DISCUSSION PAPER SERIES

DP17534

CULTURE CLASH: INCOMPATIBLE REPUTATION MECHANISMS AND INTERGROUP CONFLICT

Vasiliki Fouka and Alain Schläpfer

POLITICAL ECONOMY

CEPR

ISSN 0265-8003

CULTURE CLASH: INCOMPATIBLE REPUTATION MECHANISMS AND INTERGROUP CONFLICT

Vasiliki Fouka and Alain Schläpfer

Discussion Paper DP17534
Published 23 September 2022
Submitted 22 September 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

CULTURE CLASH: INCOMPATIBLE REPUTATION MECHANISMS AND INTERGROUP CONFLICT

Abstract

Under what 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. Reputational concerns can deter defection in one-time interactions within a group, but the informational content of reputation can differ across groups. We consider two types of information. Punishment-based reputation (a "culture of honor") represents past sanctioning behavior of individuals, while a reputation based on image scoring captures past cooperative and uncooperative acts. While either type can successfully sustain cooperation within a group, we show theoretically that interactions of individuals from a punishment-based culture with those from a culture of image scoring can lead to widespread inter-group tensions. Mutual cooperation is a more likely outcome if both cultures use a similar reputation mechanism. We find empirical support for the model's predictions across phenomena related to the emergence of social tensions. Cross-cultural differences in the importance of retaliation predict patterns of host population discrimination against immigrants and variation in bilateral conflict across ethnic groups.

JEL Classification: P0, Z1

Keywords: N/A

Vasiliki Fouka - vfouka@stanford.edu

Alain Schläpfer - alainsch@stanford.edu

Acknowledgements

For helpful comments and discussions we thank Elias Dinas, James Fearon, Karla Hoff, David Laitin, Marco Tabellini and conference participants at Bocconi, the 2019 APSA Annual Meeting, the 2020 SIOE Annual Meeting and the 2020 AEA Annual Meeting. Ethan Tiao and Hailey Weisel provided excellent research assistance.

Culture Clash: Incompatible Reputation Systems and Intergroup Conflict*

Vasiliki Fouka† Alain Schläpfer‡

September 2022

Abstract

Under what 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. Reputational concerns can deter defection in one-time interactions within a group, but the informational content of reputation can differ across groups. We consider two types of information. Punishment-based reputation (a "culture of honor") represents past sanctioning behavior of individuals, while a reputation based on image scoring captures past cooperative and uncooperative acts. While either type can successfully sustain cooperation within a group, we show theoretically that interactions of individuals from a punishment-based culture with those from a culture of image scoring can lead to widespread inter-group tensions. Mutual cooperation is a more likely outcome if both cultures use a similar reputation mechanism. We find empirical support for the model's predictions across phenomena related to the emergence of social tensions. Cross-cultural differences in the importance of retaliation predict patterns of host population discrimination against immigrants and variation in bilateral conflict across ethnic groups.

Keywords: cooperation, reputation, indirect reciprocity, discrimination, conflict

---

*For helpful comments and discussions we thank Elias Dinas, James Fearon, Karla Hoff, David Laitin, Marco Tabellini and conference participants at Bocconi, the 2019 APSA Annual Meeting, the 2020 SIOE Annual Meeting and the 2020 AEA Annual Meeting. Ethan Tiao and Hailey Weisel provided excellent research assistance.

†Stanford University, Department of Political Science. Email: vfouka@stanford.edu.

‡Stanford University, Department of Political Science. Email: alainsch@stanford.edu.

1 Introduction

The puzzle of human cooperation is central in research across the social sciences. How do genetically unrelated individuals manage to cooperate even when there are strong individual incentives to defect? And if cooperation is feasible – as the overcoming of free-rider problems in human societies time and again demonstrates – why does it sometimes break down, leading to social tensions and even conflict? In this paper, we attempt to provide an answer to both questions based on the role of reputation. 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 cooperation failure, tensions and conflict.

That reputation can facilitate large-scale cooperation in one-time interactions is a central finding in a rich literature in economics and evolutionary biology (Nowak and Sigmund, 1998a; Fehr and Gächter, 2000; Fehr and Fischbacher, 2003; Dal Bó, 2005; Henrich et al., 2006). But the informational content of reputation 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 et al., 2001; Brandt et al., 2003; dos Santos et al., 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, 1998b; Milinski et al., 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 deviant 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. Agents are heterogeneous in their cost of cooperating during the prisoner's dilemma and of sanctioning during the punishment round. We consider two types of reputational mechanisms. With punishment-based reputation, or "pun-rep", players can learn the punitive record of their opponent. Individuals then have incentives to build a reputation for punishing defection, inducing cooperative behavior by their opponent, but some players may find it too costly to do so. Reputational information can instead also represent past instances of cooperation and defection. If opponents avoid interactions with known defectors, players have incentives to cooperate and build a good image score. We refer to this as a cooperation-based reputation, or "coop-rep".

We first derive conditions under which either type of reputation system can support a cooperative equilibrium within a group, where individuals with sufficiently low cost cooperate. We then consider interactions across group boundaries, with each group relying on their distinct reputational mechanism. Reputational concerns carry over to intergroup interactions, allowing for a cooperative equilbrium across groups under somewhat tighter conditions than in the within group case, but only if both groups use the same type of reputation. When both groups rely on coop-rep, players cooperate in intergroup interactions to maintain their positive image score 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 an equilibrium does not exist if groups differ in terms of their reputation systems. Members of a coop-rep group don't punish, which implies that pun-rep group members have no incentives to cooperate in intergroup interactions, since they face no threat of punishment by their opponent, and cooperation is not rewarded by their group's reputation type.

We explore two model extensions with similar results. First, we allow for the flow of reputational information across group boundaries, which strengthens incentives to cooperate as long as both groups are of the same reputational type. In the case of

one pun-rep and one coop-rep group, however, reputation shared across groups fails to induce cooperation, as it does not resolve the underlying incompatibility of actions prescribed by the respective reputation mechanisms. Second, we explore outcomes when players can endogenously alter the probability of interacting with outgroup members. A cooperative equilibrium remains feasible with two similar groups. In a situation with group mixing, we observe exclusionary behavior, with all members of the coop-rep group avoiding interactions with members of the pun-rep group.

We demonstrate the broad applicability of the model in explaining patterns of intergroup cooperation and conflict by testing its predictions in two different contexts: attitudes of host populations towards immigrants and bilateral conflict across ethnic groups. A central challenge for the empirical analysis is the absence of systematic measures of reputation across societies. We overcome this problem by relying on a recent ethnographic database of folklore made available to economists by Michalopoulos and Xue (2021). We proxy for punishment-based reputation by the prevalence of motifs related to revenge and retaliation in a society's oral traditions. Wherever possible, we complement this measure with two additional proxies. The first one relies on collections of ethnographic material for 304 societies catalogued online by the *Human Relations Area Files* project. We code as punishment-based those societies where revenge and retaliation are mentioned in the sections of their ethnographic records discussing conflict and social control. The second proxy is a measure of negative reciprocity, constructed as an aggregate of self-reported willingness to punish and take revenge elicited in a global representative survey of economic preferences, the Global Preference Survey (Falk et al., 2018).

We conduct two separate empirical analyses. For the first one, we examine whether differences in the prevalence of revenge-related motifs between host and origin country predict perceptions of discrimination among second generation immigrants in Europe. We rely on the European Social Survey (ESS) and exploit variation in folklore across pairs of host and origin countries. This allows us to control for host and origin country fixed effects, netting out factors – such as political climate or history – that may lead to more anti-immigrant attitudes in certain European countries, as well as factors – such as religion or skin color – that may make certain groups of immigrants more likely to be discriminated.

We find that immigrants from origins that differ from the host country in the importance of revenge are more likely to report being members of a discriminated

group. This result is conditional on other measures of bilateral distance in economic development, geographic proximity, common history and culture. Consistent with the model, discrimination originates from members of relatively more cooperation-based societies and is directed towards members of relatively more punishment-based societies, rather than vice versa. The type of discrimination predicted by bilateral differences in revenge is based on cultural group affiliation, defined by characteristics such as ethnic origin, nationality or religion, and not on individual characteristics such as gender or sexual orientation.

An important concern for this analysis is that the bilateral distance in revenge-related motifs captures incompatibility in dimensions of culture other than reputation structures. We rely on the database of folklore to rule out a role for two important competing theoretical mechanisms behind intergroup tensions: overall propensity to violence, a documented feature of cultures of honor (Nisbett and Cohen, 1996; Grosjean, 2014; Cao et al., 2021), and differences in cooperative preferences (Tabellini, 2008). We find no evidence that the prevalence of either motifs related to conflict and violence or motifs related to cheating and defection predict perceptions of discrimination. The predictive power of differences in revenge motifs remains high even after controlling for differences in economic preferences from GPS and differences in twenty-one human values commonly used by psychologists to conduct cross-cultural comparisons (Schwartz, 1994). Of all economic preferences measured in the GPS, only differences in patience have an independent effect on perceptions of discrimination, but accounting for them does not diminish the role of differences in revenge. Differences in negative reciprocity strongly predict discrimination patterns only when differences in revenge motifs are not controlled for, confirming the fact that these two measures capture a similar underlying construct.

For our second empirical test, we turn to patterns of conflict between ethnic groups. We build on the observation that conflict is often the result of lower intensity disputes between individuals that escalate into larger scale tensions (Boehm et al., 1984; Horowitz, 1985; Rohner, 2011). To conduct an analysis at the ethnic group pair level, mirroring the one for discrimination, we use the Ethnic Power Relations Dataset Family (EPR) (Vogt et al., 2015). The EPR tracks at an annual frequency all politically relevant groups in the world and their status as members of the government or excluded from power. It also tracks conflict between the government and non-government groups by linking to data on conflict incidence from the Uppsala Conflict Data Program

(UCDP). We link groups in the EPR to their oral traditions and other ethnographic characteristics through the Ethnographic Atlas (Murdoch, 1967). We then examine whether the distance in the prevalence of revenge-related motifs predicts the occurrence of conflict between groups, conditional on government and rebel group unobservable fixed characteristics and country by year fixed effects.

We find that a one standard deviation increase in the absolute distance in revenge motifs increases the probability of conflict between the government and a rebel group by 0.25 standard deviations. This estimate increases when accounting for other bilateral differences shown to affect conflict incidence, such as linguistic and religious distance (Spolaore and Wacziarg, 2016; Guarnieri, 2022), differences in political sophistication and societal complexity or inequality in natural endowments.

As in the case of discrimination, the effect of differences in revenge motifs is not driven by distance in other cultural factors affecting the propensity to cooperate. Differences in the prevalence of motifs related to violent conflict and motifs related to cheating are more predictive of conflict than of discrimination against outgroups, but none of the two drives the effect of revenge-related traditions. We are also able to rule out that our measure captures differences in subsistence patterns, specifically reliance on pastoralism, which is strongly associated with honor cultures (Cao et al., 2021). Similarly, our results are not driven by differences in forms of organizational structure shown to affect conflict intensity, such as segmentary lineages (Moscona et al., 2019). The predictive power of differences in revenge-related motifs remains high in alternative empirical specifications and using different methods of clustering standard errors for inference.

The remainder of the paper proceeds as follows. Section 2 surveys literature on the role of reputation and cultural incompatibility in driving conflict and identifies our contribution to this line of research. In Section 3 we present 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 tensions. Section 4 presents the theoretical model. Section 5 reports empirical evidence consistent with the model's predictions. We present our proxies of reliance on punishment-based reputation and show that cross-cultural incompatibility in these measures predicts discrimination of host populations against immigrants in Europe and bilateral conflict across ethnic groups in Africa and Asia. Section 6 concludes and

discusses paths to future research.

2 Related literature

Our paper contributes to literatures in multiple disciplines studying cooperation, reputation, culture and conflict. Research in economics and evolutionary biology shows that the emergence of cooperation among unrelated individuals can be explained by the role of reputation (Nowak and Sigmund, 1998a; Fehr and Gächter, 2000; Fehr and Fischbacher, 2003; Dal Bó, 2005; Henrich et al., 2006). Distinct lines of inquiry on reputation-based cooperation have shown that both indirect reciprocity (Nowak and Sigmund, 1998b; Milinski et al., 2002; Sylwester and Roberts, 2013) and punishment (Sigmund et al., 2001; Brandt et al., 2003; dos Santos et al., 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.

Empirically, 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 et al. (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 et al. (2018) find that reputational considerations can impede coordination on socially optimal conventions across, but also within cultural groups. The present paper adds to this empirical literature by using observational data to show that cross-societal incompatibility in the importance of punishment has implications for multiple domains of social interactions, from low-intensity tensions and discrimination to larger scale conflict.

This study also relates to work studying the role of culture in explaining cross-societal differences in cooperation. Most prominently, Tabellini (2008) provides an explanatory framework for variation in the scope of cooperation across societies based on the interaction of values of good conduct (altruism) and material incentives. Enke (2019) builds on this framework to highlight the role of kinship systems for the evolution of different types of moral values, distinguishing between universal values – altruism

that extends to more socially distant partners – and communal values – altruism confined to a narrow ingroup. Central in this line of research is the role of preferences. A related literature in political science has instead emphasized that, even when values are identical, cooperation rates across societies may differ due to underlying differences in payoff structures related to the external environment and material constraints (Bednar and Page, 2007, 2018). In our model, societies have identical preferences and constraints, but diverge in the way they use reputation to solve cooperation problems. Different reputation mechanisms can be equally effective in sustaining cooperation within groups, but not across them. Our approach also complements this literature because it extends the scope of the research question, from identifying variation in culture, to studying the implications of this variation for cross-cultural contact.

A distinct feature of our study relative to the aforementioned work in cultural economics is that it highlights the existence of cross-cultural variation in mechanisms that support observationally equivalent group-level behaviors and outcomes. As such, our approach 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 difference.

Our theory and empirical results also relate to a voluminous literature on the determinants of conflict. Conflict frequently occurs across rather than within groups, with group membership determined by salient characteristics like ethnicity, skin color or religion.[1] With boundaries taken as given, material differences or inequalities in endowments drive clashes between groups (Mitra and Ray, 2014; Huber and Mayoral, 2014; Morelli and Rohner, 2015; Guariso and Rogall, 2017; McGuirk and Nunn, 2020;

[1]Conflict along ethnic lines can be explained by the instrumental role of ethnicity in providing excludable access to resources (Fearon, 1999; Caselli and Coleman, 2013).

Eberle et al., 2020). Yet the literature has also identified cultural drivers of intergroup conflict. Moscona et al. (2017) and Moscona et al. (2019) have shown that certain types of social organization are more conflict-prone. Cao et al. (2021) provide evidence that a culture of honor, which makes use of violent retaliation and is most commonly found among herding societies, is associated with higher conflict incidence. Our study differs from these works in two respects. First, it focuses on the instrumental role of retaliation for reputation-building and distinguishes it theoretically and empirically from other elements of culture that drive proneness to violence. Second, by shifting the focus from group-specific cultural characteristics to the interaction between different cultures, it shows that social tensions and conflict associated with certain cultural features of groups may be more common in interactions with incompatible groups.

Two studies on intergroup conflict are most related to the theoretical setup in this paper. Rohner (2011) examines how reputation based on retaliation can support cooperation in bilateral interactions. In that model, individuals can form a reputation both in their own group and in other groups they interact with; the likelihood of intergroup conflict is affected by the frequency of interactions across groups captured by measures of diversity and segregation. Our model focuses not on the type of intergroup interactions, but in the content of reputation, which varies across groups. Fearon and Laitin (1996) also view reputation as a group-specific feature. They 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 identifies conditions under which such cooperation is more likely to break down.

Empirically, our study closely relates to work connecting the incidence of conflict to cultural distance across groups. Several studies view cultural distance as distance in preferences over goods (Esteban and Ray, 2011; Spolaore and Wacziarg, 2016, 2017), with theoretically mixed conclusions for the likelihood of conflict. Similar preferences may increase conflict over rival goods, but reduce it over public goods. Guarnieri (2022) provides evidence for this by showing that the likelihood of civil conflict over power (and thus over access to the provision of public goods) increases in the cultural distance of rebel groups from the government. Our study emphasizes a concrete dimension of heterogeneity across populations, reputation, and attempts to empirically isolate that

channel from overall differences in preferences and culture. In a similar vein, but focusing on a different element of culture, Guarnieri and Tur-Prats (2020) show that differences across groups in gender norms drive conflict-related sexual violence.

Finally, and beyond conflict, our conclusions have implications for intergroup tensions in other settings of intergroup contact, such as immigration. Research has found some groups of immigrants and refugees to be more violent-prone than others due to group-specific cultural characteristics of collective experiences (Miguel et al., 2011; Grosjean, 2014; Couttenier et al., 2019). We highlight instead incompatibilities in cultural characteristics between immigrant and host populations. Citizens of western countries tend to discriminate against Middle Eastern, Balkan and Eastern European immigrants (Hainmueller and Hangartner, 2013; Bansak et al., 2016), a propensity that has been attributed to religion (Adida et al., 2010, 2014) or broad civilizational clashes (Huntington, 1996). Our framework narrows in on concrete cultural differences related to how members of different groups use reputation, as a complementary factor that can drive conflict between immigrants and natives.

3 Reputation types and their interaction across cultures

3.1 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 et al., 2002; Dal Bó, 2005). Cooperation can be efficiently sustained through different types of reputation systems. On the one hand, building a 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.

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 et al., 1984), the Middle East (Stewart, 1994) and Central Asia (Keiser, 1991), and are commonly associated with subsistence systems reliant on pastoralism (Cao et al., 2021). Cultural psychologists Nisbett and Cohen have identified a culture of honor descending from Scots-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.

3.2 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). The literature has documented multiple triggers of these conflicts, including uneven access to state institutions and fights over resources triggered by climate change (Eberle et al., 2020; McGuirk and Nunn, 2020). However, participants themselves have also 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 violence of herders has often been documented (Benjaminsen et al., 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 necessarily 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 societies of the Western world, where reputations based on punishment typically play only a minor role. 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).

