-* players. Since punishment is costly and does not yield any benefits, no player will punish in the second round.

If all players follow the above strategy, then *nice* players will have *coop+* and *bad* players will have *coop-* at all times. We can then derive conditions so that *nice* players have no incentives to deviate. The only relevant possible deviation for a *nice* player is to defect against a *coop+* player by playing d.[1] In the short run, this gives a gain in payoffs of $1 - R/2 > 0$, but comes at the cost of acquiring a bad public image *coop-*.

If in the following round, the player is paired with a *nice* player, which happens with probability $1 - p$, then because of the negative image *coop-*, the opponent will play a instead of c, leading to a loss in payoffs of $R - 1$. Furthermore, the reputation will not be updated until the player meets a *bad* opponent, in which case the player can choose c (insteady of the usual a) at a payoff loss of $1 - R/2$ and restore their reputation to *coop+*. In total, the expected loss that arises from a damaged reputation is

$$\begin{aligned}&\delta(p(1-R/2)+(1-p)(R-1)[\delta(p(1-R/2)+(1-p)(R-1)[...\\&=\frac{\delta}{1-\delta(1-p)}(p(1-R/2)+(1-p)(R-1))\end{aligned}$$

For no deviations, we then need that the short run gains of deviation are smaller than the expected losses that arise from a damaged reputation, in particular that

$$1-R/2 \leq \frac{\delta}{1-\delta(1-p)}(p(1-R/2)+(1-p)(R-1))$$

or equivalently

$$(1-p) \geq \left(\frac{1-\delta}{\delta}\right)\left(\frac{1-R/2}{R-1}\right) \tag{1}$$

It follows that an equilibrium where *nice* players cooperate with each other (and avoid inter-

[1]It is straightforward to check that any other deviation from the prescribed strategy, that is playing a against a *coop+* player or playing c or d against a *coop-* player never increase payoffs.

actions with *bad* players) is feasible as long as the share of *nice* players is sufficiently large, the discount factor δ sufficiently high, and returns to the common project R sufficiently large. In this equilibrium, the rate of first round contributions will be given by $(1-p)^2$.

We next turn to a group where individuals have a reputation based on their second round actions during the stage game. Players can use second round punishment to build a reputation for being *tough*, indicating that they will punish any future opponents that do not contribute to the joint project. This is only possible if the group records each player's second round response to their opponent's first round actions. We refer to this as a punishment-based reputation mechanism (pun-rep), which we again consider in a very simple binary form. If a player's opponent plays either d or a in the first round, and the player punishes in the second round, then the player gains a positive reputation *pun+*. The player gains a negative reputation *pun−* if they fail to punish in response to an opponent playing d or a. This reputation is not updated whenever the opponent played c in the first round.

Consider the following strategies: *nice* players play c with *pun+* and d wit *pun-* opponents, while *bad* players always play d. In the second stage *tough* players punish whenever their opponent played d or a during the first round, while *weak* players never punish. Assuming that all players follow this strategy, then all *weak* players will have a *pun-* and all *tough* players a *pun+* reputation. We need to check for the conditions that ensure that punishing is optimal for *tough* players, and that cooperating with *pun+* opponents is optimal for *nice* players, since by assumption *weak* players never punish and *bad* players never cooperate.

By not punishing, a *tough* player can save costs of k in the short run, at the price of gaining a negative reputation *pun-*. If this player meets a *nice* opponent in the next game, then this opponent will play d instead of c, reducing the player's payoff by $R/2$. Additionally, the player will need to punish in order to restore their reputation at a cost of k. If the player instead meets a *bad* opponent in the following round, then no additional costs arise, since this opponent plays d in any case, requiring punishment for sustaining *pun+*. Punishing is thus optimal for a *tough* player as long as

$$\delta(1-p)(R/2+k) \geq k$$

or

$$(1-p) \geq \frac{k}{\delta(R/2+k)} \tag{2}$$

which requires that the share of *nice* players, the discount factor and the return to cooperation are all sufficiently large, and that punishment is not too costly.

For a *nice* player to cooperate with a *pun+* player we need that damage from being punished is larger than the gains of opportunistic deviation, that is

$$d \geq 1 - R/2 \tag{3}$$

If both (2) and (3) are satisfied, then there exists an equilbrium where all *nice* players play c whenever they are paired with a *tough* player. In this equilibrium, the rate of first round contributions will be given by $(1-p)(1-q)$.

Interactions across group boundaries

We now consider what happens when players across two groups start interacting, for example because one group immigrated into the territory of another. We denote the respective shares of members of group 1 and group 2 in the resulting total population with s_1 and $s_2 = 1 - s_1$, respectively. Players tend to mainly interact with members of their own group, but occasional cross-group pairings will occur. In particular, we specify that in each round a player is paired with an opponent of the same group with probability $1-\gamma$. With probability γ, the co-player is randomly drawn from the whole population. Thus, the probability of meeting a member of the opposite society is $r_1 = \gamma s_2$ for members of group 1 and $r_2 = \gamma s_1$ for members of group 2.

We assume that reputations remain group specific, so that the relevant information is only shared among group members, without flow of information across group boundaries. This is justified if we assume that individuals gossip about group members actions only with other members

of the ingroup, but not with outgroup members. We consider first the case where groups cannot identify members of the other group, so that individuals cannot build a personal reputation across group boundaries. In what Fearon and Laitin (1996) refer to as the ingroup policing equilibrium, concerns for reputation among the ingroup can be sufficient to sustain cooperation also with outgroup members.

Consider first two groups that both use coop-rep to sustain cooperation within the group. Under what conditions can an equilibrium be sustained in which *nice* players cooperate in interactions with outgroup members? In such an equilibrium, a *nice* player would need to cooperate in all interactions with the other group, as actions in games with outgroup members cannot be conditioned on the opponent's reputation, since reputational information is only shared within group. As in the within group case, we assume that *bad* players play d always. Defecting against an outgroup member then yields short run gains of $1 - R/2$, but reputation among the ingroup would turn to *coop-*. If in the following round, the player encounters again a member of the other group, which happens with probability r, then the opponent's actions are not affected by the within group reputation, and the player can restore their reputation to *coop+* without bearing any cost. This limited reach of ingroup reputation reduces the range of parameter values for which the cooperative equilibrium is sustainable. The relevant condition is now

$$(1-p) \geq \left(\frac{1-\delta(1-r)}{\delta(1-r)}\right)\left(\frac{1-R/2}{R-1}\right) \tag{4}$$

where the right hand side is increasing in r. This condition is tighter than the single group equivalent (1) as long as $r > 0$. Since players cannot condition on an outgroup member's reputation, we also need to ensure that they prefer to unconditionally play c against members of the outgroup, rather than playing a. This requires that on average, returns to c are higher than zero, or

$$p^o \leq 2\left(\frac{R-1}{R}\right) \tag{5}$$

where p^o denotes the share of bad types in the other group.

Under what conditions can a cooperative equilibrium across groups be sustained if both groups

use pun-rep? In such a separating equilibrium across groups, *tough* players punish opponents who play d or a in intergroup pairs, in order to keep their *pun+* reputation in their own group, providing incentives for opponents to play c. As in the single group case, we assume that *weak* players never punish and thus have a *pun-* reputation in their own group. In the first round of intergroup interactions, *nice* players play c in a cooperative equilibrium to avoid being punished, while *bad* players play d always.

For a *tough* player, deviation from the equilibrium strategy saves the punishment cost k in the short run, at the price of having their reputation within their own group turn negative (*pun-*). The difference from the the single group case is that the player may in the next interaction re-encounter a member of the other group, who has no reputational information about the player. If this opponent is of the *bad* type and thus plays d, then the player will be able to update their reputation back to *pun+* without incurring extra cost. If the opponent is *nice* and plays c, then the player's reputation will not be updated, but no additional losses arise. This possibility of another encounter with a member of the other group limits the strength of the within group reputation, and the condition for punishing to be optimal for *tough* players is now

$$\left(\frac{1-r}{1-r(1-p^o)\delta}\right)\delta(1-p)(R/2+k) \geq k$$

or

$$(1-p) \geq \left(\frac{k}{\delta(R/2+k)}\right)\left(\frac{1-r(1-p^o)\delta}{1-r}\right) \tag{6}$$

This is again tighter than the single group equivalent (2) as long as $r > 0$. As in the case where both groups used cooperation-based reputation, *nice* types will have to play c against any player of the opposite group, due to the lack of reputational information across groups. Thus the cost of being punished has to be sufficiently strong to deter a *nice* player from playing d against all opponents from the other group, even though only *tough* opponents will punish. We thus need

$$d \geq \frac{1}{1-q^o}(1-R/2) \tag{7}$$

where q^o denotes the share of *weak* players in the other group.

In both cases, while conditions for cooperation across group boundaries are more stringent than within group, it is nevertheless possible that full cooperation among *nice* players of both groups can be achieved. If instead one group uses a reputation based on cooperation, and the other based on punishment, then a cooperative equilibrium is not feasible. In across-group interactions, members of the coop-rep group have no incentives to punish, since their own group does not value punishing, and they cannot build a reputation in the other group. Players from the pun-rep group then have no incentives to cooperate in these interactions, since they will not get punished and since first round actions do not affect their reputation among their peers. Instead, all members of the pun-rep group, independently of their type, will necessarily play d whenever they are paired with an opponent from the other group.

A possibility is an equilibrium where *nice* players of the coop-rep group play c in all interactions with members of the other group, while *tough* players of the pun-rep group play d and punish if the opponent plays a or d. This would constitute a semi-cooperative equilibrium, since one group cooperates in intergroup interactions while the other does not. In such an equilibrium, *nice* players of the coop-rep group have two types of incentives to contribute when paired with a member of the other group: their reputation within their own group, and the threat of being punished by their opponent. The resulting condition for *nice* players to cooperate across group boundaries is then

$$(1-p) \geq \left(\frac{1-\delta(1-r)}{\delta(1-r)}\right)\left(\frac{1-R/2}{R-1}\right) - \left(\frac{[1-\delta(1-r)(1-p)]d}{\delta(1-r)(R-1)}\right) \tag{8}$$

where the right hand side is smaller than in the equivalent condition when both groups use cooperation-based reputation (5). The additional incentives from the threat of punishment make this equilibrium more likely to be feasible.

For members of the pun-rep group, incentives to punish in intergroup interactions are identical to the case when both groups are of the same type, and we thus need that condition (6) is satisfied. Under these conditions, there is an equilibrium in which *nice* players of the coop-rep group cooperate in intergroup interactions, while their opponents defect. While this outcome is theoretically possible, it appears unlikely that this can constitute a stable equilbrium in practice,

as players from the coop-rep group receive a negative payoff from all interactions with the other group. If possible, they would therefore strongly prefer to shut down any intergroup interactions and completely segregate the two groups.

Reputation of outgroup members

While it appears reasonable in a short run perspective to assume that players struggle to identify individuals of other groups, we should expect that if groups interacted repeatedly over a longer time, cross-group identification becomes possible. In this section, we allow individuals to build a reputation in other groups, which facilitates intergroup cooperation. We retain the assumption that groups are distinct, that reputational information is not shared across group boundaries, and that groups con observe the actions of outgroup members only during interactions with players from the ingroup.

We first show that if both groups use the same reputational type, and if reputation building across group is possible, then if cooperation can be sustained within groups, it is also feasible to be sustained across group boundaries. This follows from the fact that now in intergroup pairings, a player's reputation in both groups is at stake, making deviations more costly. We first consider the case where both groups use coop-rep. We denote the value of having a good reputation, for a *nice* player, within the own as well as in the other group with V^{++}, where the first symbol denotes the players' standing among their own group. If the player defects during an interaction with a member of the other group, both reputations turn negative and we denote the resulting reputational value with V^{--}. Cooperating in an interaction across groups turns both reputations positive, while cooperating in a within group pairing is not observed by the other group and only the ingroup reputation would improve (V^{+-}). In a cooperative equilibrium, nice players cooperate with opponents that have coop+ in the relevant group. The respective value functions can then

be written as

$$\begin{aligned}
V^{++} =&(1-p)(R-1)+p(1-R/2)+\delta V^{++}\\
V^{+-} =&(1-r)(1-p)(R-1)+(1-r)p(1-R/2)+\delta((1-rp)V^{+-}+rpV^{++})\\
V^{--} =&\delta\left((1-p)V^{--}+(1-r)pV^{+-}+rpV^{++}\right)
\end{aligned}$$

For cooperation with opponents from the same group, we need to have $\delta V^{+-} \geq (1-R/2)+\delta V^{--}$, which translates to the same condition as (5). If this condition is satisfied, it follows directly that cooperation can also be sustained across groups. A nice player cooperates in pairings with a member of the other group if $\delta V^{++} \geq (1-R/2)+\delta V^{--}$, which is guaranteed to hold since $V^{++} \geq V^{+-}$. Incentives to cooperate are higher in intergroup pairings due to reputational concerns in both groups.