4 Model

In this section we present a model of intergroup cooperation and conflict. First, we describe how reputation can help sustain cooperation within a group. We distinguish two types of reputation that differ in their informational content: whether group members cooperated in past interactions, or whether group members have punished others who defected. Second, we show how cooperation can be sustained across group boundaries, as long as both groups rely on the same reputational mechanism. Finally, we show how mutual cooperation breaks down in interactions between members of two groups that are of different reputational types.

4.1 Setup

Consider a group of people playing a social matching game. Each period, all 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 simultaneously decide whether or

not to contribute to a joint project. Each contribution yields output R, $2 > R > 1$, and the pair's total output is equally split among the two players. Individuals are heterogeneous with respect to contribution cost ξ_i, where ξ_i is drawn from a distribution with cumulative distribution function G and associated density function g, $G(\xi) = 0$ for $\xi < R/2$ and $G(\xi) = 1$ for $\xi \geq R$. We assume that G is continuous everywhere.

In addition to contributing (c) and non-contributing (d), individuals may also choose action a and avoid the prisoner's dilemma round, in which case their opponent earns a payoff of zero.[2] Avoiding the first round interaction requires effort from the player who chooses to do so, and therefore entails a cost $\nu \geq 0$. For simplicity of exposition, we assume that a player who opts for a receives a payoff that is equal to the lowest possible payoff when participating in the prisoner's dilemma game, i.e. $\nu = \xi - R/2$. The complete payoff matrix for the row player in the first round of the interaction is then given by Table 1.

Table 1: First stage game

	c	d	a
c	R-ξ_i	R/2-ξ_i	0
d	R/2	0	0
a	$-\nu$	$-\nu$	$-\nu$

Aggregate payoffs are maximized if both players contribute, but players have incentives to opt for the opportunistic choice d since $R/2 > R - \xi_i$ and $0 > R/2 - \xi_i$ for $\xi_i > R/2$. We study how a reputation mechanism can help groups to achieve mutual cooperation. As will become clear, the option of avoiding the first round interaction a gives players the possibility to withhold cooperation from an uncooperative opponent without appearing opportunistic, and will play an important role in a reputational equilibrium.

After the prisoner's dilemma, each pair immediately plays a second round, where players can choose to punish their opponent. A punisher i must bear a cost $\kappa_i > 0$ to inflict damage $d > 0$ to the punishee. The individual's cost of punishment is independently drawn from a distribution with continuous cumulative distribution function $H(\kappa_i)$ with $H(\kappa) = 0$ for $\kappa \leq 0$ and $H(\kappa) = 1$ for $\kappa \geq d$. We assume that κ_i and ξ_i are

[2]This is similar to e.g. Rohner (2011).

independently distributed, private known, but not publicly observed. Both distribution functions G and H are identical across individual and publicly known.

Players are infinitely lived and maximize expected payoffs valued with a time discount factor $0 < \delta < 1$. The group is assumed to be 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 with the player in the past – can only play a minor role, and we fully abstract from it for simplicity. Instead, players can learn the reputation that an opponent has within the group. That is, cooperation may be sustained through indirect reciprocity, as reputational concerns can incentivize individuals with sufficiently low contribution cost ξ_i to cooperate. We view reputation as a collective institution, which requires members of the group to observe, record, and transmit information about past actions of members of the group. A reputation mechanism is thus a group-specific characteristic.

We consider two different types of reputation. In the first, which we refer to as cooperation-based reputation, or "coop-rep", a group keeps track of the first round actions of its members, namely of who has cooperated and who has defected in the past. Alternatively, a group with punishment-based reputation ("pun-rep") records the second-round responses of its players, keeping track of who has punished an uncooperative opponent, and who has failed to do so. We assume that a group does not have sufficient informational capacity to keep track of both types of information, but instead has to rely on either one or the other reputational mechanism.

Cooperation-based reputation. Consider a reputational mechanism based on first round actions, where a player's reputation turns positive (*coop*+) if the player cooperates and negative (*coop*−) if they defect. Reputation is not changed if either player avoids the first round. Players know the reputation of their opponent and can condition their actions on this binary information.

While there always exists a pooling equilibrium where all players defect and have *coop*−, we are interested in a separating equilibrium with a cutoff $\bar{\xi}$, so that players with $\xi_i \leq \bar{\xi}$ cooperate against players with *coop*+ and avoid interactions with players with *coop*−. Players with $\xi_i > \bar{\xi}$ defect independently of the opponent's type, and no player punishes during the second round. In this equilibrium, all players with $\xi_i \leq \bar{\xi}$ have a *coop*+ reputation and cooperate with each other, while players with $\xi_i > \bar{\xi}$ have *coop*−.

Under what conditions is this an equilibrium, so that it is optimal for players

with $\xi_i \leq \bar{\xi}$ to cooperate with a *coop*+ player and to avoid a *coop*− player? Denote the maximum expected discounted lifetime payoff of an individual i with *coop*+ and *coop*−, respectively, at the beginning of a period with $V^+(\xi_i)$ and $V^-(\xi_i)$. After being matched with an opponent, the reputational standing of the opponent is denoted in parenthesis, so that, for example, the expected discounted payoff of a *coop*+ player i that is paired with a *coop*− opponent is denoted by $V^{+(-)}(\xi_i)$.

Assuming that others play according to the suggested equilibrium strategy, an opponent with *coop*+ has $\xi_j \leq \bar{\xi}$ and therefore cooperates with a *coop*+ player and avoids a *coop*− player. Similarly, an opponent with *coop*− defects against either type of player. An opponent is of type $\xi_j \leq \bar{\xi}$ with probability $G(\bar{\xi})$. Then, $V^+ = G(\bar{\xi})V^{+(+)} + (1 - G(\bar{\xi}))V^{+(-)}$ with

$$V^{+(+)}(\xi_i) = \max_{c,d,a} \begin{cases} R - \xi_i + \delta V^+(\xi_i) & \text{if } c \\ R/2 + \delta V^-(\xi_i) & \text{if } d \\ R/2 - \xi_i + \delta V^+(\xi_i) & \text{if } a \end{cases} \tag{1}$$

$$V^{+(-)}(\xi_i) = \max_{c,d,a} \begin{cases} R/2 - \xi_i + \delta V^+(\xi_i) & \text{if } c \\ 0 + \delta V^-(\xi_i) & \text{if } d \\ R/2 - \xi_i + \delta V^+(\xi_i) & \text{if } a \end{cases} \tag{2}$$

Similarly, $V^- = G(\bar{\xi})V^{-(+)} + (1 - G(\bar{\xi}))V^{-(-)}$ with

$$V^{-(+)}(\xi_i) = \max_{c,d,a} \begin{cases} 0 + \delta V^-(\xi_i) & \text{if } c \\ 0 + \delta V^-(\xi_i) & \text{if } d \\ R/2 - \xi_i + \delta V^-(\xi_i) & \text{if } a \end{cases} \tag{3}$$

$$V^{-(-)}(\xi_i) = \max_{c,d,a} \begin{cases} R/2 - \xi_i + \delta V^+(\xi_i) & \text{if } c \\ 0 + \delta V^-(\xi_i) & \text{if } d \\ R/2 - \xi_i + \delta V^-(\xi_i) & \text{if } a \end{cases} \tag{4}$$

The optimal action is c in (1) and (4) and a in (2) if $\xi_i \leq R/2 + \delta(V^+ - V^-)$, and d in all three cases otherwise. Notice that we assume that a player chooses a over c in (2) whenever the two actions yield the same payoff. A player is always indifferent between c and d in case (3), since the opponent plays a and reputations are not updated. This

defines the threshold

$$\bar{\xi} = R/2 + \delta(V^+(\bar{\xi}) - V^-(\bar{\xi}))$$

which together with (2), (3), and (4), implies that $V^{-(-)}(\bar{\xi}) = V^{-(+)}(\bar{\xi}) = V^{+(-)}(\bar{\xi}) = 0$ and therefore $V^-(\bar{\xi}) = 0$. Further from (1) we have $V^+(\bar{\xi}) = G(\bar{\xi})V^{+(+)} = G(\bar{\xi})R/2$, so that the threshold is given by

$$\bar{\xi} = (1 + \delta G(\bar{\xi}))\frac{R}{2} \equiv M(\bar{\xi}) \tag{5}$$

We notice that $M(.)$ is non-decreasing in $G(\xi)$ and therefore also in ξ.

Proposition 1. *There is at least one coop-rep equilibrium with positive cooperation rates if there is a $\xi \geq R/2$ such that*

$$G(\xi) > \frac{2}{\delta R}\left(\xi - \frac{R}{2}\right)$$

.

Proof. Notice that $M(\xi)$ has upper bound $(1+\delta)\frac{R}{2}$, so that $\xi > M(\xi)$ for any $\xi > ((1+\delta)R/2)$. By the condition of the proposition, $\xi < M(\xi)$ for some $\xi \geq R/2$. As $M(\xi)$ is continuous, the intermediate value theorem implies that there is at least one $\bar{\xi} > R/2$ with $G(\bar{\xi}) > 0$ such that $\bar{\xi} = M(\bar{\xi})$. □

Proposition 1 states that an equilibrium with positive cooperation rates exists, as long as there is a sufficient mass of individuals with relatively low contribution cost. Notice that for any $\xi \in [R/2, R]$, the length of the support that is to the left of ξ is a share $\frac{\xi - R/2}{R/2}$ of the total support. The condition of the proposition then requires that there is a ξ where the share of the population with $\xi_i < \xi$ is larger than the share of the support to the left, scaled by a factor $1/\delta$. For the rest of this paper, we will assume that this condition is satisfied, so that a cooperative equilibrium exists.

We refer to the socially optimal equilibrium, namely the equilibrium with highest $\bar{\xi}$, as the cooperative equilibrium with associated cut-off ξ^c. That is, we define $\xi^c \leq (1+\delta)R/2$ such that $\xi^c = M(\xi^c)$ and $\xi > M(\xi)$ for all $\xi > \xi^c$. In the cooperative equilibrium, cooperation rates are given by $(G(\xi^c))^2$, while defection rates are given by $(1 - G(\xi^c))$.

Proposition 2. *In a coop-rep group, cooperation rates (defection rates) in the cooperative equilibrium are increasing (decreasing) in the return to the common project R and in the discount factor δ.*

Proof. The function M(.) is increasing in R and δ. The respective derivatives are given by

$$\begin{aligned} \frac{\partial M(.)}{\partial R} &= (1 + \delta G(\xi_c)) && > 0 \\ \frac{\partial M(.)}{\partial \delta} &= G(\xi_c)(R/2) && \geq 0 \end{aligned}$$

The result of the proposition then follows directly from the fact that $\partial M/\partial \xi < 1$ at the cooperative equilibrium. □

Punishment-based reputation. Next, we turn to a reputational mechanism that is based on second round actions, which captures whether individuals punish opponents who don't contribute. We consider again a simple reputational score. In this case, a player gains a positive reputation ($pun+$) if they punish an opponent who does not cooperate. If a player fails to punish a non-cooperator, they earn a negative reputation ($pun-$). While there is a trivial pooling equilibrium where all individuals defect and nobody punishes, we are interested in the existence of a separating equilibrium with a threshold $\tilde{\kappa}$, where players punish non-cooperation if and only if their punishment cost κ_i is below the threshold. Notice that this reputation does not contain information on whether an individual contributes. Therefore, in such an equilibrium, all players defect against a non-punishing opponent ($pun-$), since there are no incentives to cooperate. Players with contribution cost below some threshold $\tilde{\xi}$ will cooperate against players with $pun+$, in order to avoid being punished, while the others defect.

Under what conditions is this optimal? Defecting against a $pun+$ opponent increases first round payoffs, relative to cooperating, by $\xi_i - R/2$, but comes at the cost of being punished and thus suffering damage d. The threshold for contributing when paired with a $pun+$ player is then

$$\tilde{\xi} = d + R/2 \tag{6}$$

That is, a player with $\xi_i \leq \tilde{\xi}$ is a conditional cooperator, who contributes when paired with a $pun+$ player and defects otherwise. Players with $\xi_i > \tilde{\xi}$ defect inde-

pendently of the opponent's reputation. Reputation only matters if the opponent is a conditional cooperator, and the value of having *pun+* rather than *pun−* is therefore

$$V^+(\kappa_i) - V^-(\kappa_i) = G(\tilde{\xi})\left(\frac{R}{2} + \delta V^+(\kappa_i)\right) - G(\tilde{\xi}) \max_{p,np} \begin{cases} -\kappa_i + \delta V^+(\kappa_i) & \text{if } p \\ \delta V^-(\kappa_i) & \text{if } np \end{cases} \quad (7)$$

The last term represents the decision of a *pun−* player of whether to punish and restore reputation or not, and determines the threshold level for punishment cost as

$$\tilde{\kappa} = \delta[V^+(\tilde{\kappa}) - V^-(\tilde{\kappa})]$$

Using this in (7) yields

$$\tilde{\kappa} = \frac{\delta G(\tilde{\xi})}{1 - \delta G(\tilde{\xi})}(R/2) \quad (8)$$

and players punish non-cooperation if they have punishment cost $\kappa_i < \tilde{\kappa}$. Cooperation rates in the cooperative equilibrium are $G(\tilde{\xi})H(\tilde{\kappa})$, while defection rates are given by one minus the cooperation rate.

Proposition 3. *Cooperation rates (defection rates) in the cooperative equilibrium of a pun-rep group are increasing (decreasing) in returns R, the discount rate δ and inflicted damage d.*

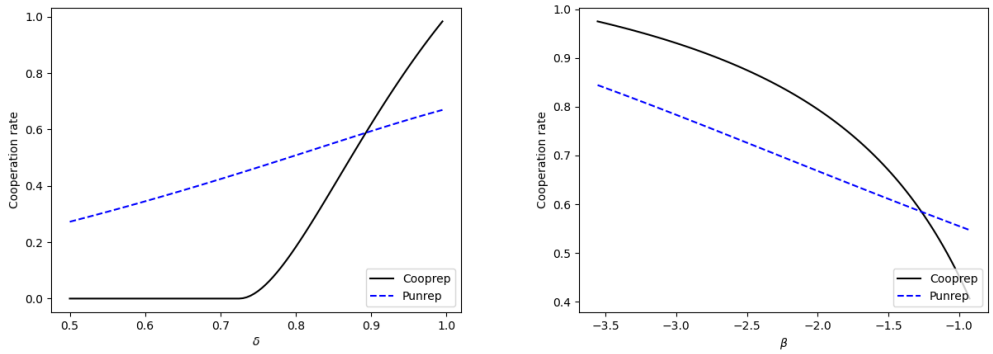
Proof. $\tilde{\xi}$ is increasing in d and in R and constant in δ, $\tilde{\kappa}$ is increasing in δ, in R and in $\tilde{\xi}$. □

4.2 Comparing coop-rep and pun-rep

Neither reputational mechanism is always preferable. Which mechanism can sustain higher cooperation rates depends on the parameter environment, and a direct comparison is difficult as the pun-rep equilibrium depends on an additional parameter (d) and an additional distribution (H).

For illustration, we plot cooperation rates for a parametrization with $R = 1.5$,

Figure 1: Cooperative equilibrium



$d = 0.5$, $\kappa \sim \Gamma(1,1)$ and a linear density function for ξ of the form

$$g(\xi) = \begin{cases} \frac{2}{R} + \left(x - \frac{3R}{4}\right)\beta & \text{if } \frac{R}{2} \leq \xi \leq R \\ 0 & \text{else} \end{cases}$$

for some $\beta \in [-8/R^2, -\delta]$. The coefficient β denotes the slope of the density function, with larger absolute values of β indicating that a larger share of the population have relatively low values of ξ. In the left panel of Figure 1 we fix $\beta = -1.35$ and vary δ, in the right panel we fix $\delta = 0.9$ and vary β. Pun-rep can sustain higher cooperation rates than coop-rep if discount factors are low, but the relationship inverses at high levels of δ. The lack of cooperation under coop-rep with low discount factors mirrors the condition of proposition 1. Our results indicate that with high discount factors, a punishment based mechanism achieves higher cooperation rates. This is in line with the observation that honor cultures, which exhibit a particularly strong emphasis on a reputation based on punishment, developed predominantly in herding economies with low population density, where interactions are rare and discount factors thus should be higher (Nisbett and Cohen, 1996). Cooperation rates are also higher with punishment-based rather than cooperation-based reputation if β is small is small in absolute value, that is, if there are only relatively few individuals with a low contribution cost. This mirrors the finding of Schläpfer (2018) that starting from a pool of defectors, cooperation is likely to arise initially with support from a punishment-based reputation mechanism.

4.3 Reputation and intergroup interactions

To study interactions across group boundaries, we extend the model to include two groups. Occasionally, individuals from different groups will be paired together to play the same game as above. A situation like this may occur when groups occupy neighboring territories, share economic resources, or when members of one group immigrate into a territory currently occupied by another group. If groups use reputation to sustain cooperation in within group pairings, when can cooperation also be sustained across groups?

Denote the share of group 1 in the total population with s_1. Players mainly interact with members of their own group, but occasional cross-group pairings will occur. In each period 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 $r_1 = \gamma(1-s_1)$ for members of group 1 and $r_2 = \gamma s_1$ for members of group 2.

Reputations are assumed to remain group-specific, so that reputational information is only shared among group members, without flow of information across group boundaries. This can be justified by assuming that individuals gossip about group members' actions only with other members of the ingroup, but not with outgroup members. We further assume that groups cannot identify outsiders, so that individuals cannot build a personal reputation across group boundaries. Individuals are aware of average choices of members of the other group in intergroup pairings however, and chose their actions accordingly. Finally, we assume that groups can observe all interactions of their members, also if these interactions take place in a pairing with the outgroup.