We find a similar result in the case where both groups use pun-rep. In a cooperative equilibrium, *tough* players punish in all interactions, and *nice* players cooperate with a player with *pun+* in the relevant group. Incentives for nice players to cooperate against a *pun+* are the same as in the single group case and given by (3). Using W^{++} to denote the value for a *nice*, *tough* player of having pun+ in both groups, we now have

$$\begin{aligned}
W^{++} =&(1-p)(k+R/2)+\delta W^{++}\\
W^{+-} =&(1-r)(1-p)(k+R/2)+\delta((1-r)W^{+-}+rW^{++})\\
W^{--} =&\delta\left((1-r)W^{+-}+rW^{++}\right)
\end{aligned}$$

For sustaining punishing within a group, we need that $\delta W^{+-} \geq k+\delta W^{--}$ or equivalently

$$(1-p) \geq \frac{k}{\delta(1-r)(k+R/2)} \tag{9}$$

which again is tighter than the single group counterpart (2) as long as $r > 0$. As in the case

when both groups use coop-rep, cooperation across group boundaries can always be sustained if the within group condition is satisfied. Players are more likely to punish defectors from other groups, as failing to do so would damage their reputation in both groups. Formally, *tough* players punish in intergroup interactions as long as $\delta W^{++} \geq k + \delta W^{--}$, which is milder than (9) since $W^{++} \geq W^{+-}$

Does cooperation among group members always imply that it can also be sustained across group boundaries? We find that this is not the case if groups use different reputational mechanism. We consider two groups, one group relying on coop-rep and the other on pun-rep, and a cooperative equilibrium where *nice* players of both groups cooperate in interactions across groups. *Tough* players from pun-rep will punish as long as (9) is satisfied, giving players from coop-rep incentives to cooperate on top of reputational concerns within their own group. Thus, as long as (5) is satisfied and cooperation can be sustained within the coop-rep group, its *nice* members will also cooperate with pun-rep members. It is harder to incentivize pun-rep players to cooperate across group boundaries, which can be achieved in two ways: by concerns about their reputation in the coop-rep group, or by punishment by members of the coop-rep group, who in turn are concerned about their reputation within the pun-rep group. The problem is that across-group reputational concerns are weak unless intergroup pairings are frequent. For pun-rep players to cooperate because of concerns about their reputation in the coop-rep group, we need that

$$(1-p) \geq \left(\frac{1-\delta r}{\delta r}\right)\left(\frac{1-R/2}{R-1}\right) \tag{10}$$

which is substantially more demanding than the single group counterpart (1) if r is small. Similarly, for coop-rep group members to have sufficient incentives to punish in across-group interactions we need

$$(1-p) \geq \frac{k}{\delta r(R/2+k)} \tag{11}$$

It follows that unlike in cases where both groups use the same reputational type, sustaining cooperation across groups is more demanding than within groups also if players can build a reputation in the other group, as long as intergroup pairings are less frequent than within group pairings.

Sustained intergroup cooperation is only feasible if r is sufficiently large. Thus, an implication of this extension is that an increased frequency of cross-group interactions can relax the conditions required for successful cooperation and mitigate the negative effects of intergroup incompatibility. Segregation and the maintenance of enclaves instead can contribute to conflict. These conclusions echo those of contact theory (Allport 1954), but highlight a different mechanism behind the positive effect of intergroup contact.

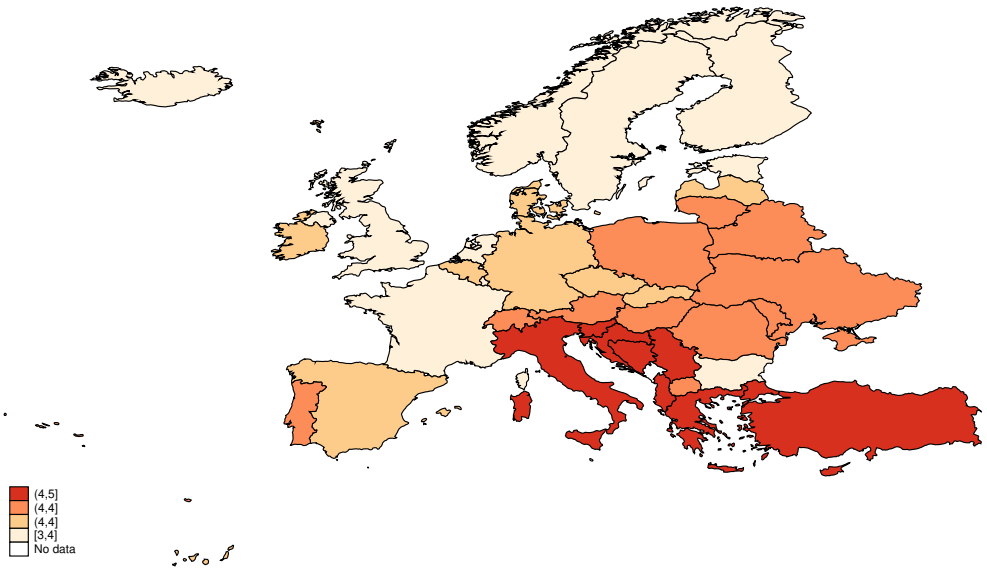
Empirical patterns

In this section we present a number of empirical patterns consistent with our theory. In particular, we provide evidence from observational data consistent with two predictions of the model. First, that among immigrants to cooperation-based societies, those coming from punishment-based societies are more likely to engage in deviant behavior. Second, that cultural incompatibility across societies in terms of reputation systems is correlated with outgroup discrimination. Discrimination is more likely to originate from members of coop-rep societies and be directed against members of pun-rep societies, rather than the other way around.

A major challenge in testing our argument is that we have no good empirical way of classifying societies into punishment-based and cooperation-based. That would require knowledge not only of behavioral patterns, but also of reputational considerations, such as what type of reputation society members try to build, or what reputational aspect interests them the most about parties they transact with (whether they are likely to punish deviations or exhibit cooperative behavior). In the absence of such knowledge, we employ an imperfect proxy of punishment-based cooperation from survey data. The European Social Survey (ESS) asks respondents the following question: "Now I will briefly describe some people. Please listen to each description and tell me how much each person is or is not like you. Use this card for your answer. It is important to her/him to get respect from others." The right to respect is a fundamental component of the concept of honor (Stewart 1994), and lack of respect, in words or in acts, is an insult that often calls for violence in honor cultures. As such, though imperfect, this questions proxies for the importance

of punishment-based reputation.[2] Answers range from 1 (very much like me) to 6 (not at all like me). We reverse-code the variable, so that higher values indicate a higher importance on being respected by others and compute country-level averages across all 8 ESS waves (2002-2016). As can be seen in Figure 1, Balkan and Middle Eastern countries, which represent the traditional societies of honor, are overrepresented among societies that place a high importance on respect.

Figure 1: Importance of respect, by country



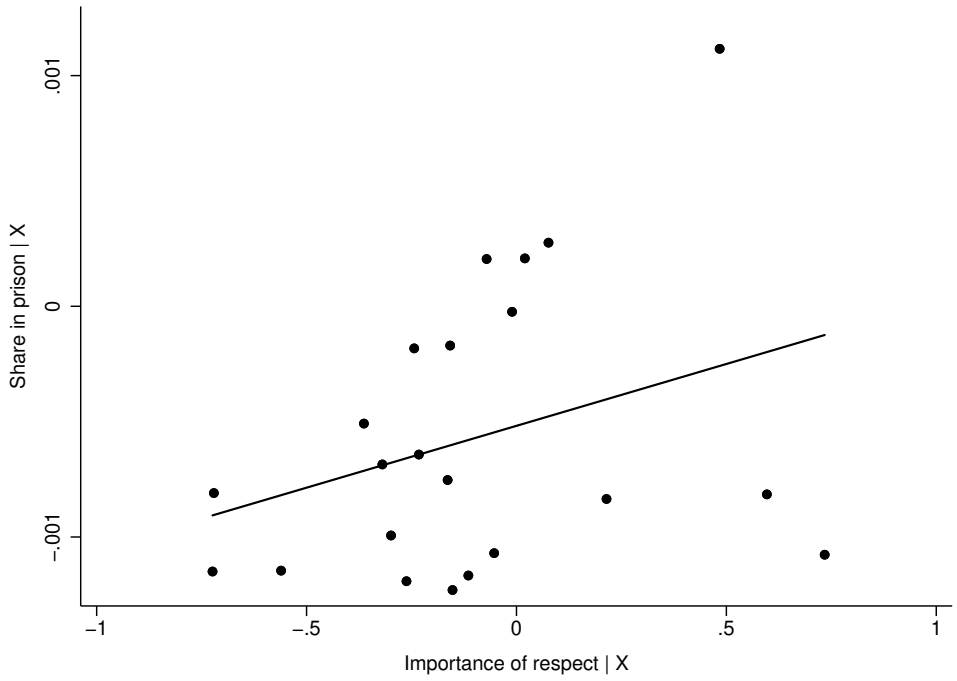
Notes: Average importance of respect by country. Data is from the ESS waves 1-8 (2002-2016).

One of the main predictions of our model is that interaction between coop-rep and pun-rep cultures results in the latter deviating in transactions with the former. We provide evidence consistent with this prediction in the context of crime rates in the US, a canonical example of dignity society (Leung and Cohen 2011), in which the importance of punishment is expected to be low. In the absence of detailed data on crime rates by nationality among immigrant groups we take an alternative approach. The US census provides information on the type of group quarters in which respondents reside, and allows for the identification of institutional quarters such as prisons.

[2]The question is part of the Schwartz value survey, a module of the ESS aimed at measuring basic human values recognized across all cultures according to the theory in Schwartz (1994). The importance of respect taps into the value of Tradition, which also encompasses acceptance and commitment to the values and behaviors dictated by one's culture.

We use data from the default samples of the Integrated Public Use Microdata Series (IPUMS) of the census for the years 1940-1980 (Ruggles et al. 2010) and identify US-born individuals whose father was born outside the US.[3] We assign to each individual the average value of the importance of respect in the father's country of origin. Figure 2 presents the correlation between the share of each ethnicity that is imprisoned at census time and the importance of respect in the origin country, accounting for origin country log GDP per capita (from the World Bank) and crime rates (from the UN Survey on Crime Trends). Consistent with our theory, the share of second generation immigrants imprisoned is higher for pun-rep societies (as proxied by the importance of respect).

Figure 2: Share of nationality in prison and importance of respect in country of origin



Notes: The figure plots residuals of all variables from an individual-level regression on log GDP per capita in 2000 and the crime rate in the country of origin averaged at the country of origin level. Data is from the 1940-1980 1% and 5% IPUMS samples and consists of US-born individuals with a foreign-born father.

Table 2 presents a more systematic analysis of this pattern. We run regressions of the form:

[3]We use 5% samples for 1960 and 1980 and 1% samples for the remaining decades. We focus on the second generation following the empirical literature in cultural economics (Giuliano 2007; Fernández 2011). This strategy attenuates the potential problem of differential selection of migrants across cultures. Results are similar when including the first generation in the sample.

$$P_{ic} = \alpha + \beta R_c + \mathbf{X}_{ic}\gamma_1 + \mathbf{Z}_c\gamma_2 + \epsilon_{ic}$$

where subscripts i and c denote individuals and countries, respectively. P_{ic} is an indicator for individuals in prison at census time and R_c is the average importance of respect in the individual's country of origin. $\mathbf{X}_{ic}$ and $\mathbf{Z}_c$ are individual and origin country-level controls, respectively. We control for sex, age and age squared, as well as race indicators. To ensure that the importance of respect does not capture differences in levels of development across countries, which could directly affect imprisonment rates, we control for origin country log GDP per capita in 2000 and crime rate. We also control for the share of each major religion in the origin country from the Pew Center, to account for other cultural differences correlated with both human values and propensity to engage in crime.