As Fearon and Laitin (1996) show for the case of two coop-rep groups, mutual cooperation across groups can be sustained if reputational concerns extend to intergroup pairings. If players fear that they will loose their *coop*+ reputation within their group if they defect against a player from the other group, then they have incentives to cooperate. As we show here, a similar logic applies to pun-rep groups: if failing to punish a defecting outgroup member lowers one's status in their own group, then they have incentives to punish, which in turn incentivizes one's opponent to cooperate.

4.3.1 Two coop-rep groups

Consider first two groups that both rely on coop-rep to sustain cooperation within the group. Under what conditions can an equilibrium be sustained in which players

cooperate in interactions with outgroup members?

Assume that in group 2, a share p_2 of players cooperate in intergroup interactions, while the rest defect. Cooperation is necessarily unconditional, since reputations cannot be formed across group boundaries, and players therefore have no information about their opponent's past actions in an intergroup pairing. We are then looking for a separating equilibrium in group 1, where players with ξ smaller than some threshold $\bar{\xi}_1$ cooperate within their group with a *coop*+ opponent and avoid otherwise, and cooperate unconditionally in intergroup pairings. Players with $\xi > \bar{\xi}_1$ always defect. We consider a mechanism where a player's reputation turns negative whenever they defect, be it in interactions withing the group or in outgroup pairings, and turns positive after cooperation. Note that in the equilibrium that we are considering here, such a reputational mechanism is informationally efficient, since a player that defects against an outgroup member will also defect in a within-group pairing. That is, actions in intergroup interactions reveal the type of the player, which is then accurately reflected in a player's reputation.

Expected discounted payoffs of an individual with positive reputation that is matched with an outgroup member are

$$V_O^+(\xi_i) = \max_{c,d,a} \begin{cases} p_2(R-\xi_i) + (1-p_2)(R/2-\xi_i) + \delta V^+(\xi_i) & \text{if } c \\ p_2(R/2) + \delta V^-(\xi_i) & \text{if } d \\ R/2 - \xi_i + \delta V^+(\xi_i) & \text{if } a \end{cases} \tag{9}$$

Notice that this boils down to the same decision problem as in the ingroup interaction case described in (1), and a player thus cooperates if $\xi_i \leq \frac{R}{2}\delta[V^+(\xi_i) - V^-(\xi_i)]$. Expected payoffs of an individual with a negative reputation are the same, except that reputation remains negative in case the individual plays a. Since a is never optimal, we have that $V_O^+(\xi_i) = V_O^-(\xi_i) = V_O(\xi_i)$.

Encounters with conditional cooperators occur with probability $(1-r_1)G_1(\bar{\xi}_1)$, so that the threshold $\bar{\xi}_1$ is

$$\bar{\xi}_1 = [1 + \delta(1-r_1)G_1(\bar{\xi}_1)]\frac{R}{2} \tag{10}$$

with an equivalent equation defining $\bar{\xi}_2$ so that $p_2 = G_2(\bar{\xi}_2)$.

Proposition 4. *There is at least one coop-rep equilibrium with positive cooperation*

rates for group k if there is a $\xi \leq R/2$ such that

$$G_k(\xi) > \frac{2}{\delta(1-r_k)R}\left(\xi - \frac{R}{2}\right)$$

In particular, there is at least one coop-rep equilibrium with positive cooperation rates if $G_k(R/2) > 0$.

Cooperation rates in the cooperative equilibrium, if it exists, are non-increasing in the share of intergroup pairings r.

Proof. The proof of the existence statement is equivalent to the proof of Proposition 1. For the second part, consider two rates of intergroup interactions r^A and r^B, $r^B > r^A$ and associate cooperative equilibria with threshold levels $\bar{\xi}^A$ and $\bar{\xi}^B$, respectively. Then $\bar{\xi}^A = [1 + \delta(1 - r^A)G(\bar{\xi}^A)]R/2$ and therefore $\bar{\xi}^A < [1 + \delta(1 - r^B)G(\bar{\xi}^A)]R/2$. Since $\xi > [1+\delta(1-r^B)G(\xi)]R/2$ for any $\xi > [1+\delta(1-r^B)]R/2$, it must be that $\bar{\xi}^B > \bar{\xi}^A$. □

Reputation only pays off in ingroup pairings. More frequent intergroup interactions then limit the value of being in good reputational standing, thus reducing cooperation rates.

4.3.2 Two pun-rep groups

Under what conditions can a cooperative equilibrium across groups be sustained with two pun-rep groups? In such a separating equilibrium, *tough* players punish opponents who play d or a in intergroup pairings, in order to keep their *pun+* reputation within their own group, providing incentives for opponents to play c.

Assume that a share π_2 of players from group 2 cooperate in intergroup interactions, while the rest defects. Further, assume that a share q_2 punishes non-cooperation. A player from group 1 will then cooperate in an intergroup pairing if

$$\xi_i \leq \frac{R}{2} + q_2 d \tag{11}$$

which implies that we now have two thresholds given by

$$\tilde{\xi} = \frac{R}{2} + d$$
$$\tilde{\tilde{\xi}} = \frac{R}{2} + q_2 d$$

$\tilde{\tilde{\xi}} \leq \tilde{\xi}$. Individuals with $\xi_i \leq \tilde{\xi}$ cooperate in in-group pairings with a *pun+* opponent, while those with $\xi_i \leq \tilde{\tilde{\xi}}$ additionally cooperate in intergroup pairings.

Opponents from the outgroup do not condition their actions on reputation, so the advantage of entering an intergroup pairing with positive reputation is reduced to costlessly conserving this reputation in case the opponent cooperates, which happens with probability π_2. Therefore we have

$$V_O^+ - V_O^- = \pi_2 \delta (V^+ - V^-)$$

As in the within group case, the surplus value of a good reputation when meeting a conditional cooperator is given by

$$V_{cc}^+ - V_{cc}^- = \frac{R}{2} + \delta (V^+ - V^-)$$

therefore

$$\begin{aligned} V^+ - V^- &= (1 - r_1) G_1(\tilde{\xi}_1) \frac{R}{2} + ((1 - r_1) G_1(\tilde{\xi}_1) + r_1 \pi_2) \delta (V^+ - V^-) \\ &= \frac{(1 - r_1) G_1(\tilde{\xi}_1)}{1 - [(1 - r_1) G_1(\tilde{\xi}_1) + r_1 \pi_2] \delta} \left(\frac{R}{2} \right) \end{aligned}$$

and we find the threshold value of punishing non-cooperation

$$\tilde{\kappa} = \frac{(1 - r_1) G_1(\tilde{\xi}_1) \delta}{1 - [(1 - r_1) G_1(\tilde{\xi}_1) + r_1 \pi_2] \delta} \left(\frac{R}{2} \right) \tag{12}$$

where

$$\pi_2 = G_2 \left(\frac{R}{2} + H_1(\tilde{\kappa}_1) d \right) \tag{13}$$

Proposition 5. *There is at least one solution to the system defined by* (12) *and* (13).

Proof. Define

$$M(\kappa) = \frac{(1 - r_1) G_1(\tilde{\xi}_1) \delta}{1 - [(1 - r_1) G_1(\tilde{\xi}_1) + r_1 G_2 \left(\frac{R}{2} + H_1(\kappa) d \right)] \delta} \left(\frac{R}{2} \right)$$

First notice that since H_1 is non-decreasing in κ and G_2 is non-decreasing in ξ, $M'(\kappa) \geq$

0. Further for any κ we have

$$0 \leq \frac{(1-r_1)G_1(\tilde{\xi}_1)\delta}{1-(1-r_1)G_1(\tilde{\xi}_1)\delta}\left(\frac{R}{2}\right) \leq M(\kappa) \leq \frac{(1-r_1)G_1(\tilde{\xi}_1)\delta}{1-[(1-r_1)G_1(\tilde{\xi}_1)+r_1]\delta}\left(\frac{R}{2}\right)$$

Therefore

$$\begin{aligned} &\kappa \leq M(\kappa) && \text{if } \kappa \leq \frac{(1-r_1)G_1(\tilde{\xi}_1)\delta}{1-(1-r_1)G_1(\tilde{\xi}_1)\delta}\left(\frac{R}{2}\right) \\ &\kappa \geq M(\kappa) && \text{if } \kappa \geq \frac{(1-r_1)G_1(\tilde{\xi}_1)\delta}{1-[(1-r_1)G_1(\tilde{\xi}_1)+r_1]\delta}\left(\frac{R}{2}\right) \end{aligned}$$

and the proposition follows from the continuity of $M(\kappa)$. □

We define the equilibrium with highest levels of $\tilde{\kappa}$, and therefore with the highest level of $\tilde{\tilde{\xi}}$ in the other group, as the cooperative equilibrium $\tilde{\kappa}^C$. That is, $\tilde{\kappa}^C = M(\tilde{\kappa}^C)$ and $\kappa > M(\kappa)$ for any $\kappa > \tilde{\kappa}^C$.

Cooperation rates in pairings within and across groups are then given by, respectively,

$$\begin{aligned} &G_i(\tilde{\xi}_i)H_i(\tilde{\kappa}_i^C) && \text{within group } i \\ &\frac{1}{2}(G_i(\tilde{\tilde{\xi}}_i^C)+G_j(\tilde{\tilde{\xi}}_j^C)) && \text{across groups } i \text{ and } j. \end{aligned}$$

Notice that it is possible that cooperation rates across groups are higher than within group.

Proposition 6. *If the groups have the same size and the same distributions of ξ and κ, then cooperation rates across and within group are both decreasing in the rate of intergroup interactions r.*

Proof. Since $\tilde{\tilde{\xi}} \leq \tilde{\xi}$, $G(\tilde{\tilde{\xi}}) \leq G(\tilde{\xi})$.Consider any r_A, r_B with $r_A > r_B$, and associated cooperative equilibrium thresholds $\tilde{\kappa}_A^C$ and $\tilde{\kappa}_B^C$.Then $\tilde{\kappa}_A^C = M(\tilde{\kappa}_A^C, r_A) < M(\tilde{\kappa}_A^C, r_B)$, which implies that $\tilde{\kappa}_A^C < \tilde{\kappa}_B^C$ by the definition of the cooperative equilibrium. □

This follows the same logic as in the two coop-rep groups case, where more intergroup interactions limit the scope of reputaitonal concerns. Nevertheless, an equilibrium with mutual cooperation in some intergroup pairings remains feasible.

4.3.3 One coop-rep and one pun-rep group

In the case of one coop-rep and one pun-rep group, mutual cooperation in intergroup pairings cannot be sustained. 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, since they will not get punished and 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.

In order to sustaining a *pun+* reputation withing their group, members of the pun-rep group have incentives to punish in intergroup pairings. In particular, a pun-rep group member punishes non-cooperation if $\kappa_i \leq \tilde{\kappa}$, where $\tilde{\kappa}$ is defined by (12), when referring to the pun-rep group as group 1 and to the coop-rep group as group 2.

Coop-rep group members have two sources of incentives to cooperate in intergroup pairings: to sustain their reputation withing their own group, and to avoid being punished in case their opponent has punishment cost $\kappa_i \leq \tilde{\kappa}$. Since opponents from the other group always defect, the expected lifetime payoff when matched in an intergroup pairing is given by

$$V_O = \max_{c,d,a} \begin{cases} \frac{R}{2} - \xi_i + \delta V^+ & \text{if } c \\ -q_1 d + \delta V^- & \text{if } d \\ \frac{R}{2} - \xi_i - q_1 d + \delta V^+ & \text{if } a \end{cases}$$

where $q_1 = H_1(\tilde{\kappa}_1)$. The threshold on contribution cost ξ_i for cooperating in intergroup pairings is therefore

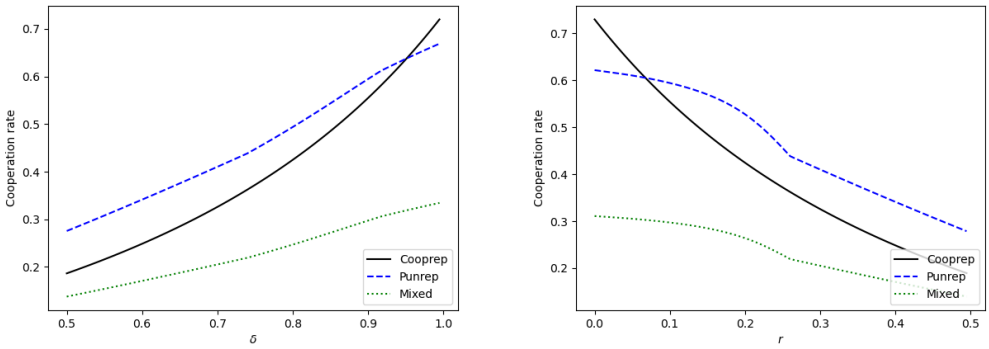
$$\tilde{\tilde{\xi}} = \frac{R}{2} + q_1 d + \delta(V^+ - V^-)$$

while

$$\tilde{\xi} = \frac{R}{2} + \delta(V^+ - V^-)$$

is the relevant threshold in pairings within group 2. Clearly $\tilde{\tilde{\xi}} > \tilde{\xi}$ whenever $q_1 > 0$. This implies that a share $G_2(\tilde{\tilde{\xi}}) - G_2(\tilde{\xi})$ will cooperate in intergroup pairings but not if paired with a member of their own group. But this cannot be an equilbrium, since it is no longer informationally efficient for actions in intergroup pairings to affect reputation. If a member of the coop-rep group is paired with an individual from their own group

Figure 2: Cooperation in intergroup pairings



who has defected in the past in ingroup pairings, but has cooperated with members of the outgroup, then this opponent must have an ξ_i with $\tilde{\tilde{\xi}} \geq \xi_i > \tilde{\xi}$. Therefore, this individual will defect again with ingroup members, and the action in the outgroup games should be discarded. The reputational information will therefore only represent past actions in within group pairings.

The incentives for playing c when paired with a member of the pun-rep group are then only due to the avoidance of punishment, and the relevant threshold is therefore

$$\tilde{\tilde{\xi}} = \frac{R}{2} + q_1 d$$

Punishment of group 1 members and cooperation of group 2 members in intergroup pairings are determined by (12) and (13), and cooperation rates are exactly half of those in the case with two pun-rep groups.

Figure 2 shows cooperation rates in intergroup pairings for our baseline calibration with varying levels of δ (left panel, $r = 0.2$) and r (right panel, $\delta = 0.9$). As in the single group case, the parametrization determines whether coop-rep or pun-rep performs better. The smallest cooperation rates are unambiguously found in the case of mixed reputational systems, where one side, the pun-rep group members, always defect.

4.3.4 Selective intergroup interactions

Can cooperation survive if players can influence the probability with which they are matched to outgroup members? It is possible in the case of two coop-rep groups, but only as a borderline case. Assuming that the two groups are identical, so that

$G_1(\bar{\xi}_1) = G_2(\bar{\xi}_2)$, we find that conditional cooperators are indifferent between interacting with ingroup or with outgroup members. Defectors on the other hand strictly prefer intergroup pairings, since members of the other group cannot selectively avoid interacting with them. If players can influence the chance of being paired across group boundaries, then players with $\xi > \bar{\xi}$ will seek out intergroup pairings, and the cooperative equilibrium only survives under the assumption that conditional cooperators, who are indifferent, will do the same. Otherwise, the share of defectors in intergroup pairings will increase, so that conditional cooperators prefer within group pairings.

In the case of two pun-rep groups, two opposing forces are at play. Ingroup pairings benefit conditional cooperation, as having access to the opponent's reputation allows for defecting against an opponent that does not punish. But non-punishers prefer pairings with outgroup members, since then their negative reputation cannot be exploited. This in turn makes intergroup pairings more attractive for all players, including conditional cooperators. Which of the two forces prevails, or whether they will balance out, depends on the parameter specification.

Results are unambiguous in the case of one coop-rep group and one pun-rep group. All members of the coop-rep group will avoid interactions across group boundaries, since their opponents will always defect and some also punish non-cooperation. If possible, the coop-rep group will therefore opt to shut down any interactions with the pun-rep group.

4.3.5 Gossip sharing across group boundaries

So far we assumed that reputational information is not shared across group boundaries. Here, we show that the main result holds if we relax this assumption, and instead assume that in inter-group pairings, players can learn the reputational status of their opponent within the other group.

In the cases of two coop-rep grous, full information sharing implies that player can condition their actions in intergroup pairings in the same way as in pairings within group. As a result, a player has the same information and faces the same incentives with all opponents. With respect to reputation, the two groups effectively merge into one large group. Under the assumption that both groups are identical, so that $G_{1+2}(\xi) = G_1(\xi) = G_2(\xi)$, the new equilibrium will then be defined by the same cutoff (5) as in the single group case, and cooperation rates in intergroup interactions are the same as within single groups.

Similarly, if both groups are pun-rep, learning whether an opponent is a punisher or not allows players to apply the same strategy in across-group pairings as within-group. Assuming identical distribution of parameters within groups, the resulting equilibrium is defined be (6) and (8), and cooperation rates are the same as in the single group case.

In both these cases, sharing of reputational information across groups is beneficial for cooperation. But mutual cooperation remains unfeasible in the case of a coop-rep and a pun-rep group. In intergroup interactions, pun-rep group members can now learn whether their opponent is known to be a cooperator of defector, but this provides them with no incentives to adjust their behavior. Namely, they lack incentives to cooperate, and will thus defect in all intergroup pairings. Members of the coop-rep group instead are now able to condition their actions in intergroup pairings on the pun-rep status of their opponent. Equation (6) then determines whether the coop-rep player cooperates or defects across group, and cooperation rates in interactions across group will be half of what they would be with two pun-rep groups, assuming that parameter distributions are the same across groups.

5 Empirical patterns

In this section, we provide evidence that our theory is consistent with empirical patterns pertaining to different types of intergroup tensions. First, we show that cultural incompatibility across societies in terms of reputation systems is correlated with the intensity of 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. Second, we show that incompatibility in reputation mechanisms is a consistent predictor of bilateral conflict across ethnic groups. Differences in reputation systems are not mere proxies for bilateral distances in other dimensions of culture or for intergroup differences in other common predictors of conflict.

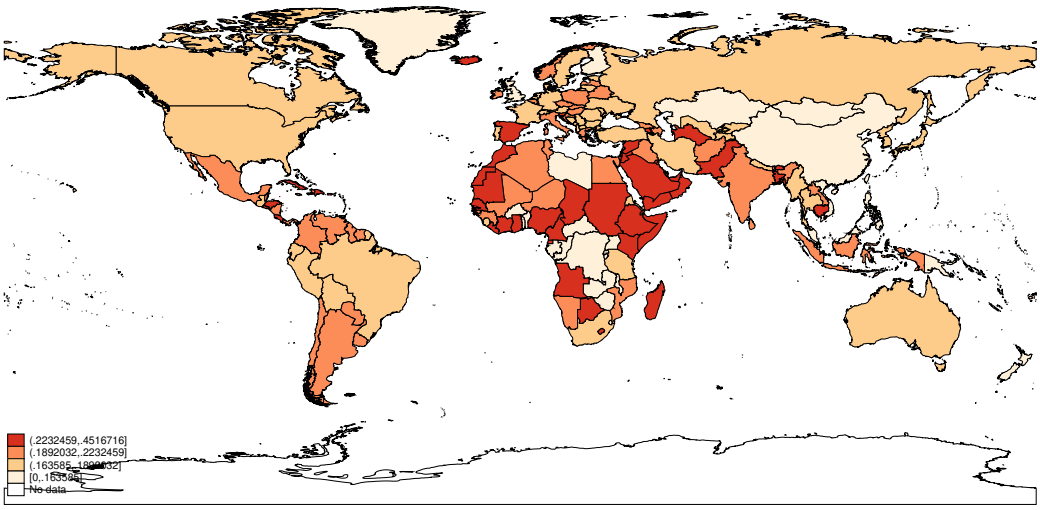
5.1 Measuring reputation mechanisms

A major challenge in testing our argument is that we have no precise empirical way of classifying societies as pun-rep or coop-rep. 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 information, we employ an imperfect proxy of punishment-based reputation using the prevalence of motifs related to retaliation and revenge in a society's oral traditions. We rely on a catalog of folklore, defined as stories and beliefs orally transmitted across generations, which was recently introduced in the economics literature by Michalopoulos and Xue (2021) and which records thousands of motifs for roughly 1,000 societies worldwide. Michalopoulos and Xue (2021) provide frequencies of motifs related to different concepts at the level of a country and at the level of an ethnic group in the *Ethnographic Atlas* (Murdoch, 1967), one of the most widely used anthropological datasets in cross-cultural research in economics. This allows us to use differences in the prevalence of revenge-related motifs between pairs of countries or ethnic groups as a proxy for differences in reputation mechanisms.

Figure 3: Revenge motifs by country



Notes: Revenge-related motifs relative to all motifs recorded in the folklore catalog of Michalopoulos and Xue (2021), by country.

For each society (country or ethnic group), we compute the total number of folklore motifs that relate to revenge, by summing up motifs from the catalog that are tagged

by the keywords *revenge*, *punish*, *punishment*, *penalty*, *retaliate*, *retaliation*.[3] We divide this number by the total number of motifs recorded for each society to account for variation in the richness of oral traditions across societies. Figure 3 depicts the distribution of our measure of revenge-related oral traditions across countries. Unsurprisingly, revenge-related themes are more prevalent in regions more commonly associated with honor cultures, such as the Middle East, large parts of Africa and Southern Europe. However, the measure displays variation within both continents and sub-continental regions.

When possible, we complement this measure with two additional proxies for pun-rep cultures. These measures are available for a limited number of societies or at the country level only. This allows us to use them for validation of our baseline results relying on folklore only in analyses that involve country pairs.

The first measure is an indicator for societies known to employ retaliatory practices, as documented by the ethnographic record. Since information on revenge is not available in commonly used ethnographic databases like the Ethnographic Atlas, we construct this measure based on *eHRAF World Cultures*, a large collection of ethnographic material for over 300 societies, organized and indexed by subject at the paragraph level. The catalog of ethnographic work is available online and a summary is provided for each culture's main aspects of social and cultural life.[4] We crawl the summaries for mentions of keywords related to retaliation, focusing on the sections on conflict and social control, which pertain to societies' interactions with outgroups and means of maintaining ingroup cohesion and cooperation.[5] We tag societies as pun-rep if they include any of the selected keywords in either of the two relevant sections

[3]We follow Cao et al. (2021) in the selection of keywords. Their interest is in using folklore to quantify the prevalence of a culture of honor, which they argue is characterized by both retaliation and a general tendency to violence. Consequently, their folklore-based measure combines concepts related to revenge and concepts related to violence and conflict. We distinguish between the two sets of keywords, because we are interested specifically in second-party punishment that serves a reputation-building purpose. In our model, this type of reputation is not related to a higher prevalence of violence in equilibrium. In fact, we show empirically that differences between societies in the prevalence of violence-related motifs do not predict bilateral tensions once differences in reputation are accounted for.

[4]See ehrafworldcultures.yale.edu.

[5]These include practices like gift-giving, mediation by elders, as well as feuds and other retaliatory norms used to maintain social order through the threat of conflict escalation.

of their culture's summary.[6] We are able to construct this measure for 304 societies available in the eHRAF World Cultures collection. We match societies in the eHRAF to ethnic groups in the Ethnographic Atlas, and aggregate the eHRAF indicator to the country level following a procedure similar to Giuliano and Nunn (2018). Further details on the construction of this measure are provided in Section B of the Appendix. The eHRAF-derived indicator of punishment-based reputation is moderately, but significantly correlated with the prevalence of revenge-related motifs in folklore at the ethnic group level, with a correlation coefficient of 0.223 (p-value < 0.0005).

The second alternative proxy of punishment-based reputation comes from the *Global Preference Survey* (GPS) (Falk et al., 2018). The GPS is a representative survey of economic preferences conducted in 76 countries. We use the measure of negative reciprocity, which is a weighted average of responses to three questions. Two of them capture self-assessments of the willingness to take revenge and to punish unfair behavior towards one's self, and correspond well to the role of punishment-based reputation in our model. The third question, willingness to punish unfair behavior towards others, captures third-party punishment and is less related to the mechanisms in our framework. Since the GPS does not provide disaggregated responses by component question, the overall measure of negative reciprocity is a noisy proxy of reputation built on second-party punishment. Nonetheless, we find this measure to have predictive power in our empirical analysis.[7]

Pairwise correlations in absolute bilateral distances between negative reciprocity, revenge-related motifs and the eHRAF revenge indicator at the country level are shown in Table A.2. Summary statistics for all variables used are provided in Tables A.1 (country level) and A.3 (ethnic group level).

[6]Section B of the Appendix provides examples of descriptions of societies tagged as employing revenge practices, as well as descriptions of societies with a zero value in the eHRAF indicator, whose practices correspond to coop-rep cultures in our model.

[7]The predictive power of differences in negative reciprocity for bilateral perceptions of discrimination further improves when we instrument this measure with differences in revenge-related motifs from folklore. This IV strategy isolates variation in the measure that does not relate to third-party punishment.

5.2 Discrimination against immigrants

5.2.1 Data and empirical strategy

Our model predicts that incompatibility in reputation systems leads to the breakdown of cooperation, which manifests as avoidance of pun-rep group members by coop-rep group members. To test this prediction, we proxy for cooperation failure with discrimination against members of one society by those of another.

We use data from eight waves of the *European Social Survey* (ESS), a representative survey of social attitudes conducted in 32 European countries every two years between 2002 and 2016. 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 immigrants to cultural differences between their country of residence and the country of origin of their parents. 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.[8] We run regressions of the form:

$$D_{icot} = \beta_1 R_{co} + \mathbf{X}_{icot}\beta_2 + \mu_c + \nu_o + \theta_t + \mathbf{Z}_{co}\beta_3 + \epsilon_{icot} \tag{14}$$

where subscripts i, c, o and t denote individuals, residence countries, origin countries (of immigrants' parents) and survey years, respectively. D_{icot} is an indicator for second generation immigrants who report belonging to a discriminated group. R_{co} is the absolute difference in the share of revenge-related motifs between residence and origin country, our main measure of difference in reputation mechanisms. The vector $\mathbf{X}_{icot}$ includes a parsimonious set of individual-level controls (age, age squared and gender).

Our identification leverages variation at the country-pair level. μ_c are residence country fixed effects that account for time-invariant characteristics of European host societies, such as low tolerance of immigrants or conservative political attitudes. ν_o are origin country fixed effects which account for the fact that discrimination may be higher against immigrants from poorer or linguistically and religiously more distant

[8] We define as second generation immigrants those with either parent born abroad. When both parents are born abroad, we assign origin country characteristics based on the country of birth of the mother.

societies. θ_t are year fixed effects, that net out time trends in perceptions of discrimination common across all countries. $\mathbf{Z}_{co}$ is a set of controls at the country-pair level. Conditional on the characteristics of the host population and immigrants' origin countries, pair-level characteristics such as a common historical past or geographic proximity may make discrimination more intense against some groups of immigrants than others. Throughout, we cluster standard errors at the country-pair level, but results are robust to two-way clustering at the level of origin and residence countries.

5.2.2 Results

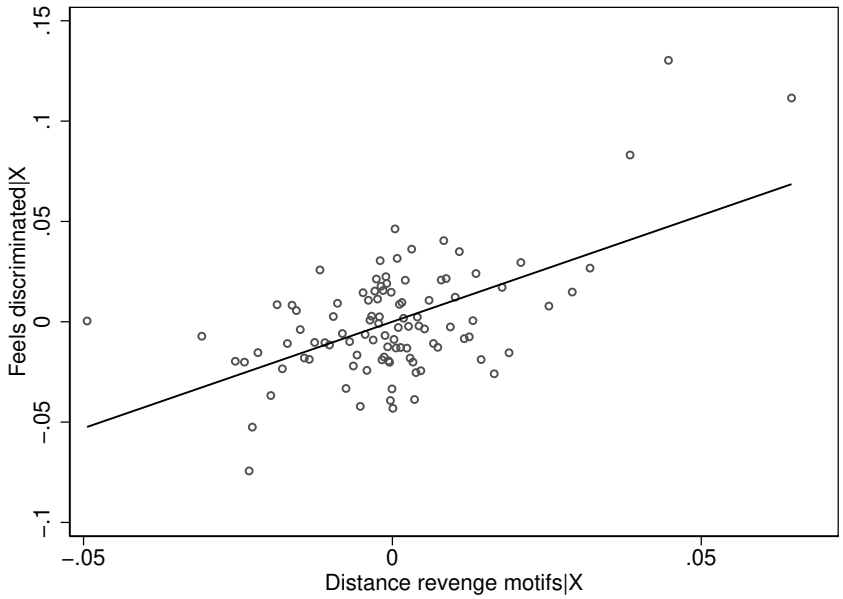
Figure 4 is a graphical overview of our baseline result. It depicts the correlation between perceptions of discrimination and bilateral distance in the prevalence of revenge-related motifs, net of residence, origin and survey year fixed effects. Consistent with our theoretical predictions, the higher the difference between countries in terms of our proxy of reputation systems, the higher is the likelihood of intergroup avoidance, measured as (perceived) discrimination. The relationship is tight and not driven by outlier country pairs.[9]

We explore these patterns more systematically in Table 2. Column 1 reports the baseline correlation between distance in revenge motifs and perceptions of discrimination, accounting for fixed effects and individual controls. Columns 2 to 4 successively add controls at the host-origin country pair level that could affect attitudes and patterns of bilateral interactions between immigrants and host societies. Following Guiso et al. (2009), in column 2 we control for geographic distance and for whether countries share a common border, official language, and colonial past. Consistent with the findings of Guiso et al. (2009) for bilateral trust, immigrants from more distant countries perceive more discrimination; the opposite is true of immigrants from neighboring countries. Immigrants from former European colonies do not report significantly higher discrimination, but immigrants sharing a common language with their host societies do. This is consistent with higher labor market competition between immigrants and natives when their profiles are more similar (Borjas, 2003). Conditional on origin from a former colony, immigrants who share a language with their destination country are

[9]The slope coefficient is $\beta = 1.06$, $p = 0.001$. After trimming the top and bottom 1% in terms of distance in revenge motifs, we have $\beta = 1.15$, $p < 0.001$.

also more likely to be better integrated, a feature correlated with a higher propensity to report discrimination (Lajevardi et al., 2020). In column 3 we control for the absolute difference in log GDP per capita (measured in 2000) between residence and origin country. The relationship between discrimination and bilateral income differences is negative, but not significant. In column 4, 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.[10] Lower religious distance is strongly associated with lower perceptions of discrimination, consistent with work identifying religion as one of the main drivers of discrimination against immigrants in Europe (Adida et al., 2014; Bansak et al., 2016). Across specifications, the positive point estimate of the distance in revenge-related motifs remains highly significant.

Figure 4: Discrimination and differences in revenge prevalence, binned scatterplot



Notes: Perceptions of discrimination among second generation immigrants in the ESS and distance between host and origin country in the prevalence of revenge-related motifs. Figure plots binned residuals of variables net of host and origin country fixed effects and survey year fixed effects.

[10]This measure is from Guiso et al. (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.

Magnitudes are small, but meaningful. As an illustration, consider the case of immigrants from Egypt, a country with a relatively high prevalence of revenge-related motifs (0.205). According to the estimate in column (1), an Egyptian immigrant is around 4 percentage points more likely to feel discriminated in the UK, where the share of revenge motifs is 0.142, compared to Greece, where the respective share is 0.202. Indeed, in our data two out of six immigrants of Egyptian origin in the UK feel discriminated, but only one out of eight in Greece. Conversely, an immigrant from Singapore, a country with a share of revenge motifs equal to 0.106, is predicted to be 1.8 percentage points less likely to feel discriminated in the UK than an immigrant from Egypt. In our data, none out of four Singaporean immigrants in the UK report belonging to a discriminated group.

Table 2: Differences in revenge prevalence and perceptions of discrimination

Dep. variable	Feels discriminated			
	(1)	(2)	(3)	(4)
Distance revenge motifs	0.104***	0.0913***	0.0951***	0.0653***
	(0.0303)	(0.0308)	(0.0315)	(0.0236)
Contiguous		-0.0364***	-0.0393***	-0.0248**
		(0.0129)	(0.0131)	(0.0124)
Common official language		0.0573***	0.0592***	0.0355**
		(0.0168)	(0.0168)	(0.0157)
Former colonial relation		0.00414	0.00543	-0.00418
		(0.0140)	(0.0140)	(0.0141)
Distance between capitals		0.120**	0.135***	0.0551
		(0.0528)	(0.0524)	(0.0462)
Difference log GDP p.c.			-0.00565	-0.00901
			(0.0227)	(0.0232)
Religious similarity				-0.153***
				(0.0268)
Observations	23896	23896	23693	23693
R-squared	0.0649	0.0666	0.0664	0.0692

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Distance revenge motifs* measures the absolute difference between the respondent's residence country and their parents' country of birth in the proportion of folklore motifs that are revenge-related. All columns control for gender, age and age squared, host and origin country fixed effects and survey year fixed effects. Standardized beta coefficients reported. Standard errors are clustered at the country pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

An additional testable implication of our theory is that discrimination, as a proxy of avoidance of interactions, should originate from members of coop-rep societies and be directed towards members of pun-rep societies, rather than vice versa. We ex-

plicitly test this prediction in Table 3 by decomposing the distance in the share of revenge motifs in two parts. We denote by *Distance revenge motifs (+)* the positive difference between the origin and host country's share of folklore motifs related to revenge, and with *Distance revenge motifs (−)* the (absolute value of) the respective negative difference. Higher values of *Distance revenge motifs (+)* imply that the share of revenge-related motifs is higher in the respondent's country of origin than in the residence country. Higher values of *Distance revenge (−)* imply instead the opposite pattern. This specification does not allow us to simultaneously control for residence and origin country fixed effects, which are collinear with the (linear) distance variables. We report results from specifications that separately control for origin country (columns 1–2) and host country fixed effects (3–4).

Consistent with the prediction of the model, perceptions of discrimination are primarily expressed by immigrants from countries whose oral traditions include more mentions of revenge than those of their dyadic partner. The estimated effect of *Distance revenge motifs (−)* instead is never significant and is inconsistent in sign and magnitude across specifications. Discrimination and avoidance appears to be predominantly the response of coop-rep societies towards members of pun-rep societies.

Table 3: Discrimination by higher/lower revenge prevalence

Dep. variable	Feels discriminated			
	(1)	(2)	(3)	(4)
Distance revenge motifs (+)	0.144***	0.115***	0.0781***	0.0454***
	(0.0340)	(0.0264)	(0.0194)	(0.0155)
Distance revenge motifs (-)	0.0238	0562	-0.0111	0.0136
	(0.0180)	(0.0212)	(0.0206)	(0.0170)
Observations	23896	23693	23919	23715
R-squared	0.0542	0.0646	0.0375	0.0541
Origin country FE	✓	✓		
Host country FE			✓	✓
Controls		✓		✓

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Distance revenge motifs (+)* (resp. −) measures the difference in the proportion of folklore motifs between the respondent's residence country and their parents' country of birth when that difference is positive (resp., negative). All columns control for gender, age and age squared and survey year fixed effects. Columns 2 and 4 include all bilateral economic, geo-cultural and religious controls. Standardized beta coefficients reported. Standard errors are clustered at the country pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

In Table 4 we conduct a number of falsification exercises to better justify a role

for our proposed mechanism of incompatibility in reputation mechanisms. Column 1 replicates, for comparability, the baseline result with host and origin country fixed effects and bilateral controls (column 4 of Table 2). In columns 2 and 3 we examine the grounds on which immigrants perceive discrimination in the host country. The ESS asks respondents whether they feel discriminated due to membership in a number of group categories. In column 2 of Table 4 we construct a binary dependent variable for respondents perceiving discrimination on the basis of color or race, nationality, language ethnicity or religion. These categories correspond to social groups, which form the unit of culture in our theoretical framework. In column 3 we examine discrimination on the basis of any of the remaining categories, which include age, gender, sexuality and disability. As expected, discrimination derives primarily from social (ethnoreligious or racial) group belonging and not from other attributes.