As it appears, there is a positive correlation between the importance of respect in the the country of origin and the likelihood of being in prison. Magnitudes are relatively stable, and increase with the inclusion of origin country controls. In the most parsimonious specification reported in column 5 that includes indicators for state of residence and census decade, a one standard deviation increase in the importance of respect is associated with a 0.05 percentage point increase in the mean of this variable ($\mu = 0.1\%$), a large effect.

Table 2: Imprisonment and importance of respect in country of origin

Dep. variable	In prison				
	(1)	(2)	(3)	(4)	(5)
Respect	0.00116*	0.000835	0.00135**	0.00169***	0.00168**
	(0.000616)	(0.000552)	(0.000502)	(0.000553)	(0.000745)
Observations	71130	71130	64444	64444	64444
R-squared	0.000155	0.00134	0.00153	0.00156	0.00477
Individual controls		✓	✓	✓	✓
Origin country controls			✓	✓	✓
Religious controls				✓	✓
State FE					✓
Census year FE					✓

Notes: Data is from the 1940-1980 1% and 5% IPUMS samples and consists of US-born individuals with a foreign-born father. Individual controls are sex, age, age squared and seven indicators for race (white, black, American Indian or Alaska native, Chinese, Japanese, other Asian or Pacific islander, other). Origin country controls are log GDP p.c. in 2000 and the crime rate in 1980. Religious controls are the share of the origin country that belongs to each of five major religious groups (Christians, Buddhists, Muslims, Hindus, Jews). Standard errors are clustered at the country of origin level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

We next turn to our model's major prediction, incompatibility between reputation systems that leads to the breakdown of cooperation. We proxy for this failure of cooperation using discrimination of members of one society towards those of another. The ESS asks respondents the following question: "Would you describe yourself as being a member of a group that is discriminated against in this country?" Responses are binary. We relate responses of second generation immigrants to cultural differences between their country of residence and the country of origin of their parents.

Column 1 of Table 3 shows that the (absolute) distance between countries in the average importance of respect is positively correlated with perceptions of discrimination. This pattern is consistent with our theoretical predictions. The higher the difference in terms of reputation systems between countries the is higher the likelihood of cooperation breaking down and intergroup avoidance ensuing. Columns 2 and 3 successively add more bilateral controls that could affect patterns of bilateral discrimination. Following Guiso, Sapienza, and Zingales (2009), in column 2, we control for countries' geographic distance and whether they share a common border, official language, and colonial past. We also control for the absolute difference between two countries in log GDP per capita in 2000. In column 3, we add a control for religious similarity, measured as the empirical probability that two randomly chosen individuals in two countries will share the same religion.[4] The remaining columns of Table 3 replicate the specifications of columns 1–3, but control for origin country (columns 4–6) and host country fixed effects (7–9). The coefficient is strongly significant and remarkably stable across specifications.

[4]This measure is from Guiso, Sapienza, and Zingales (2009) and it is computed as the product of the fraction of individuals in country j and in country i who have religion k, summed across all religions. Data on religious adherence is from the World Values Survey.

Table 3: Cultural distance and perceptions of discrimination

Dep. variable	Feel discriminated								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance respect	0.0529***	0.0573***	0.0434***	0.0576***	0.0541***	0.0266***	0.0436	0.0466*	0.0382**
	(0.00884)	(0.0108)	(0.00912)	(0.00899)	(0.00859)	(0.00918)	(0.0269)	(0.0235)	(0.0148)
Observations	24132	23895	23895	24106	23872	23872	24132	23895	23895
R-squared	0.0179	0.0308	0.0410	0.0525	0.0537	0.0606	0.0330	0.0444	0.0520
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geo-cultural controls		✓	✓	✓	✓	✓	✓	✓	✓
Economic controls		✓	✓		✓	✓		✓	✓
Religious controls			✓			✓			✓
Origin country FE				✓	✓	✓			
Host country FE							✓	✓	✓

Notes: Data from ESS waves 1-8. The sample contains all second generation immigrants. *Feel discriminated* is the respondent's agreement to being part of a discriminated group. Individual controls include sex, age and age squared. Geo-cultural controls include geographic distance between the capitals of the host country and the country of origin, a common border indicator, an indicator for sharing a common official language and an indicator for a common colonial past. Economic controls are absolute differences in log per capita GDP in 2000. Religious controls are absolute differences in shares of major religions. *Distance respect* measures the absolute difference in the share of a country's population that states it is important to be respected by others. Two-way clustering of all standard errors at the host country and the country of origin level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Our theory has an additional testable implication. To the extent that discrimination proxies for the avoidance of interactions, the model predicts that discrimination should originate from members of coop-rep societies and be directed towards members of pun-rep societies, rather than vice versa. We explicitly test this prediction in Table 4 by decomposing the distance in the importance of respect in two parts. We denote by *Respect (+)* the positive difference between a country's average importance assigned to respect and that of the corresponding partner and with *Respect (−)* the (absolute value of) the respective negative difference. Higher values of *Respect (+)* imply that the importance of respect is higher in the country of origin of the respondents than in the residence country. Higher values of *Respect (−)* imply instead the opposite pattern.