Bilateral distance in terms of the prevalence of revenge-related motifs could proxy for cultural distance more broadly, rather than differences specific to the role of reputation. We assess the relevance of this important concern in the remainder of the table. First, we explore whether the salience of retaliation in a society is a proxy for other features of cultures of honor. Empirical research shows that honor cultures are prone to interpersonal within-group violence (Nisbett and Cohen, 1996; Grosjean, 2014) and more belligerent and likely to be involved in intergroup conflict (Cao et al., 2021). Our model instead emphasizes punishment as a reputation mechanism that sustains a cooperative equilibrium, and does not imply a higher equilibrium association of punishment-based reputation with violence and conflict. Conflict in our framework is driven by bilateral differences in reputation types, and not by differences in propensity to violence. In column 4 of Table 4, we compute the share of a country's folklore motifs that relate to violence and conflict, by counting the number of motifs tagged by the concepts *violence*, *violent*, *battle*, *fighting*, *attach*, *soldier*, *guard*, *troop*, *army*, *enemy*, *fighter*, *invasion*, *invade*, *defender*. There is no significant association between bilateral differences in violence-related motifs and perceptions of discrimination.

An equally important cultural difference could be in the propensity for cooperative behavior. In contrast to frameworks like Tabellini (2008), societies in our model do not differ in terms of the distribution of agents' types, but in terms of the reputation mechanisms that sustain cooperation. In practice, punishment-based motifs may be more prevalent in societies where deception is more widespread, since retaliation may develop as a mechanism to deal with uncooperative behavior. Residents of more altruis-

Table 4: Falsification tests

Dep. variable	Feel discriminated														
	Baseline	Ethnoreligious	Placebo										Baseline		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Distance revenge motifs	0.0588***	0.0540**	0.0245			0.0718***							0.0674**	0.0486**	0.0809**
	(0.0203)	(0.0220)	(0.0155)			(0.0222)							(0.0312)	(0.0210)	(0.0345)
Distance violence motifs				-0.0161		-0.0347***									
				(0.0114)		(0.0131)									
Distance cheating motifs					-0.0333	-0.0313									
					(0.0251)	(0.0255)									
Distance patience							0.0919***						0.102***		
							(0.0307)						(0.0288)		
Distance risk-taking								00479					-0.0205		
								(0.0212)					(0.0192)		
Distance positive reciprocity									0414				0.0280		
									(0.0189)				(0.0205)		
Distance negative reciprocity										0.0499***			0.0264		
										(0.0192)			(0.0205)		
Distance altruism											0.0162		0655		
											(0.0325)		(0.0301)		
Distance trust												-0.0161	-0.0134		
												(0.0183)	(0.0157)		
Observations	23693	23997	23997	23693	23693	23693	15951	15951	15951	15951	15951	15951	15849	23693	15849
R-squared	0.0634	0.0815	0.0152	0.0625	0.0625	0.0639	0.0566	0.0553	0.0553	0.0562	0.0553	0.0554	0.0589	0.0672	0.0631
Schwartz Human Values														✓	✓
GPS preferences															✓
Other folklore motifs															✓

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Distance revenge motifs*, *Distance violence motifs* and *Distance cheating motifs* measures the absolute difference between the respondent's residence country and their parents' country of birth in the proportion of folklore motifs that are revenge-, violence- and cheating-related, respectively. Regressions include individual controls for gender, age and age squared, bilateral economic, geo-cultural and religious controls from column 4 of Table 2, origin and host country fixed effects and survey year fixed effects. Column 15 controls for differences in the share of motifs related to violence and cheating, in 21 human values from the Schwartz index and distances in preferences from GPS (Falk et al., 2018). Standardized beta coefficients reported. Standard errors are clustered at the country pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

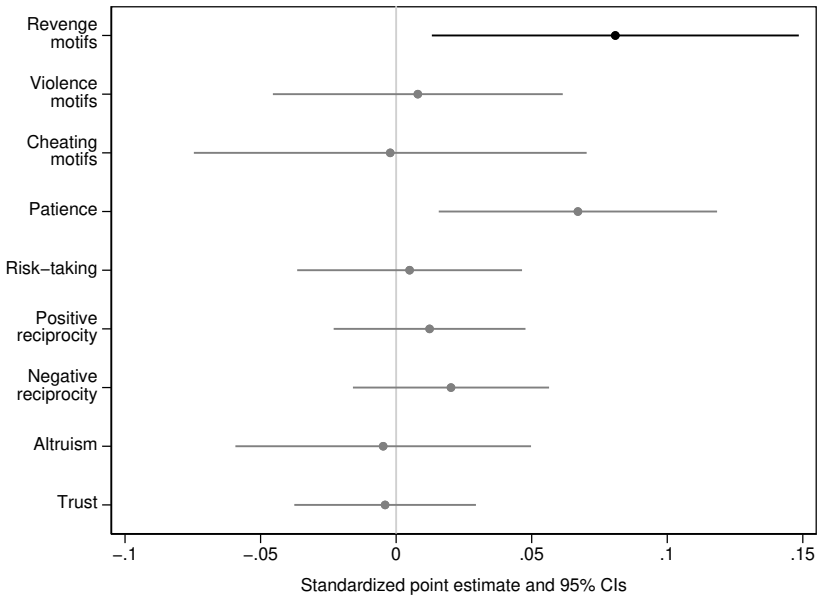
tic or cooperative societies may then be likely to avoid immigrants from societies where cooperative types are less frequent. We proxy for differences in cooperative behavior using the share of motifs related to deception, tagged by the concepts *cheat*, *deceive* and *trick*. As was the case with differences in violence, column 3 of Table 4 reveals that bilateral differences in the prevalence of deception-related motifs are not significantly correlated with perceptions of discrimination. In column 6 we control for distances in all three types of motifs simultaneously. Bilateral distances in revenge continue to predict discrimination. Conditional on the role of revenge, perceived discrimination is decreasing in the distance in violence-related motifs. A potential explanation for this finding is that, for given differences in reputation mechanisms, conflict is highest between two groups with a strong tendency for (non revenge-related) violence; it is instead diminished if one of the groups is less conflict and violence-prone.

In columns 7 to 12 we control for bilateral differences in economic preferences from the GPS. Out of the six economic preferences and beliefs measured as part of the survey, only differences in patience and negative reciprocity correlate with perceptions of discrimination. As discussed in Section 5.1, negative reciprocity encompasses measures of second-party punishment closely related to punishment-based reputation in our model and is a reasonable alternative proxy for reputation types. In column 13 we control simultaneously for differences in all preference dimensions and distance in revenge-related motifs. In this specification, negative reciprocity loses in magnitude and significance. This is consistent with this measure being a less accurate proxy for reputation differences. Patience remains a strong predictor of perceptions of discrimination. Falk et al. (2018) show that patience is strongly correlated with income, as well as with a set of preferences related to individualism and the spirit of capitalism. They demonstrate that this measure is more predictive of comparative development than most other commonly employed attitudinal measures. The results of Table 4 indicate that differences in this central driver of comparative development are also important for intergroup interactions. Yet, these differences appear independent from differences in reputation types, whose estimated effect remains unaffected.

Column 14 adds controls for the absolute bilateral distance in each of twenty-one personal human values, such as the importance of being respected or successful. These questions form 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).[11] Figure 5 visualizes the results for relevant coefficients and shows that the distance in revenge motifs remains one of the two most important predictors of perceptions of discrimination. In Column 15, we jointly control for differences in other folklore motifs, differences in economic preferences, and differences in human values. Overall, these results are encouraging for our analysis as they suggest that avoidance and discrimination are not the product of general cultural differences, but rather the specific outcome of cross-cultural differences in the role of punishment.

Figure 5: Other measures of cultural distance



Notes: Point estimates and 95% intervals from the specification in column 15 of Table 4.

5.2.3 Robustness

In Table 5 we present results using the two alternative proxies of punishment-based reputation. In columns 1 and 2 we use the absolute bilateral distance in negative reciprocity from Falk et al. (2018). As shown in Table 4, distance in negative reciprocity

[11]These questions are phrased as follows: "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." Answers range from 1 (Very much like me) to 6 (Not like me at all).

is positively correlated with perceptions of discrimination. To purge this measure of the component related to third-party punishment, which is extraneous to our theory, we instrument it with the share of revenge motifs in columns 3 and 4. Consistent with measurement error, IV estimates are larger in magnitude than OLS ones.

Table 5: Other proxies of difference in reputation types

Dep. variable	Feel discriminated						
	OLS		2SLS		OLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance negative reciprocity	0.0429**	0.0430**	0.498**	0.334*			0.0619*
	(0.0204)	(0.0170)	(0.210)	(0.174)			(0.0350)
Distance revenge (eHRAF)					0.0344	0.144*	0.238***
					(0.0690)	(0.0737)	(0.0695)
Distance revenge motifs							0.0127
							(0.0382)
Observations	15971	15951	15869	15849	9572	9446	9440
R-squared	0.0558	0.0622			0.0647	0.0677	0.0676
F-statistic			16.03	11.97			
Controls		✓		✓		✓	✓

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Distance negative reciprocity* measures the absolute difference between the respondent's residence country and their parents' country of birth in negative reciprocity, from Falk et al. (2018). *Distance revenge (eHRAF)* is the absolute difference in mentions of revenge in eHRAF societies averaged across all ethnic groups in a country. *Distance revenge motifs* measures the absolute difference between the respondent's residence country and their parents' country of birth in the proportion of folklore motifs that are revenge-related. All columns control for gender, age and age squared, origin and host country fixed effects and survey year fixed effects. Columns 2, 4, 6 and 7 include bilateral economic, geo-cultural and religious controls. Standard errors are clustered at the country pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

We next turn to the proxy of revenge-based reputation we construct using the collection of ethnographies in eHRAF. This measure is available for a limited number of societies, resulting to a loss of more than one third of our sample. While positively correlated with perceptions of discrimination, distance in mentions of revenge in the ethnographic record is only marginally significant after the inclusion of bilateral controls. In column 7, we simultaneously include all three measures of punishment-based reputation in the regression, for the restricted sample for which all three are available. Distance in the eHRAF measure emerges as the strongest predictor of discrimination. Revenge motifs remain positive, but diminish in both magnitude and significance. This may indicate that the eHRAF proxy for pun-rep cultures, which relies on detailed anthropological studies of each cultural group, most precisely captures the dimension of

interest. Nonetheless, these results should be viewed with caution given the limited number of country pairs for which there is overlap in measures.

As a further check, we explore an alternative specification for our main explanatory variable. In place of the absolute distance in revenge motifs R_{co} in equation (14), we use the product of the shares of revenge-related motifs in residence and origin country, $R_c \times R_o$. According to the theory, an increase in the relevance of retaliation in one country should only have a detrimental effect on discrimination if it is of little relevance in the other country. We thus expect the coefficient on $R_c \times R_o$ to be negative. Table A.5 shows results from regressions equivalent to those of Table 5, with all bilateral absolute differences replaced by interactions of residence and origin country controls, consistent with the format used for the main independent variable. The coefficient on the interaction term is significantly negative for all three measures of the salience of revenge. As in Table 5, the eHRAF-derived measure proves to be the most robustly predictive of perceptions of discrimination.

Finally, Table A.4 shows regression estimates from models equivalent to those from our main results in Table 2, but with two-way clustering of standard errors at the host and origin country level. This allows for any correlation of errors across individuals with the same origin or the same current residence. Standard errors are somewhat larger with two-way clustering, but results remain significant at the 5% level.

5.3 Bilateral conflict

Our theory predicts that incompatible reputation mechanisms lead to avoidance and the breakdown of cooperation across groups. If we consider, as Rohner (2011), high rates of defection in intergroup interactions to be a measure of social tensions, then an implication of the framework is that intergroup clashes are more common when reputation mechanisms differ across groups. In this section, we test this prediction by examining whether differences in reputation predict the frequency of intergroup conflict.

5.3.1 Data

Because our interest is in predicting the likelihood of conflict between pairs of groups that differ in their reputation systems, our analysis requires a dyadic dataset of conflict occurrence. We rely on the *Ethnic Power Relations Dataset Family* (EPR) (Vogt et al., 2015), which provides annual data on all politically relevant ethnic groups for

all countries in the world, comprising a total of over 800 groups between 1946 and 2021. Groups are defined as politically relevant if "at least one political organization has claimed to represent their interests at the national level or if their members are subjected to state-led political discrimination" (Vogt et al., 2015). We use the EPR to construct a dyadic dataset in which the unit of analysis in each year is a pair formed by the group dominating the government and each group without government power.

Information on the incidence of conflict between group pairs comes from ACD2EPR, a component of the EPR Dataset Family that links ethnic groups in EPR to rebel organizations in the *Uppsala Conflict Data Program's* (UCDP) actor database. The UCDP is a widely used dataset of georeferenced conflict events that covers the entire world in the period 1989-2016. Events are registered as conflict and included in the UCDP when they have resulted in at least 25 fatalities. This imposes a relatively high bar for testing the implications of our model, which pertain to social tensions that do not necessarily result in widespread high-fatality clashes. Additionally, because the EPR dataset is restricted to politically relevant groups, the types of conflicts that we can examine are armed civil conflicts, that always include the government on the one side and rebel groups inhabiting the same country on the other. Both of these limitations of the setup should be kept in mind when interpreting the results of this empirical exercise.

We link ethnic groups in the EPR to the share of revenge-related motifs in folklore via the Ethnographic Atlas. For Africa, we rely on the concordance table between the two data sources constructed by Michalopoulos and Papaioannou (2016). We extend the match between EPR groups and groups in the Ethnographic Atlas to Asia following a two-step procedure. First, we use the coordinates of each group's ethnic homeland provided in the Ethnographic Atlas to identify all groups that inhabit the area covered by a polygon associated to a group in EPR. Polygons of each group's location are from GeoEPR (Wucherpfennig et al., 2011). We also use information on the language spoken by each group which is available in the Ethnographic Atlas and for EPR groups through the EPR Ethnic Dimensions Dataset (EPR-ED) (Bormann et al., 2017). We consider groups as matched when they inhabit the same polygon and speak the same language. Next, we complement this location and language-based match with a manual match of remaining groups using online sources. Based on this procedure, we are able to match 61% of Ethnographic Atlas groups in Asia.

Having assigned each group in EPR to their counterpart in the Ethnographic Atlas,

we rely on frequencies of folklore motifs at the ethnic group level from Michalopoulos and Xue (2021) to compute the absolute distance in the share of revenge-related motifs between pairs of ethnic groups. Limited overlap between the groups in EPR and the societies available in eHRAF World Cultures prohibits us from using the eHRAF indicator of revenge practices as an alternative proxy of punishment-based reputation.

We complement this data with bilateral distances between groups in other folklore motifs, as well as distances in additional characteristics of ethnic groups from the Ethnographic Atlas. We also construct a number of geographic control variables (terrain ruggedness, elevation, suitability for agriculture) for polygons assigned to each group in EPR. Further details on the construction of the dataset as well as variable descriptions and sources are provided in Section B of the Appendix. Table A.3 reports descriptive statistics.

5.3.2 Empirical strategy

To examine whether differences in reputation mechanisms predict conflict occurrence between ethnic groups, we estimate the following equation:

$$C_{dct} = \delta_{g1} + \delta_{g2} + \lambda_c + \kappa_t + \gamma_1 R_d + + \mathbf{\Omega}_d \gamma_2 + \varepsilon_{dct} \tag{15}$$

where C_{dct} is an indicator for conflict between two groups in pair d in country c at year t. δ_{g1} and δ_{g2} are fixed effects for the group in government and the rebel group, respectively. λ_c are country fixed effects and κ_t are year fixed effects. This specification then accounts for any time-invariant group-specific characteristics that make it more likely that any given group, whether in power or not, is more likely to engage in conflict. Those include time-invariant ethnographic and cultural confounders linked to the propensity to engage in conflict, such as main mode of subsistence (Cao et al., 2021) or kinship structure (Moscona et al., 2019). Country fixed effects rule out the influence of fixed country characteristics that affect conflict propensity, such as level of development, institutional quality or historical legacies of colonialism. Year fixed effects account for shocks common across countries and ethnic groups, such as global economic downturns. In more parsimonious specifications we account for a full set of country-year fixed effects, netting out any time-variant factors that affect all groups within a country. Given that pairs are always formed between ethnic groups out of power and the government, these latter fixed effects rule out the confounding role of changing government characteristics, such as overall quality or military strength.

Our interest lies in the effect of R_d, the absolute distance between ethnic group-pairs in the prominence of revenge in their oral traditions. The vector $\mathbf{\Omega}_d$ includes dyad-specific controls that can explain the occurrence of conflict within a given group pair. These include overall cultural distance, geographic distance, differences in geographic endowments and differences in cultural characteristics relevant for conflict. We discuss these controls as we introduce them below. Throughout, we cluster standard errors at the pair level, but our results are robust to clustering at the country level.

The effect of R_d, the absolute difference of a one-dimensional variable, is identified from two sources of variation. Within a country, we can estimate γ_1 as long as we observe at least two different groups in power and two different groups out of power, and as long as the groups in power do not all rank higher or lower with respect to their share of revenge motifs than all of the out-of-power groups. Additionally, we exploit variation coming from the fact that the same ethnic group is present in more than one country. Table A.6 lists all countries in our data set that exhibit the type of variation we use for identification. We also report results for less stringent specifications that allow for exploiting the variation in the entire data set.

5.3.3 Results

Column 1 of Table 6 presents the correlation between conflict incidence and distance in the prevalence of revenge motifs in folklore, conditional on group, country and year fixed effects. A one standard deviation increase the share of revenge-related motifs increases the likelihood of conflict by 0.25 of a standard deviation. This effect is only significant at the 10% level, but its magnitude is appreciable: it stands between the effect of between-group inequality in rainfall (0.12 standard deviations) (Guariso and Rogall, 2017) and that of linguistic distance (0.36 standard deviations) (Guarnieri, 2022).

Next, we progressively introduce pair-level controls that have both been shown to affect the likelihood of conflict and are plausibly correlated with cross-group differences in the importance of retaliation. In column 2, we control for cultural distances between groups, proxied by linguistic and religious distance. To compute linguistic distance, we merge ethnic groups to the Ethnologue, a dataset of world languages and their classification into linguistic trees, and follow Fearon (2003) to compute cladistic distance between language pairs. We compute religious distance in a similar way, by using the information on religious affiliations of each group provided in the EPR-ED dataset and

following Guarnieri and Tur-Prats (2020). Details on the construction of these variable are provided in the Appendix Section B.3. We also control for the logged geodesic distance between group centroids, based on the polygons of each group's settlement region provided by GeoEPR.

Table 6: Difference in revenge prevalence and conflict incidence

Dep. variable	Conflict incidence					
	(1)	(2)	(3)	(4)	(5)	(6)
Distance revenge motifs	0.249*	0.329**	0.799***	1.117***	1.121***	1.097***
	(0.143)	(0.147)	(0.274)	(0.267)	(0.265)	(0.336)
Linguistic distance		-0.268***	-0.298***	-0.241***	-0.238***	-0.276***
		(0.0994)	(0.0464)	(0.0522)	(0.0530)	(0.0925)
Religious distance		0.990	2.543***	3.576***	3.170***	2.220*
		(0.601)	(0.902)	(0.818)	(0.842)	(1.155)
Log geographic distance		0.421***	0.587***	0.564***	0.560***	0.513***
		(0.110)	(0.134)	(0.125)	(0.124)	(0.166)
Distance jurisdictional hierarchy			-0.778*	-1.140***	-1.132***	-1.160**
			(0.425)	(0.405)	(0.404)	(0.471)
Distance settlement complexity			-0.0337	-0.159	-0.199	-0.304
			(0.217)	(0.223)	(0.222)	(0.287)
Distance High Gods			0.0866	0.658*	0.684*	0.677
			(0.142)	(0.351)	(0.345)	(0.450)
Distance agricultural suitability				-0.139*	-0.153*	-0.173
				(0.0787)	(0.0773)	(0.106)
Distance ruggedness				0.982	1.010	1.144
				(0.740)	(0.728)	(1.254)
Distance elevation				-0.345	-0.311	-0.279
				(0.598)	(0.592)	(1.048)
Distance group size					0.443*	1.210
					(0.246)	(0.840)
Observations	3968	3915	2985	2811	2811	2157
R-squared	0.390	0.393	0.366	0.366	0.368	0.633
Country × Year FE						✓

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Distance revenge motifs* measures the absolute difference between the two ethnic groups in the proportion of folklore motifs that are revenge-related. All columns control for group (government and rebels) fixed effects, country and year fixed effects. Standard errors are clustered at the group pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Conditional on country fixed effects, higher geographic distance between groups increases the likelihood of conflict, suggesting that civil war is more likely to occur between the government and geographically peripheral ethnic groups. Linguistic distance instead lowers the likelihood of conflict, consistent with the findings in Spolaore and

Wacziarg (2016) and Guarnieri (2022), who find that more closely related populations are more prone to engage in conflict as they share similar preferences and may clash over rival goods. Importantly, neither geodesic nor overall cultural distance drive the effect of differences in revenge motifs.

In column 3, we account for differences in a set of group characteristics recorded in the Ethnographic Atlas, which capture important dimensions of development and institutional sophistication. We take the absolute difference in jurisdictional hierarchy, which ranges from 0 (none/autonomous bands or villages) to 4 (complex state), and in settlement complexity, which ranges from 1 (fully migratory or nomadic bands) to 8 (complex settlements). We also control for the absolute difference in the presence of High Gods. Belief in High Gods has been shown to contribute to cooperation (Norenzayan, 2013) and is considered an important predictor of conflict prevalence (McGuirk and Nunn, 2020). These differences have little independent effect on the likelihood of bilateral conflict, though groups more similar in their jurisdictional hierarchy are more likely to engage in conflict. In column 4, we account for differences in agricultural suitability, ruggedness and elevation. Column 5 controls for differences in group size between the government and the rebels; the estimated effect of this difference suggests that conflict is more likely between groups of unequal size. Finally, in column 6 we control for a full set of country by year fixed effects. Inclusion of all controls and fixed effects only increases the magnitude and significance of differences in the prevalence of revenge motifs.

The estimated effects are large. According to the richest specification with country-year fixed effects in column 6, a pair at the 75^{th} percentile of differences in revenge motifs, such as the Mbundu and the Bakongo in Angola (0.109), have a 40.7% higher probability of conflict each year than pairs with equal revenge values.

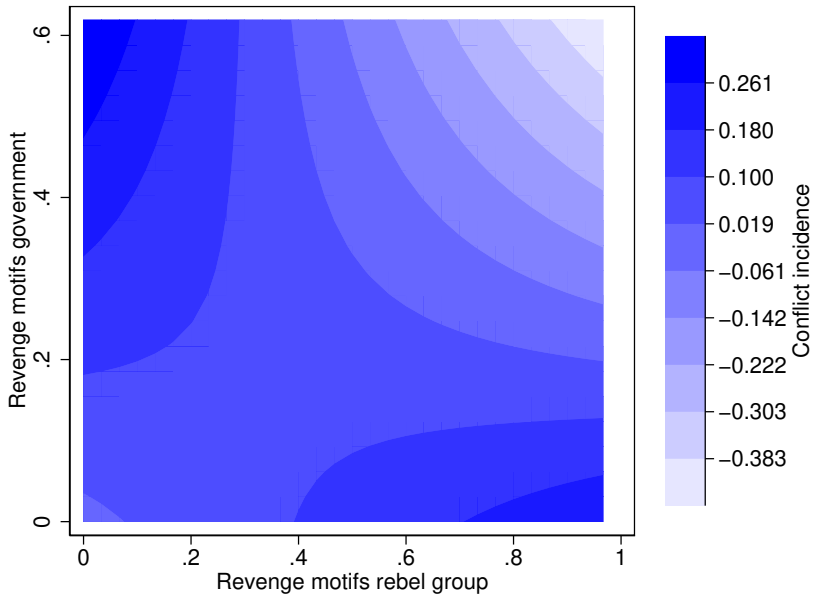
To provide a graphical illustration of our main result, we rely on an alternative empirical specification of the from

$$C_{dct} = \lambda_{c,t} + \delta_1 R_{g_1} + \delta_2 R_{g_2} + \gamma_1 R_{g_1} \times R_{g_2} + \mathbf{\Omega}_d \gamma_2 + \epsilon_{dct} \tag{16}$$

where R_{g_i} is the share of group i's folklore motifs related to revenge. Figure 6 is a contour plot of the predicted probability of conflict resulting from the specification in 16, estimated for different values of R_{g_1} and R_{g_2}. Conflict becomes more likely as groups diverge in terms of the prevalence of revenge in their folklore, and the result is symmetric: conflict is equally likely when the government is more revenge-prone than

the rebels and vice versa. Interestingly, the lowest likelihood of conflict is observed between groups with a high proportion of revenge motifs, in line with the theoretical result that the cooperative equilibrium across groups is most stable in the case of two pun-rep groups. Given that retaliation often spirals in societies of honor, the deterrent effect of punishment and escalation may be more powerful in interactions between such cultures. Table A.8 presents regression results from the specification underlying Figure 6, but accounting for all bilateral controls and fixed effects as in Table 6. As predicted by the model, the interaction term is robustly negative, indicating that conflict increases with divergence in the salience of revenge motifs within an ethnic group pair.

Figure 6: Difference in revenge prevalence and conflict incidence, contour plot



Notes: Contour plot of the predicted probability of conflict as a function of the share of revenge motifs in the government (y-axis) and rebel group (x-axis). Predicted values are from the regression in 16, omitting group fixed effects and accounting for the following bilateral controls: religious, linguistic and log geodesic distance and interactions between government and rebel group characteristics (jurisdictional hierarchy, settlement complexity, High Gods, agricultural suitability, ruggedness, elevation, group size).

As in the analysis of Section 5.2, we want to rule out that the effect of differences in the salience of revenge is driven by differences in cultural characteristics correlated with retaliation. We undertake this exercise in Table 7. Column 1 reports the estimate

in column 5 of Table 6 for comparability. In columns 2 and 3, we replace the distance in revenge motifs with distances in motifs of violence and deception, respectively. The former should capture differences in the overall belligerence of ethnic groups or in features of honor societies related to higher violence, but not retaliatory second-party punishment. This variable does not significantly affect conflict incidence. Motifs of violence that predict conflict-proneness at the group level are already captured by group fixed effects; differences in these motifs across groups do not have additional explanatory power.

Table 7: Falsification tests

Dep. variable	Conflict incidence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance revenge motifs	1.097***			0.861*	1.021**	354.4***	37.17***
	(0.336)			(0.494)	(0.400)	(105.4)	(12.55)
Distance violence motifs		0.176		-0.653			
		(0.195)		(0.548)			
Distance cheating motifs			0.636*	0.808			
			(0.356)	(0.776)			
Distance herding dependence					0.213		-127.1***
					(0.487)		(37.45)
Distance segmentary lineage						-65.89***	-8.112***
						(19.59)	(2.639)
Observations	2157	2157	2157	2157	2157	550	550
R-squared	0.633	0.631	0.632	0.633	0.633	0.647	0.647

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Distance revenge motifs*, *Distance violence motifs* and *Distance cheating motifs* measures the absolute difference between ethnic groups in the proportion of folklore motifs that are revenge-, violence- and cheating-related, respectively. *Distance herding dependence* is the absolute difference between ethnic groups in dependence on herding, constructed following Becker (2021) and Cao et al. (2021). *Distance segmentary lineage* is the absolute difference between ethnic groups in an indicator for the presence of segmentary lineages, from Moscona et al. (2019). All columns control for group (government and rebels) fixed effects, country by year fixed effects and the following bilateral controls: religious, linguistic and log geodesic distance and absolute differebces between government and rebel group characteristics (jurisdictional hierarchy, settlement complexity, presence of High Gods, agricultural suitability, ruggedness, elevation, group size). Standard errors are clustered at the group pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Differences in the prevalence of deception-related motifs instead do predict conflict incidence (column 3). Yet when jointly controlling for all three types of motifs in column 4, only revenge-related ones remain significant. This result highlights the role of a specific mechanism underlying the finding in existing work that cultures of honor are associated with violence. Conditional on the higher prevalence of violence in environments where honor cultures thrive – characterized by lower penetration of formal

institutions, tribal structures and lower average levels of development (Couttenier et al., 2017; Cassar et al., 2014) – honor cultures may clash more frequently in interactions with societies that do not have the same role for retaliation. Indeed, this is consistent with observations in the ethnographic literature that the use of violence within groups that employ retaliation practices is not as widespread as the focus on known cases of vendettas makes it to be (Gould, 2000). Referring to the pastoral group of the Nuer of Southern Sudan, Bates (1983) says "Insofar as the Nuer raided cattle, they tended to raid the cattle of others; raids within the tribe were rare." (cited in Rohner (2011)).

Consistent with the previous observation, column 5 shows that our results are not driven by differences in groups' main mode of subsistence. Cao et al. (2021) find that dependence on herding is correlated with the prevalence of revenge-related motifs in folklore, as well as with self-reported willingness to take revenge and a higher prevalence of conflict. Our specifications include group fixed effects and should thus account for any variation in group-specific conflict-proneness correlated with pastoralism. Following Becker (2021) and Cao et al. (2021), we construct a variable measuring each group's dependence on herding using the Ethnographic Atlas and compute the absolute difference in this variable between groups. The results in column 5 reveal no significant effect of differences in herding dependence on conflict incidence. Pastoral groups are indeed more likely to clash with groups less reliant on herding, but this effect appears to be driven specifically by the role of incompatible reputation mechanisms proxied by revenge-related motifs.[12]

In column 6, we control for another societal feature known to affect the frequency and intensity of conflict: segmentary lineage. Moscona et al. (2019) show that ethnic groups organized in segmentary lineages can quickly mobilize large numbers of individuals who trace their origins to the same common ancestor, leading to conflict escalation. Data on segmentary lineage societies has been compiled by Moscona et al. (2019) for a small number of groups in Africa. This severely limits the size of our dataset and results in this analysis should be interpreted given the limited remaining variation after accounting for all bilateral controls and fixed effects. Differences in segmentary lineage structure are not what drives the effect of differences in the prevalence of revenge. In

[12]The beta coefficient of the independent effect of herding dependence on conflict incidence is 1.335 (p-value < 0.000).

fact, the effect of the distance in the salience of revenge remains consistent with the model, even after simultaneously accounting for differences in herding dependence and segmentary lineage in column 7.

5.3.4 Robustness

Our main results remain significant when clustering standard errors at the country level, accounting for unobserved dependence of group-pairs within the same country (Table A.7). Given the small number of clusters (16), we report p-values from the wild bootstrap following Cameron et al. (2008). All estimates lose in precision, but the effect of differences in the prevalence of revenge motifs remain significant at the 5% level, while otherwise only linguistic and religious distance remain consistently and significantly predictive of conflict incidence.

Tables A.8 and A.9 show that both our main results and falsification tests remain robust to an alternative specification that follows equation 16.

Our main specifications include fixed effects for both government and out-of-power ethnic groups, ensuring that we control for all observable and unobservable group-specific characteristics, and identify effects of distances in revenge motifs from dyadic group pairings. This, however, limits the set of countries in our dataset that provide sufficient variation for identification. Reassuringly, our findings do not depend on this stringent specification. Columns 1-3 in Table A.10 report results from specifications that include fixed effects only for the out-of-power groups and, respectively, year, country and year, and country-by-year effects. Columns 5-7 verify robustness of these estimates to replacing absolute distances between groups with interactions in group characteristics, as in equation 16. Columns 4 and 8 report estimates from an even more stringent specification, which controls for group by country fixed effects for both ethnic groups in a pair. In these last regressions, we shut down any variation due to the presence of the same ethnic group in different countries. Results remain robust throughout.

6 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 perceptions of discrimination of immigrants in European countries, and of inter-ethnic conflict in Africa and Asia, are consistent with the model's predictions.

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, where institutionalized punishment crowds out incentives for individual retaliation. 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

Adida, Claire L, David D Laitin, and Marie-Anne Valfort, "Identifying Barriers to Muslim Integration in France," *Proceedings of the National Academy of Sciences*, 2010, *107* (52), 22384–22390.

_ , _ , and _ , "Muslims in France: Identifying a Discriminatory Equilibrium," *Journal of Population Economics*, 2014, *27* (4), 1039–1086.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, 2013, *128* (2), 469–530.

Allport, Gordon, *The Nature of Prejudice*, Reading, MA: Addison-Wesley, 1954.

Aronowitz, A, *Assimilation, Acculturation, and Juvenile Delinquency among Second Generation Turkish Youths in Berlin*, Dartmouth: Dartmouth Publishing Company,

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

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

Becker, Anke, "On the Economic Origins of Restricting Women's Promiscuity," *The Review of Economic Studies*, 2021.

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

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

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

Birnir, Jóhanna K, David D Laitin, Jonathan Wilkenfeld, David M Waguespack, Agatha S Hultquist, and Ted R Gurr, "Introducing the AMAR (All Minorities at Risk) Data," *Journal of Conflict Resolution*, 2018, *62* (1), 203–226.

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

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

Borjas, George J, "The Labor Demand Curve is downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *The Quarterly Journal of Economics*, 2003, *118* (4), 1335–1374.

Bormann, Nils-Christian, Lars-Erik Cederman, and Manuel Vogt, "Language, Religion, and Ethnic Civil War," *Journal of Conflict Resolution*, 2017, *61* (4), 744–771.

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

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

Brooks, Benjamin A, Karla Hoff, and Priyanka Pandey, "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*, 2018, *115* (45), 11385–11392.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.

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

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

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

Cao, Yiming, Benjamin Enke, Armin Falk, Paola Giuliano, and Nathan Nunn, "Herding, Warfare, and a Culture of Honor: Global Evidence," NBER Working Paper No. w29250 2021.

Caselli, Francesco and Wilbur John Coleman, "On the Theory of Ethnic Conflict," *Journal of the European Economic Association*, 2013, *11* (S1), 161–192.

Cassar, Alessandra, Giovanna d'Adda, and Pauline Grosjean, "Institutional Quality, Culture, and Norms of Cooperation: Evidence from Behavioral Field Experiments," *The Journal of Law and Economics*, 2014, *57* (3), 821–863.

Couttenier, Mathieu, Pauline Grosjean, and Marc Sangnier, "The Wild West is Wild: The Homicide Resource Curse," *Journal of the European Economic Association*, 2017, *15* (3), 558–585.

_ , Veronica Petrencu, Dominic Rohner, and Mathias Thoenig, "The Violent Legacy of Conflict: Evidence on Asylum Seekers, Crime, and Public Policy in Switzerland," *American Economic Review*, 2019, *109* (12), 4378–4425.

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

Eberle, Ulrich, Dominic Rohner, and Matthias Thoenig, "Heat and Hate: Climate Security and Farmer-Herder Conflicts in Africa," CEPR Working Paper DP15542 2020.