Table 4: Cultural distance and perceptions of discrimination

Dep. variable	Feel discriminated								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Respect (+)	0.0603***	0.0597***	0.0458***	0.0646***	0.0606***	0.0304**	0.0809**	0.0709**	0.0561***
	(0.0111)	(0.00911)	(0.00784)	(0.00937)	(0.00778)	(0.0121)	(0.0305)	(0.0279)	(0.0178)
Respect (−)	-0.0171	-0.0119	-0.00297	-0.0336	-0.0394*	-0.0152	-0.00890	0.00477	0.0121
	(0.0211)	(0.0186)	(0.0110)	(0.0211)	(0.0205)	(0.0146)	(0.0174)	(0.0116)	(0.00997)
Observations	24194	23895	23895	24170	23872	23872	24194	23895	23895
R-squared	0.0254	0.0354	0.0456	0.0535	0.0566	0.0641	0.0362	0.0463	0.0532
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geo-cultural controls		✓	✓	✓	✓	✓	✓	✓	✓
Economic controls		✓	✓		✓	✓		✓	✓
Religious controls			✓			✓			✓
Origin country FE				✓	✓	✓			
Host country FE							✓	✓	✓

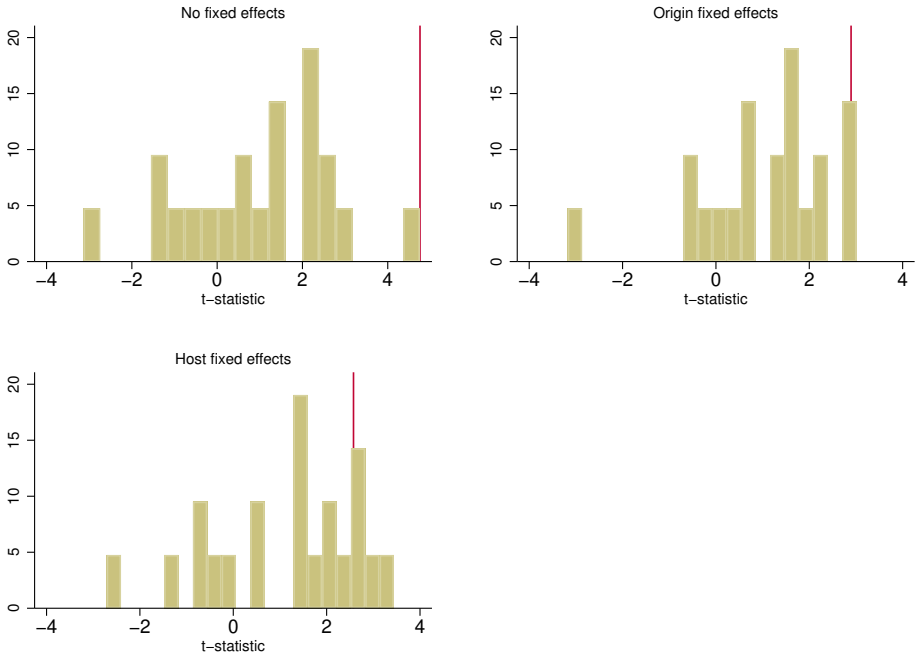
Notes: Data from ESS waves 1-8. The sample contains all second generation immigrants. *Feel discriminated* is the respondent's agreement to being part of a discriminated group. Individual controls include sex, age and age squared. Geo-cultural controls include geographic distance between the capitals of the host country and the country of origin, a common border indicator, an indicator for sharing a common official language and an indicator for a common colonial past. Economic controls are differences in log per capita GDP in 2000. Religious controls are differences in shares of major religions. *Distance respect* measures the difference in the share of a country's population that states it is important to be respected by others. A positive distance implies that respect is more important in the country of origin. All economic and religious controls are allowed to have different effects for positive and negative differences. Two-way clustering of all standard errors at the host country and the country of origin level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

The pattern revealed by Table 4 is consistent with the predictions of the theory. Perceptions of discrimination are primarily expressed by immigrants from countries that value respect more than their dyadic partner implying that members of coop-rep societies are more likely to discriminate against members of pun-rep societies. The estimated effect on *Respect*(-) instead, is never significant and is inconsistent in sign and magnitude across specifications. Discrimination and avoidance appears to be the response of coop-rep societies towards pun-rep societies, rather than vice versa.

Bilateral distance in terms of the importance placed on respect could proxy for cultural distance more broadly, rather than differences specific to the role of reputation. To assess the relevance of this concern we replicate the specifications reported in columns 3, 6 and 9 of Table 3 by replacing the difference in the importance of respect with differences in the importance of each of twenty other basic human values. Together, these twenty-one values constitute the Schwartz module of the ESS. Figure 3 plots histograms of t-statistics for coefficients estimated as part of this exercise. Few other values significantly predict perceptions of discrimination, and none of them does so consistently across specifications. Given the high correlation among human values, these results are encouraging for our analysis as they suggest that avoidance and discrimination are not the

general product of cultural differences, but rather the specific outcome of cross-cultural differences in the role of punishment and honor.

Figure 3: Other measures of cultural distance



Notes: Data from ESS waves 1-8. Panels display t-statistics for coefficient estimates of the effect of various measures of cultural distance on perceptions of discrimination, from specifications in columns 3, 6 and 9 of Table 3. The red line indicates t-statistics when the right hand side variable is the distance in the importance of respect.

Conclusion

In this paper, we study conditions for cooperation and conflict when societies with different reputational considerations (punishment-based vs cooperation-based) interact. Using a game theoretic model, we show that members of a punishment-based society have little incentive to cooperate in such intergroup interactions, while members of a society with cooperation-based reputation might in consequence avoid such interactions alltogether. Empirical patterns on immigrants' imprisonment rates in the US and perceptions of discrimination of immigrants in European countries are consistent with the model's predictions.

Our paper is most related to Fearon and Laitin (1996), who explain interethnic cooperation by

showing theoretically the existence of equilibria in which groups avoid escalation of conflict and equilibria in which ingroups "police their own". By adding the element of cross-cultural difference we show that the existence of such equilibria is more likely when the groups that interact share the same reputational considerations. While their theory explains why intergroup cooperation is common, ours explains when such cooperation is more likely to break down.

By highlighting the existence of cross-cultural variation in mechanisms that support observationally equivalent group-level behaviors and outcomes, our study is very close in spirit to Greif (1994). Using the comparison between Genoese and Maghribi traders, Greif (1994) shows that the same fundamental economic problem can be solved equally effectively in different ways, each supported by different cultural beliefs. To solve the problem of agency, the Genoese relied on contracts and impersonal relations, while the Maghribis relied on a collectivist system of information-sharing. Yet intergroup contact in the context of international trade favored one group (the Genoese) and led the other one (the Maghribis) to extinction. Our framework similarly considers isolated cultures that effectively manage to solve a societal problem (cooperation) and their interaction, but focuses on reputation as the crucial aspect of cultural differences.