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

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

Esteban, Joan and Debraj Ray, "Linking Conflict to Inequality and Polarization," *American Economic Review*, 2011, *101* (4), 1345–74.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences," *The Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

Fearon, James D, "Why Ethnic Politics and 'Pork' Tend to Go Together," in "an SSRC-MacArthur sponsored conference on "Ethnic Politics and Democratic Stability," University of Chicago" Citeseer 1999.

Fearon, James D., "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 2003, *8* (2), 195–222.

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

_ and _ , "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 2003, *97* (1), 75–90.

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

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

Fernández, Raquel, "Does Culture Matter?," in Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, Elsevier, 2011, pp. 481–510.

Figueredo, Aurelio José, Ilanit Robin Tal, Prentiss McNeil, and Alfonso Guillén, "Farmers, Herders, and Fishers: The Ecology of Revenge," *Evolution and Human Behavior*, 2004, *25* (5), 336–353.

Fischer, Günther, HT Van Velthuizen, MM Shah, and Freddy O Nachtergaele, *Global Agro-Ecological Assessment for Agriculture in the 21st Century: Methodology and Results*, Rome: Food and Agriculture Organization of the United Nations, 2002.

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

_ and Nathan Nunn, "Ancestral Characteristics of Modern Populations," *Economic History of Developing Regions*, 2018, *33* (1), 1–17.

Gordon, Raymond G and Barbara F Grimes, "Ethnologue: Languages of the World. 2005," *Dallas, Texas: SIL International, Fifteenth*, 2009.

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

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

Grosjean, Pauline, "A History of Violence: The Culture of Honor and Homicide

in the US South," *Journal of the European Economic Association*, 2014, *12* (5), 1285–1316.

Guariso, Andrea and Thorsten Rogall, "Rainfall Inequality, Political Power, and Ethnic Conflict in Africa," LICOS Discussion Paper Series No. 391 2017.

Guarnieri, Eleonora, "Cultural Distance and Ethnic Civil Conflict," Working Paper 2022.

_ **and Ana Tur-Prats**, "Cultural Distance and Conflict-Related Sexual Violence," 2020.

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

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

Hamamura, Takeshi, Zita Meijer, Steven J Heine, Kengo Kamaya, and Izumi Hori, "Approach–Avoidance Motivation and Information Processing: A Cross-Cultural Analysis," *Personality and Social Psychology Bulletin*, 2009, *35* (4), 454–462.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich et al., "Costly Punishment across Human Societies," *Science*, 2006, *312* (5781), 1767–1770.

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

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

Huber, John and Laura Mayoral, "Inequality, Ethnicity and Civil Conflict," Working Paper, Columbia University 2014.

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

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

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

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

Lajevardi, Nazita, Kassra AR Oskooii, Hannah L Walker, and Aubrey L Westfall, "The Paradox between Integration and Perceived Discrimination among American Muslims," *Political Psychology*, 2020, *41* (3), 587–606.

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

Little, Peter D, *The Elusive Granary: Herder, Farmer, and State in Northern Kenya*,

Cambridge University Press, 1992.

McGuirk, Eoin F. and Nathan Nunn, "Transhumant Pastoralism, Climate Change and Conflict in Africa," NBER Working Paper No. 28243 2020.

Michalopoulos, Stelios and Elias Papaioannou, "The Long-Run Effects of the Scramble for Africa," *American Economic Review*, 2016, *106* (7), 1802–48.

_ and Melanie Meng Xue, "Folklore," *The Quarterly Journal of Economics*, 2021, *136* (4), 1993–2046.

Miguel, Edward, Sebastián M Saiegh, and Shanker Satyanath, "Civil War Exposure and Violence," *Economics & Politics*, 2011, *23* (1), 59–73.

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

Mitra, Anirban and Debraj Ray, "Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India," *Journal of Political Economy*, 2014, *122* (4), 719–765.

Morelli, Massimo and Dominic Rohner, "Resource Concentration and Civil Wars," *Journal of Development Economics*, 2015, *117*, 32–47.

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

_ , _ , _ et al., "Kinship and Conflict: Evidence from Segmentary Lineage Societies in Sub-Saharan Africa," Unpublished Manuscript, Harvard University 2019.

Murdoch, George Peter, *Ethnographic Atlas*, Pittsburgh: University of Puttsburgh Press, 1967.

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

Norenzayan, Ara, *Big Gods: How Religion Transformed Cooperation and Conflict*, Princeton: Princeton University Press, 2013.

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

Nowak, Martin A. and Karl Sigmund, "Evolution of Indirect Reciprocity by Image Scoring," *Nature*, 6 1998, *393* (6685), 573–577.

Nunn, Nathan and Diego Puga, "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

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

Rohner, Dominic, "Reputation, Group Structure and Social Tensions," *Journal of Development Economics*, 2011, *96* (2), 188–199.

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

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

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

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

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

Spolaore, Enrico and Romain Wacziarg, "War and Relatedness," *Review of Economics and Statistics*, 2016, *98* (5), 925–939.

_ and _ , "The Political Economy of Heterogeneity and Conflict," Technical Report, NBER Working Paper No. 23278 2017.

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

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

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

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

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

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

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

_ , Augustine A Ayantunde, ED Patterson, and KP Patterson, *Farmer-Herder Relations and Conflict Management in Agro-Pastoral Zone of Niger* 2006.

Vega, William A, Andres G Gil, George J Warheit, Rick S Zimmerman, and Eleni Apospori, "Acculturation and Delinquent Behavior among Cuban American Adolescents: Toward an Empirical Model," *American Journal of Community Psychology*, 1993, *21* (1), 113–125.

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker, and Luc Girardin, "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family," *Journal of Conflict Resolution*, 2015, *59* (7), 1327–1342.

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

Wucherpfennig, Julian, Nils B Weidmann, Luc Girardin, Lars-Erik Cederman, and Andreas Wimmer, "Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset," *Conflict Management and Peace Sci-*

ence, 2011, *28* (5), 423–437.

Appendix

A Additional Figures and Tables

Table A.1: Summary statistics, ESS

Variable	Mean	Std	Min	Max	N
Feels discriminated	0.12	0.32	0	1	24,221
Female	0.54	0.50	0	1	24,536
Age	43.89	17.86	13	101	24,442
Contiguous	0.50	0.50	0	1	24,509
Common official language	0.23	0.42	0	1	24,509
Former colonial relation	0.33	0.47	0	1	24,509
Distance between capitals	1.81	2.39	0.06	19.15	24,509
Difference log GDP p.c.	0.76	0.72	0	4.39	24,313
Religious similarity	4.26	2.40	0	9.47	24,536
Distance revenge motifs	0.03	0.03	0	0.29	24,319
Distance violence motifs	0.02	0.02	0	0.17	24,319
Distance cheating motifs	0.04	0.05	0	0.26	24,319
Distance revenge (eHRAF)	0.12	0.29	0	1	9,769
Distance patience	0.42	0.32	0.01	1.56	16,220
Distance risk-taking	0.18	0.18	0	1.32	16,220
Distance positive reciprocity	0.28	0.18	0	1.35	16,220
Distance negative reciprocity	0.28	0.25	0	1.03	16,220
Distance altruism	0.33	0.27	0	1.36	16,220
Distance trust	0.21	0.16	0	1.01	16,220

Table A.2: Pairwise correlations, measures of reputation

	Distance revenge motifs	Distance revenge (eHRAF)	Distance negative reciprocity
Distance revenge motifs	1		
Distance revenge (eHRAF)	0.2011	1	
	(0.000)		
Distance negative reciprocity	-0.0124	0.1204	1
	(0.7220)	(0.0203)	

Notes: P-values in parentheses.

Table A.3: Summary statistics, civil conflict

Variable	Mean	Std	Min	Max	N
Conflict incidence	0.09	0.28	0	1	3,969
Distance revenge motifs	0.10	0.08	0	0.41	3,969
Distance cheating motifs	0.11	0.09	0	0.50	3,969
Distance violence motifs	0.06	0.06	0	0.30	3,969
Linguistic distance	0.44	0.29	0	1	3,915
Religious distance	0.34	0.31	0	1	3,969
Geographic distance	0.92	0.72	0	4.02	3,969
Distance jurisdictional hierarchy	1.21	1.06	0	4	3,630
Distance settlement complexity	1.23	1.62	0	6	3,713
Distance High Gods	0.19	0.36	0	1	2,986
Distance herding dependence	10.10	11.90	0	55.75	3,969
Distance segmentary lineage	0.54	0.44	0	1	959
Distance agricultural suitability	0.12	0.12	0	0.44	3,794
Distance ruggedness	509.94	621.96	0	3,234.46	3,794
Distance elevation	519.85	585.06	0	2,791.41	3,795
Distance group size	0.39	0.31	0	0.98	3,969

Table A.4: Differences in revenge prevalence and perceptions of discrimination, two-way clustered standard errors

Dep. variable	Feels discriminated			
	(1)	(2)	(3)	(4)
Distance revenge motifs	0.104**	0.0913*	0.0951**	0.0653**
	(0.0473)	(0.0463)	(0.0464)	(0.0294)
Contiguous		-0.0364***	-0.0393***	-0.0248**
		(0.0120)	(0.00982)	(0.0103)
Common official language		0.0573**	0.0592**	0.0355*
		(0.0216)	(0.0217)	(0.0187)
Former colonial relation		0.00414	0.00543	-0.00418
		(0.0259)	(0.0256)	(0.0241)
Distance between capitals		0.120	0.135*	0.0551
		(0.0718)	(0.0675)	(0.0459)
Difference log GDP p.c.			-0.00565	-0.00901
			(0.0299)	(0.0282)
Religious similarity				-0.153***
				(0.0357)
Observations	23896	23896	23693	23693
R-squared	0.0649	0.0666	0.0664	0.0692

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Distance revenge motifs* measures the absolute difference between the respondent's residence country and their parents' country of birth in the proportion of folklore motifs that are revenge-related. All columns control for gender, age and age squared, host and origin country fixed effects and survey year fixed effects. Standardized beta coefficients reported. Standard errors are clustered two-way at the host and origin country level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.5: Differences in revenge prevalence and perceptions of discrimination, interaction of host and origin characteristics

Dep. variable	Feel discriminated										
	OLS				2SLS		OLS				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Revenge motifs	-0.0647***	-0.0345*							-0.0480*		0.0088
	(0.0234)	(0.0179)							(0.0256)		(0.0222)
Negative reciprocity			-0.0474***	-0.0366***	-0.451**	-0.331			-0.0274*	-0.0390	-0.0380
			(0.0173)	(0.0138)	(0.225)	(0.207)			(0.0146)	(0.0294)	(0.0294)
Revenge (eHRAF)							-0227	-0.0121**		-0.0239***	-0.0239***
							(0559)	(0605)		(0710)	(0711)
Patience									-0.318**	-0.0410	-0.0474
									(0.160)	(0.127)	(0.125)
Risk-taking									0.443	0.303	0.309
									(0.395)	(0.434)	(0.436)
Positive reciprocity									0.333	-1.041	-1.035
									(0.440)	(0.724)	(0.728)
Altruism									0.247	0.531	0.530
									(0.222)	(0.375)	(0.373)
Trust									0206	-0.108	-0.136
									(0.294)	(0.409)	(0.415)
Violence motifs									0914	-0398	-0594
									(0.0231)	(0.0282)	(0.0294)
Cheating motifs									0.0128	0.0320	0.0348*
									(0.0164)	(0.0198)	(0.0206)
Observations	23896	23693	15971	15951	15869	15849	9572	9446	15849	6026	6026
R-squared	0.0644	0.0689	0.0563	0.0623			0.0647	0.0677	0.0643	0.0587	0.0588
F-statistic					5.42	5.07					

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Revenge motifs*, *Violence motifs* and *Cheating motifs* are interactions between origin and residence country proportions of folklore motifs that are revenge-, violence- and cheating-related, respectively. *Revenge (eHRAF)* is the interaction between host and origin country values of the eHRAF revenge indicator (averaged across all ethnic groups in a country). *Patience*, *Risk-taking*, *Positive reciprocity*, *Negative reciprocity*, *Altruism* and *Trust* are interactions between host and origin economic preferences from GPS (Falk et al., 2018). All regressions include individual controls for gender, age and age squared, origin and host country fixed effects, and survey year fixed effects, Columns 2, 4, 6, 8-11 also include interactions of bilateral economic, geo-cultural and religious controls from column 4 of Table 2. Standardized beta coefficients reported. Standard errors are clustered at the country pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.6: Sources of variation, civil conflict data

	Nr. obs.	Nr. dyads	First year	Last year
Afghanistan	85	3	1946	2020
Angola	184	4	1975	2020
Benin	19	4	1964	2006
Burundi	31	1	1966	2001
Central African Republic	27	3	1960	2020
China	339	5	1950	2020
Congo	5	1	1964	1968
Congo, Democratic Republic of (Zaire)	559	43	1960	2020
Cote D'Ivoire	64	6	1960	2020
Ethiopia	183	5	1946	2019
Gabon	13	2	1960	2005
Georgia	30	1	1991	2020
Iran (Persia)	150	2	1946	2020
Iraq	17	1	2004	2020
Japan	136	3	1953	2020
Kenya	95	7	1963	2020
Liberia	11	2	1981	2020
Mali	32	1	1960	1991
Niger	163	6	1960	2020
Nigeria	207	10	1960	2020
Rwanda	59	2	1962	2020
Sierra Leone	10	2	1965	2015
Togo	46	2	1960	2020
Uganda	171	24	1967	2020
Zimbabwe (Rhodesia)	21	3	1982	2008

Table A.7: Difference in revenge prevalence and conflict incidence, alternative clustering

Dep. variable	Conflict incidence					
	(1)	(2)	(3)	(4)	(5)	(6)
Distance revenge motifs	0.249	0.329*	0.799*	1.117**	1.121**	1.097**
	(0.122)	(0.093)	(0.082)	(0.015)	(0.014)	(0.010)
Linguistic distance		-0.268	-0.298**	-0.241*	-0.238*	-0.276*
		(0.451)	(0.022)	(0.094)	(0.071)	(0.068)
Religious distance		0.990	2.543	3.576**	3.170**	2.220**
		(0.155)	(0.484)	(0.016)	(0.023)	(0.024)
Log geographic distance		0.421	0.587*	0.564	0.560	0.513
		(0.443)	(0.071)	(0.173)	(0.156)	(0.178)
Distance jurisdictional hierarchy			-0.778	-1.140	-1.132	-1.160
			(0.133)	(0.162)	(0.167)	(0.157)
Distance settlement complexity			-0.0337	-0.159	-0.199	-0.304
			(0.720)	(0.405)	(0.338)	(0.363)
Distance High Gods			0.0866	0.658	0.684*	0.677
			(0.634)	(0.133)	(0.098)	(0.1201)
Distance agricultural suitability				-0.139	-0.153	-0.173*
				(0.116)	(0.108)	(0.0801)
Distance ruggedness				0.982	1.010	1.144
				(0.119)	(0.137)	(0.118)
Distance elevation				-0.345	-0.311	-0.279
				(0.352)	(0.387)	(0.340)
Distance group size					0.443*	1.210*
					(0.054)	(0.065)
Observations	3968	3915	2985	2811	2811	2157
R-squared	0.390	0.393	0.366	0.366	0.368	0.633
Country × Year FE						✓

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Distance revenge motifs* measures the absolute difference between the two ethnic groups in the proportion of folklore motifs that are revenge-related. All columns control for group (government and rebels) fixed effects, country and year fixed effects. Standard errors are clustered at the country level using the wild bootstrap (Cameron et al., 2008). P-values in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.8: Difference in revenge prevalence and conflict incidence, interaction of government and rebel group characteristics

Dep. variable	Conflict incidence					
	(1)	(2)	(3)	(4)	(5)	(6)
Revenge motifs	-0.181	-0.353***	-1.472**	-1.675***	-1.566***	-1.449***
	(0.129)	(0.134)	(0.644)	(0.392)	(0.432)	(0.540)
Linguistic distance		-0.281***	-0.242**	-0.248***	-0.307**	-0.442***
		(0.107)	(0.113)	(0.0753)	(0.119)	(0.140)
Religious distance		0.515	2.540**	3.402***	2.811**	1.712
		(0.521)	(1.215)	(0.881)	(1.255)	(1.296)
Log geographic distance		0.506***	0.716***	0.657***	0.645***	0.623***
		(0.120)	(0.206)	(0.147)	(0.150)	(0.185)
Settlement complexity			-0.0169	-0.161	-0.169	-0.0327
			(0.317)	(0.276)	(0.275)	(0.314)
Jurisdictional hierarchy			-0.0722	0.273	0.309	0.461*
			(0.271)	(0.201)	(0.209)	(0.235)
High Gods			-0.0277	0.149	0.130	0.0706
			(0.211)	(0.282)	(0.285)	(0.341)
Agricultural suitability				-0.100	-0.0622	0643
				(0.104)	(0.118)	(0.121)
Ruggedness				-3.305**	-3.299**	-3.706*
				(1.469)	(1.463)	(2.051)
Elevation				2.424	2.426	2.673
				(1.496)	(1.490)	(2.101)
Group size					-0.489	-1.188***
					(0.708)	(0.394)
Observations	3968	3915	2985	2811	2811	2157
R-squared	0.390	0.393	0.365	0.367	0.368	0.634
Country × Year FE						✓

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Revenge motifs* is the interaction between government and rebel group proportions of folklore motifs that are revenge-related. All other reported estimates are interactions between government and rebel group characteristics. All columns control for group (government and rebels) fixed effects, country and year fixed effects. Standard errors are clustered at the group pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.9: Falsification tests, interaction of government and rebel group characteristics

Dep. variable	Conflict incidence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Revenge motifs	-1.449***			-1.423**	-1.705***	-523.1***	-133.5***
	(0.540)			(0.668)	(0.636)	(175.5)	(40.31)
Violence motifs		-0.316		-0.0238			
		(0.285)		(0.208)			
Cheating motifs			2.880	1.968			
			(13.66)	(10.15)			
Dependence on herding					0.801		-9.668***
					(0.846)		(3.361)
Segmentary lineage						41.76***	10.80***
						(14.10)	(3.353)
Observations	2157	2157	2157	2157	2157	550	550
R-squared	0.634	0.633	0.633	0.634	0.634	0.647	0.647

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Revenge motifs*, *Violence motifs* and *Cheating motifs* are interactions between government and rebel group proportions of folklore motifs that are revenge-, violence- and cheating-related, respectively. *Dependence on herding* is the interaction between government and rebel group in dependence on herding, constructed following Becker (2021) and Cao et al. (2021). *Segmentary lineage* is the interaction between a government and a rebel group indicator for the presence of segmentary lineages, from Moscona et al. (2019). All columns control for group (government and rebels) fixed effects, country by year fixed effects and interactions of bilateral controls from column 6 of Table 6. Standard errors are clustered at the group pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.10: Differences in revenge prevalence and conflict incidence, alternative specifications of fixed effects

Dep. variable	Conflict incidence							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance revenge motifs	0.457***	0.498***	0.764***	0.768***				
	(0.123)	(0.128)	(0.246)	(0.193)				
Revenge motifs (interaction)					-0.441***	-0.320*	-1.924***	-1.568***
					(0.134)	(0.172)	(0.589)	(0.440)
Observations	2811	2811	2157	2811	2811	2811	2157	2811
R-squared	0.361	0.365	0.636	0.370	0.368	0.369	0.637	0.373
Powerless group FE	✓	✓	✓		✓	✓	✓	
Year FE	✓	✓			✓	✓		
Country FE		✓				✓		
Country × Year FE			✓	✓			✓	✓
Groups × Country FE				✓				✓

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Distance revenge motifs* measures the absolute difference between the two ethnic groups in the proportion of folklore motifs that are revenge-related. Specifications in columns 1-4 include all bilateral distance controls from column 6 of Table 6. Specifications in columns 5-8 include the same controls as interactions between characteristics of the government and rebel group. Standard errors are clustered at the country pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

B Data construction

B.1 Measures of punishment-based reputation

B.1.1 Folklore

Our main measure of punishment-based reputation is the frequency of revenge-related motifs in a society's oral tradition. We rely on the catalog of folklore organized and disseminated by Michalopoulos and Xue (2021). The primary material for this catalog is the work of anthropologist Yuri Berezkin, who compiled a collection of 2,564 motifs in the traditions of 958 groups. A motif "reflects a combination of images, episodes, or structural elements found in two or more texts" (Michalopoulos and Xue, 2021). Michalopoulos and Xue (2021) tagged each motif in the catalog as relating to a concept, using the top 10,000 concepts from ConceptNet, a free online semantic network designed to create word embeddings. A motif is tagged as relevant to a concept if it contains a mention of the concept or of one of the top-50 words returned by ConceptNet when the concept is used as a seed term. A motif may be assigned to different concepts. Michalopoulos and Xue (2021) provide a dataset with the frequency (number) of motifs related to different concepts aggregated at the level of the country and at the level of the ethnic group, for 1,245 groups in the *Ethnographic Atlas* (EA). We rely on these frequencies for the construction of our main independent variable.

We proceed as follows: first, we identify a set of concepts related to revenge and retaliation in the folklore dataset of Michalopoulos and Xue (2021). For consistency with Cao et al. (2021), who also study cultures of honor, we select the following terms: *revenge*, *punish*, *punishment*, *penalty*, *retaliate*, *retaliation*. We then compute the total number of motifs tagged by each of the selected keywords and divide that by the total number of motifs catalogued for each group, to adjust for the fact that some societies have richer oral traditions than others. Our final measure is the relative frequency of revenge-related motifs in the folklore of a given society (country or ethnic group).

We follow the same approach to compute the relative frequency of motifs related to violence and deception. For violence, we select the same terms as Cao et al. (2021): *violence*, *violent*, *battle*, *fighting*, *attach*, *soldier*, *guard*, *troop*, *army*, *enemy*, *fighter*, *invasion*, *invade*, *defender*. For deception, we follow Michalopoulos and Xue (2021) and select the terms *cheat*, *deceive* and *trick*.

B.1.2 Human Relations Area Files (HRAF)

For the analysis at the country-pair level, we use a complementary measure of punishment-based reputation that draws from the scholarly ethnographic record. We rely on *eHRAF World Cultures*, a database of ethnographic collections covering cultures across the world, produced and maintained by *Human Relations Area Files* (HRAF), a non-profit research agency based at Yale University. The online database provides a summary of the ethnographic record on each culture, organized in sections covering different types of basic information (e.g. location, language), domains of social organization (e.g. marriage and family, inheritance) or activity (e.g. subsistence, trade). We crawl through the list of summaries for any mentions of revenge and retaliation in the two sections most closely related to our theoretical framework: *conflict* and *social control*. The former covers relations to external groups, with a focus on the frequency of war and military activity; the latter relates to inter-community relations, including within-group disputes and sanctions for social deviance. We manually inspect a number of summaries and select the following list of terms as most appropriate to capture the importance of second-party punishment: *veng*, *feud*, *vendetta*, *retali*, *remedies*, *retrib*, *blood*, *repris*, *honour*, *kin*, *lineage*. We tag as pun-rep all cultures containing at least one of these terms in at least one of the two relevant sections of their culture's summary. We are able to construct this indicator for the 304 societies in eHRAF with recorded information on conflict and social control.

The following examples illustrate what is captured by this method. Societies tagged as pun-rep include some of the most well-known groups that use blood revenge, such as the Pashtun of southeastern Afghanistan and western Pakistan and the Nuer of South Sudan. From the description of social control in the Pashtun society:

"Traditionally social control was maintained by a code of behavior and honor called PAKHTUNWALI. It combines the principles of revenge, hospitality to guests, defense of those who have sought protection in one's care, the chastity of married women, and restraint toward those considered weak or helpless (Hindus, women, and boys). PAKHTUNWALI in some cases contradicts and generally takes precedence over Islamic law. It is harsh – the penalty for illicit sexual behavior, for example, is death – and it is enforced by strong social pressure. Violations of law outside of the activities the code encompasses are dealt with by the JIRGA or the government administration."

And for the Nuer:

"Within the Nuer, homicide is common and is usually related to cattle. Nuer say

that more people have died for the sake of a cow than for any other cause. Acts of homicide can be immediately avenged or held as blood-feuds until such time when the two sides finally square even, and the mechanism to deter homicide and revenge has been the imposition of blood wealth, which is payable in cattle. The norm has been 30 cows paid to the family of the slain person. It can therefore, be said that because cattle are a source of turmoil, a threat of one's cattle being taken away in punishment induces prudence in the relations between people."

Cambodians instead are not tagged as a pun-rep culture based on our selected keywords. The description of their society fits the coop-rep equilibrium in our model:

"At the community level, social control is maintained through socialization from childhood into norms of proper conduct and through use of informal sanctions such as gossip or ostracism. Individuals seek to avoid the embarrassment or shame of improper behavior, as well as to earn religious merit by following the major Buddhist rules of conduct (do not lie, steal, drink alcoholic beverages, fornicate, or kill living creatures). Certain kinds of misbehavior are thought to bring punishment from supernatural beings, usually in the form of illness. Although police and law courts exist, many people avoid using them except when absolutely necessary."

For the analysis of conflict, we match cultures in eHRAF to groups in the Ethnographic Atlas using the correspondence table provided by eHRAF.[13] For the analysis of ESS data, we aggregate the information on ethnic groups to the country level, following a procedure similar to Alesina et al. (2013) and Giuliano and Nunn (2018). First, we use the match created by Giuliano and Nunn (2018) between the Ethnographic Atlas and the *Ethnologue: Languages of the World* (Gordon and Grimes, 2009) dataset, to assign the revenge indicator to modern-day language groups. Giuliano and Nunn (2018) provide a shapefile that associates each of the approximately 7,500 language groups in the Ethnologue to a polygon; polygons are non-overlapping and mutually exclusive. Next, we intersect this shapefile with a shapefile of country borders and overlay the resulting polygons of ethnic groups in a specific country with a raster file of the world's population taken from *LandScan Global Population Database* (LandScan 2016). The LandScan raster reports population estimates for each 30 arc-second by 30 arc-second cell. By summing up all cells within each group-country-specific polygon,

[13]The table can be found at https://hraf.yale.edu/resources/reference/ea-cases-in-ehraf/.

we compute the total population of an ethnic group in a country. Finally, we aggregate the eHRAF revenge indicator at the country level, by taking the average of revenge indicators across all ethnic-groups with non-missing values, using the groups' shares of the country's total population as weights.

B.2 Construction of conflict dataset

Data on conflict between the government and groups not in power comes from the *Ethnic Power Relations Dataset Family* (EPR) (Vogt et al., 2015). The EPR tracks politically relevant groups for all countries in the world and for every year between 1946 and 2021. Groups are defined as politically relevant if they hold government power, or are officially excluded from power or discriminated. The status of groups in government is listed as "dominant", "monopoly," "senior partner" or "junior partner." Groups outside the government are classified as "discriminated", "powerless' or "self-excluded." In cases where a senior partner and one or several junior partners jointly govern, we use the senior partner to classify the government.[14] For each country and year, we then have a set of pairs between the government and each other politically relevant group.

Data on conflict comes from the *Uppsala Conflict Data Program* (UCDP). The UCDP is a dataset of geo-referenced conflict incidents with at least 25 fatalities between 1989 and 2016, with information on involved actors. Actors in the UCDP are matched to groups in the EPR through the ACD2EPR dataset. Our main dependent variable is a conflict indicator that takes on the value one if the government is in conflict with another group in a given year, and zero otherwise.

Ethnic groups in the EPR are linked to variables from the folklore dataset and other ethnographic controls through the Ethnographic Atlas. Michalopoulos and Papaioannou (2016) provide a concordance table between EPR and the Ethnographic Atlas for Africa. We extend this match to Asia by combining spatial and language information on groups from various sources. Information on groups' location comes from the GeoEPR dataset (Wucherpfennig et al., 2011), which assigns groups to spa-

[14] A procedure like the one followed in Guarnieri (2022), that aggregates characteristics of groups in multi-group governments and weights senior partners more than junior partners yields very similar results. Available upon request.

tial polygons; information on language comes from from the EPR Ethnic Dimensions (EPR-ED) Dataset (Bormann et al., 2017). Using the coordinates on groups' ethnic homelands and information on language provided by the Ethnographic Atlas, we identify groups in the Ethnographic Atlas that inhabit the polygons of groups in the EPR and speak the same language. We consider those groups as matched.[15] For remaining groups, we perform a manual match based on online searches.

B.3 Variable definitions and sources

Folklore motifs

Distance revenge motifs. Absolute difference between ethnic groups or countries in the number of revenge-related motifs as a proportion of total motifs, from Michalopoulos and Xue (2021). Revenge-related motifs are those tagged by the following keywords: *revenge*, *punish*, *punishment*, *penalty*, *retaliate*, *retaliation*.

Distance violence motifs. Absolute difference between ethnic groups or countries in the number of violence-related motifs as a proportion of total motifs, from Michalopoulos and Xue (2021). Violence-related motifs are those tagged by the following keywords: *violence*, *violent*, *battle*, *fighting*, *attach*, *soldier*, *guard*, *troop*, *army*, *enemy*, *fighter*, *invasion*, *invade*, *defender*.

Distance cheating motifs. Absolute difference between ethnic groups or countries in the number of deception-related motifs as a proportion of total motifs, from Michalopoulos and Xue (2021). Deception-related motifs are those tagged by the following keywords: *cheat*, *deceive*, *trick*.

[15] In this step, we are also aided in finding a correspondence between the Ethnographic Atlas and the EPR dataset by the A-MAR project (Birnir et al., 2018). The Ethnographic Atlas contains groups that are often aggregated into "upper level" groups in the EPR. The A-MAR project makes distinctions between "upper" and "lower" level groups, allowing us to assign Ethnographic Atlas groups to more aggregated groups that can then be matched to the EPR. For example, according to A-MAR the Semang are included among the Orang Asli, which appear in the EPR.

Country-level variables

Contiguous. Indicator for contiguous host-origin country pairs, from Guiso et al. (2009).

Common official language. Indicator for common official or primary language between host-origin country pairs, from Guiso et al. (2009).

Former colonial relation. Indicator for host-origin country pairs that have ever been in a colonial relationship, from Guiso et al. (2009).

Distance between capitals. Distance (as the crow flies) between host-origin country capitals in km, from Guiso et al. (2009).

Difference log GDP per capita. Absolute difference in log GDP per capita in 2000. GDP data is from the World Bank.

Religious similarity. Measure from Guiso et al. (2009). 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.

Distance patience. Absolute difference between host and origin country in patience from the GPS (Falk et al., 2018). The country-level measure of patience is a weighted average of a self-assessment of the willingness to wait and an intertemporal choice sequence using staircase method.

Distance risk-taking. Absolute difference between host and origin country in risk-taking preference from the GPS (Falk et al., 2018). The country-level measure of risk-taking is a weighted average of a self-assessment of the willingness to take risks in general and a lottery choice sequence using staircase method.

Distance positive reciprocity. Absolute difference between host and origin country in positive reciprocity from the GPS (Falk et al., 2018). The country-level measure of positive reciprocity is a weighted average of a self-assessment of the willigness to return a favor and agreement with a statement on gift-giving in exchange for help.

Distance negative reciprocity. Absolute difference between host and origin country in negative reciprocity from the GPS (Falk et al., 2018). The country-level measure of negative reciprocity is a weighted average of three self-assessment questions on the willingness to take revenge, the willingness to punish unfair behavior toward one's self and the willingness to punish unfair behavior toward others.

Distance altruism. Absolute difference between host and origin country in altruism from the GPS (Falk et al., 2018). The country-level measure of altruism is a weighted average between a donation decision and a self-assessment of the willingness to give to good causes.

Distance trust. Absolute difference between host and origin country in trust from the GPS (Falk et al., 2018). The country-level measure of trust is based on agreement with the statement that people have only the best intentions.

Conflict

Conflict incidence. Pair-level indicator for conflict in a given year between the government and another politically relevant group. Based on data from the the ACD2EPR dataset (Vogt et al., 2015).

Ethnic group-level variables

Linguistic distance. We compute cladistic linguistic distance following Fearon (2003). First, we merge the language of each ethnic group to the Ethnologue. The Ethnologue includes information on each language's position in a linguistic tree that starts with a broad language family and branches out to intermediate nodes before ending with the language itself. Distance between a pair of languages i and j is computed as

$$d_{i,j} = 1 - \left(\frac{\#\text{of common nodes between } i \text{ and } j}{\frac{1}{2}(\text{of nodes of language } i + \text{of nodes of language } j)} \right)^{\lambda}$$

where λ is a concavity factor used to attribute higher weight to earlier common nodes, as separation in earlier nodes implies higher linguistic divergence. We set $\lambda = 0.5$

following existing literature (Fearon and Laitin, 2003; Guarnieri and Tur-Prats, 2020; Guarnieri, 2022).

Religious distance. We implement the same formula as above to compute religious distance following Guarnieri and Tur-Prats (2020). We use religions reported for each ethnic group in the EPR-ED dataset (Bormann et al., 2017).

Geographic distance. Simple geodesic distance between centroids of group polygons, from GeoEPR (Wucherpfennig et al., 2011).

Distance jurisdictional hierarchy. Absolute difference between ethnic groups in variable *v33* of the Ethnographic Atlas, capturing jurisdictional hierarchy beyond the local community. The variable takes on values between 1 (no hierarchy beyond local community) and 5 (four levels of hierarchy).

Distance settlement complexity. Absolute difference between ethnic groups in variable *v30* of the Ethnographic Atlas. The variable takes on values from 1 (nomadic or fully migratory) to 8 (complex settlements), with higher values indicating more permanent and complex settlements.

Distance High Gods. Absolute difference between ethnic groups in the presence of High Gods, based on *v34* of the Ethnographic Atlas. We code High Gods as present if the variable takes on values between 3 (Not active in human affairs) and 5 (Supportive of human morality). Results are similar when we code the presence of High Gods only when High Gods are moralizing ($v34 = 5$).

Distance herding dependence. Absolute distance between ethnic groups in dependence on herding. We code this variable for each ethnic group following Becker (2021) and Cao et al. (2021). We use the variable *v4* of the Ethnographic Atlas, which codes societies by levels of dependence on animal husbandry, ranging from 0-5% to 86-100%. We recode this to the median of each category and multiply with a dummy that equals one if the predominant animal is a herding animal (sheep, cattle, horses, reindeer, alpacas, or camels, based on *v40* of the Ethnographic Atlas).

Distance segmentary lineage. Absolute difference between ethnic groups in an indicator for segmentary lineages. This variable comes from Moscona et al. (2019) and is only available for 145 ethnic groups in Africa.

Distance agricultural suitability. Absolute distance between ethnic groups in average suitability for low-input rainfed agriculture in the ethnic groups' territories. Suitability data is from FAO (Fischer et al., 2002), averaged over polygons in the GeoEPR dataset (Wucherpfennig et al., 2011).

Distance ruggedness. Absolute distance between ethnic groups in average ruggedness in the ethnic groups' territories. We compute terrain ruggedness following Nunn and Puga (2012).Underlying elevation data are from *GTOPO30*, a global digital elevation model (DEM) of 30 arc-second by 30 arc-second cells. Ruggedness is averaged over polygons in the GeoEPR dataset (Wucherpfennig et al., 2011).

Distance elevation. Absolute distance between ethnic groups in average elevation in the ethnic groups' territories. Data on elevation is provided by EPR (Vogt et al., 2015).

Distance group size. Absolute distance in group size between ethnic groups. Information on group size is provided by EPR (Vogt et al., 2015).