The theoretical framework presented here can be extended in various ways. One such extension concerns the role of institutions and the rule of law. So far we have abstracted from any institutional mechanisms that enforce cooperation and punish deviant behavior. It can be shown that allowing for a third party enforcer has different effects in punishment-based vs cooperation-based societies. External punishment is compatible with cooperation-based reputation and can enhance cooperation by further increasing the cost of deviant behavior. It is instead less compatible with punishment-based reputation. Since typically not only defection, but also punishment would be sanctioned by law, external enforcement reduces the incentives to punish and can theoretically even increase defection rates in the short run in punishment-based societies. This insight is consistent with the empirical observation that the rule of law is stronger in cooperation-based societies rather than in societies of honor. It also echoes, albeit through a different mechanism, the findings of Tabellini (2008) that local enforcement has potentially pernicious effects for cooperation. We are exploring this interaction between culture and institutions in parallel research.

References

Allport, Gordon. 1954. "The Nature of Prejudice". (Reading, MA).

Aronowitz, A. 2002. "Assimilation, Acculturation, and Juvenile Delinquency among Second Generation Turkish Youths in Berlin". In *Migration, Culture Conflict and Crime*, ed. by Joshua D. Freilich and Moshe Addad, 233–260. Dartmouth: Dartmouth Publishing Company.

Bansak, Kirk, Jens Hainmueller, and Dominik Hangartner. 2016. "How Economic, Humanitarian, and Religious Concerns shape European Attitudes toward Asylum Seekers". *Science* 354 (6309): 217–222.

Bates, Robert H. 1983. "Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa". In *State versus Ethnic Claims: African Policy Dilemmas*, ed. by Donald Rothchild and Victor A. Olorunsola, 152:171. Westview Press Boulder, CO.

Bednar, Jenna, and Scott Page. 2007. "Can Game(s) Theory Explain Culture? The Emergence of Cultural Behavior within Multiple Games". *Rationality and Society* 19 (1): 65–97.

Bednar, Jenna, and Scott E Page. 2018. "When Order Affects Performance: Culture, Behavioral Spillovers, and Institutional Path Dependence". *American Political Science Review* 112 (1): 82–98.

Benjaminsen, Tor A, Faustin P Maganga, and Jumanne Moshi Abdallah. 2009. "The Kilosa Killings: Political Ecology of a Farmer–Herder Conflict in Tanzania". *Development and Change* 40 (3): 423–445.

Boehm, Christopher, et al. 1984. *Blood Revenge: The Anthropology of Feuding in Montenegro and other Tribal Societies*. Lawrence: University Press of Kansas Lawrence, KS.

Bovenkerk, Frank, and Tineke Fokkema. 2016. "Crime among Young Moroccan Men in the Netherlands: Does their Regional Origin Matter?" *European Journal of Criminology* 13 (3): 352–371.

Brandt, Hannelore, Christoph Hauert, and Karl Sigmund. 2003. "Punishment and Reputation in Spatial Public Goods Games". *Proceedings of the Royal Society of London B: Biological Sciences* 270 (1519): 1099–1104.

Brooks, Benjamin A, Karla Hoff, and Priyanka Pandey. 2018. "Cultural Impediments to Learning to Cooperate: An Experimental Study of High- and Low-Caste Men in rural India". *Proceedings of the National Academy of Sciences* 115 (45): 11385–11392.

Campbell, Bradley, and Jason Manning. 2014. "Microaggression and Moral Cultures". *Comparative Sociology* 13 (6): 692–726.

— . 2018. *The Rise of Victimhood Culture: Microaggressions, Safe Spaces, and the New Culture Wars*. Springer.

Campbell, J. K. 1964. *Honour, Family, and Patronage: A Study of Institutions and Moral Values in a Greek Mountain Community*. Oxford: Clarendon.

Dal Bó, Pedro. 2005. "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games". *American Economic Review* 95 (5): 1591–1604.

dos Santos, Miguel, Daniel J Rankin, and Claus Wedekind. 2011. "The Evolution of Punishment through Reputation". *Proceedings of the Royal Society of London B: Biological Sciences* 278 (1704): 371–377.

Ellickson, Robert C. 1991. *Order without Law*. Harvard University Press.

Enke, Benjamin. 2019. "Kinship, Cooperation, and the Evolution of Moral Systems". *The Quarterly Journal of Economics* 134 (2): 953–1019.

Fearon, James D, and David D Laitin. 1996. "Explaining Interethnic Cooperation". *American Political Science Review* 90 (4): 715–735.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism". *Nature* 425 (6960): 785.

Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments". *American Economic Review* 90 (4): 980–994.

Fernández, Raquel. 2011. "Does Culture Matter?" In *Handbook of Social Economics*, ed. by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 1:481–510. Elsevier.

Figueredo, Aurelio José, et al. 2004. "Farmers, Herders, and Fishers: The Ecology of Revenge". *Evolution and Human Behavior* 25 (5): 336–353.

Gagnon Jr, Valere P. 1994. "Ethnic Nationalism and International Conflict: The Case of Serbia". *International Security* 19 (3): 130–166.

Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca, NY: Cornell University Press.

Giuliano, Paola. 2007. "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association* 5 (5): 927–952.

Gould, Roger V. 2000. "Revenge as Sanction and Solidarity Display: An Analysis of Vendettas in Nineteenth-Century Corsica". *American Sociological Review* 65 (5): 682–704.

Greif, Avner. 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies". *Journal of Political Economy* 102 (5): 912–950.

Grosjean, Pauline. 2014. "A History of Violence: The Culture of Honor and Homicide in the US South". *Journal of the European Economic Association* 12 (5): 1285–1316.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *The Quarterly Journal of Economics* 124 (3): 1095–1131.

Hainmueller, Jens, and Dominik Hangartner. 2013. "Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination". *American Political Science Review* 107 (1): 159–187.

Hamamura, Takeshi, et al. 2009. "Approach–Avoidance Motivation and Information Processing: A Cross-Cultural Analysis". *Personality and Social Psychology Bulletin* 35 (4): 454–462.

Henrich, Joseph, et al. 2006. "Costly Punishment across Human Societies". *Science* 312 (5781): 1767–1770.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment across Societies". *Science* 319 (5868): 1362–1367.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkeley: University of California Press.

Huntington, Samuel P. 1996. *The Clash of Civilizations and the Remaking of World Order*. London: Simon & Schuster.

Junger-Tas, Josine. 2001. "Ethnic Minorities, Social Integration and Crime". *European Journal on Criminal Policy and Research* 9 (1): 5–29.

Keiser, R. Lincoln. 1991. *Friend by Day, Enemy by Night: Organized Vengeance in a Kohistani Community*. Fort Worth, TX: Holt, Rinehart / Winston.

Kim, Young-Hoon, and Dov Cohen. 2010. "Information, Perspective, and Judgments about the Self in Face and Dignity Cultures". *Personality and Social Psychology Bulletin* 36 (4): 537–550.

Laitin, David D. 1995. "National Revivals and Violence". *European Journal of Sociology/Archives Européennes de Sociologie* 36 (1): 3–43.

Leung, Angela K-Y, and Dov Cohen. 2011. "Within-and between-Culture Variation: Individual Differences and the Cultural Logics of Honor, Face, and Dignity Cultures." *Journal of Personality and Social Psychology* 100 (3): 507.

Little, Peter D. 1992. *The Elusive Granary: Herder, Farmer, and State in Northern Kenya*. Cambridge University Press.

Milinski, Manfred, Dirk Semmann, and Hans-Jürgen Krambeck. 2002. "Reputation Helps Solve the Tragedy of the Commons". *Nature* 415 (6870): 424–426.

Moscona, Jacob, Nathan Nunn, and James A Robinson. 2017. "Keeping it in the Family: Lineage Organization and the Scope of Trust in Sub-Saharan Africa". *American Economic Review, Papers and Proceedings* 107 (5): 565–71.

Moscona, Jacob, Nathan Nunn, James A Robinson, et al. 2019. *Kinship and Conflict: Evidence from Segmentary Lineage Societies in Sub-Saharan Africa*. Unpublished Manuscript), number=. Harvard University.

Nisbett, Richard E., and Dov Cohen. 1996. *Culture of Honor: The Psychology of Violence in the South*. Boulder, CO: Westview Press.

Nowak, Martin A., and Karl Sigmund. 1998a. "Evolution of Indirect Reciprocity by Image Scoring". *Nature* 393 (6685): 573–577.

Nowak, Martin A, and Karl Sigmund. 1998b. "The Dynamics of Indirect Reciprocity". *Journal of Theoretical Biology* 194 (4): 561–574.

Ofuoku, Albert U, and Benjamin I Isife. 2009. "Causes, Effects and Resolution of Farmers-Nomadic Cattle Herders Conflict in Delta State, Nigeria". *Journal of Agricultural Practice and Science* 69 (1-2).

Ruggles, Steven, et al. 2010. *Integrated Public Use Microdata Series: Version 5.0 [Machine-Readable Database]*. Minneapolis: University of Minnesota.

Sambanis, Nicholas, and Moses Shayo. 2013. "Social Identification and Ethnic Conflict". *American Political Science Review* 107 (2): 294–325.

Schläpfer, Alain. 2018. "The Emergence and Selection of Reputation Systems that Drive Cooperative Behaviour". *Proceedings of the Royal Society B: Biological Sciences* 285 (1886): 20181508.

Schwartz, Shalom H. 1994. "Are There Universal Aspects in the Structure and Contents of Human Values?" *Journal of Social Issues* 50 (4): 19–45.

Sellin, Thorsten. 1938. "Culture Conflict and Crime". *American Journal of Sociology* 44 (1): 97–103.

Shoham, Shlomo. 1962. "The Application of the "Culture-Conflict" Hypothesis to the Criminality of Immigrants in Israel". *The Journal of Criminal Law, Criminology, and Police Science* 53 (2): 207–214.

Sigmund, Karl, Christoph Hauert, and Martin A Nowak. 2001. "Reward and Punishment". *Proceedings of the National Academy of Sciences* 98 (19): 10757–10762.

Stewart, Frank Henderson. 1994. *Honor*. Chicago: University of Chicago Press.

Sylwester, Karolina, and Gilbert Roberts. 2013. "Reputation-Based Partner Choice is an Effective Alternative to Indirect Reciprocity in Solving Social Dilemmas". *Evolution and Human Behavior* 34 (3): 201–206.

Tabellini, Guido. 2008. "The Scope of Cooperation: Values and Incentives". *The Quarterly Journal of Economics* 123 (3): 905–950.

Thébaud, Brigitte, and Simon Batterbury. 2001. "Sahel Pastoralists: Opportunism, Struggle, Conflict and Negotiation. A Case Study from Eastern Niger". *Global Environmental Change* 11 (1): 69–78.

Tonah, Steve. 2006. "Migration and Farmer-Herder Conflicts in Ghana's Volta Basin". *Canadian Journal of African Studies/La Revue canadienne des études africaines* 40 (1): 152–178.

Turchin, Peter. 2005. *War and Peace and War: The Life Cycles of Imperial Nations*. Pi.

Turner, Matthew D. 2004. "Political Ecology and the Moral Dimensions of ?Resource Conflicts?: The Case of Farmer–Herder Conflicts in the Sahel". *Political Geography* 23 (7): 863–889.

Turner, Matthew D, et al. 2006. "Farmer-Herder Relations and Conflict Management in Agro-Pastoral Zone of Niger". In *The Future of Transhumance Pastoralism in West and Central Africa: Strategies, Dynamic, Conflicts and Interventions: Proceeding of the International Conference, Held in Abuja, Nigeria, 20-24 November 2006.* Ed. by J.O. Gefu, C.B.I Alawa, and B. Maisamari.

Vega, William A, et al. 1993. "Acculturation and Delinquent Behavior among Cuban American Adolescents: Toward an Empirical Model". *American Journal of Community Psychology* 21 (1): 113–125.

Whyte, William Foote. 2012. *Street Corner Society: The Social Structure of an Italian Slum*. Chicago: University of Chicago Press.