Explaining Interethnic Cooperation
Author(s): James D. Fearon and David D. Laitin
Source: *The American Political Science Review*, Vol. 90, No. 4 (Dec., 1996), pp. 715-735
Published by: American Political Science Association
Stable URL: http://www.jstor.org/stable/2945838
Accessed: 11-10-2016 10:22 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

American Political Science Association, Cambridge University Press are collaborating with JSTOR to digitize, preserve and extend access to *The American Political Science Review*

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

Explaining Interethnic Cooperation

JAMES D. FEARON and DAVID D. LAITIN *University of Chicago*

Though both journalists and the academic literature on ethnic conflict give the opposite impression, peaceful and even cooperative relations between ethnic groups are far more common than is large-scale violence. We seek to explain this norm of interethnic peace and how it occasionally breaks down, arguing that formal and informal institutions usually work to contain or "cauterize" disputes between individual members of different groups. Using a social matching game model, we show that local-level interethnic cooperation can be supported in essentially two ways. In spiral equilibria, *disputes between individuals are correctly expected to spiral rapidly beyond the two parties, and fear of this induces cooperation "on the equilibrium path." In* in-group policing equilibria, *individuals ignore transgressions by members of the other group, correctly expecting that the culprits will be identified and sanctioned by their own ethnic brethren. A range of examples suggests that both equilibria occur empirically and have properties expected from the theoretical analysis.*

I grew up in a Jewish section of Flatbush that bordered an Italian neighborhood. Sometimes on our way to school, some Italian kids—nearly all of them went to parochial schools—would hassle and even attack us. Although they lived only a few blocks away, we didn't even know their names. We just called them "the St. Brennan's kids." Our parents would see our injuries and report the incidents to our school principal, who was Jewish, but from a different neighborhood. He contacted the relevant authorities at St. Brennan's, who would investigate the matter and punish the culprits. The funny thing was, no one ever seemed to think of calling the police. They were Irish.[1]

—David Laitin

A compelling theory of ethnic conflict needs, first, to explain why interethnic relations are frequently characterized by a tension and mistrust which are relatively absent in intraethnic relations and which seem related to spirals of violence that mark interethnic more than intraethnic affairs. Second, for interethnic relations, a good theory of ethnic conflict should be able to explain why, despite the greater tensions, peaceful and cooperative relations are by far the more typical outcome than is large-scale violence (a claim we substantiate below). Third, a good theory should be able to explain why in some cases interethnic tensions are occasionally punctuated by spirals of violence, while in other cases the tensions exist, but interethnic disputes are more often "cauterized" short of war.

Among existing theories of ethnic conflict, accounts focusing on past tensions between groups that are memorialized in narratives of blame and threat tend to dramatic overprediction of violence.[2] Such narratives are almost always present, but large-scale interethnic violence is extremely rare. Theories that focus on how state authority prevents spirals of hostility or enables credible commitments to interethnic political bargains may also overpredict violence.[3] In those many spaces where state authority is absent or weak—for example, in many of the post-Soviet republics—interethnic relations frequently remain cooperative. Here, we seek to advance a theory that can account for ethnic violence without overpredicting its occurrence. We argue that decentralized, nonstate institutional mechanisms may often arise to mitigate problems of opportunism in interactions between individuals from different ethnic groups. Furthermore, we show how different decentralized mechanisms that work to preserve interethnic peace most of the time may have very different consequences for ethnic violence in the (relatively rare) event of breakdowns.

We should emphasize at the start that we are not offering a full causal theory of either ethnic peace or violence. We specify what we believe are important causal mechanisms that appear to have been systematically neglected in the anthropological, sociological, and political science literatures.[4] But we do not pretend that our formulation of the problem of opportunism in interethnic interactions or the decentralized institutional mechanisms we identify tell the whole causal story. A richer story would surely include those narratives of interethnic injury. It might also include the motivations stemming from indignities suffered by peoples who are considered of lower rank and who seek to overturn a rigid social ordering (Horowitz 1985); the motivations

James D. Fearon is Assistant Professor of Political Science, and David D. Laitin is Professor of Political Science, University of Chicago, Chicago, IL 60637.

This is a revised version of a paper delivered at the Annual Meeting of the American Political Science Association, Chicago, September 2–5, 1995. For valuable comments we wish to thank Arun Agrawal, Robert Bates, Gary Becker, Rupen Cetinyan, Michael Chwe, Jeff Frieden, Arman Grigorian, Jennifer Hochschild, Timur Kuran, Margaret Levi, Roger Petersen, Kenneth Shepsle, Curt Signorino, Alan Zuckerman, and seminar participants at the University of Chicago and Harvard University.

[1] This childhood memory of Laitin's nicely illustrates our main arguments.

[2] See, for an example, Kaplan 1993. A striking exception in the classical anthropological literature is Gluckman 1955 (138–9).

[3] See, for example, Fearon 1994 and Posen 1993. Neither of the authors argues that an absent or weakened state authority will necessarily make for conflict between groups. For Posen, conditions concerning geographic settlement patterns must be met; for Fearon, the minority must expect a permanent decline in relative military capability *and* significantly fear what the majority may do in the future. Nonetheless, both arguments may still overpredict interethnic violence; see below.

[4] Anthropologists, for instance, have spent many pages carefully describing the institutional mechanisms that regulate disputes within tribal societies but hardly any concerning disputes across groups.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

stemming from the angers and frustrations that result from differential rates of economic development and career opportunities (Bates 1983, Gellner 1983); the ambitions of political entrepreneurs within an ethnic group who seek to enhance their power or influence through the political and/or military mobilization of their ethnic group (Gagnon 1994/95, Laitin 1995b); and, as a necessary condition for all ethnic violence, the availability of (usually young) thugs willing to risk their future (possibly to enhance their in-group status as worthy warriors) to become fully engaged in a culture of violence (Laitin 1995b). Finally, a thorough account of how ethnic disputes are contained would more systematically investigate the role of state policing (though we return to this mechanism in the conclusion).

The structure of the article is as follows. We first provide some evidence to support our claim that interethnic violence is the exception rather than the rule. Next, we informally sketch a theory of interethnic peace and violence that starts from individual-level problems of opportunism rather than from group-level animosities, as do existing rationalist and psychological theories of ethnic relations. We then make this argument more precise by developing a game model of everyday interactions within and across groups and by analyzing the two types of cooperative equilibria that emerge. In the fourth section we check the model's robustness and extract additional insights by considering alternative assumptions. We then consider empirical examples that suggest how the model and its conceptions correspond to real-world dilemmas and solutions to ethnic tensions. In the Conclusion, we summarize our arguments and offer suggestions on how the model may be extended.

THE RARITY OF ETHNIC VIOLENCE

A great many analysts of interethnic relations, including ourselves, agree that ethnic tensions are pervasive and commonplace. The standard view in political science, however, goes farther by suggesting that ethnic violence and active conflict also are ubiquitous. Violence is assumed to follow ethnic tensions as night follows day. For example, Daniel P. Moynihan (1993, 5) observes that "nation states no longer seem inclined to go to war with one another, but ethnic groups fight all the time." In the second sentence of his classic book on the subject, Donald Horowitz (1985, xi) writes: "By one reckoning, ethnic violence since World War II has claimed more than ten million lives, and in the last two decades ethnic conflict has become especially widespread."

This widely accepted view seems to be based on a biased selection of cases. Scholars have focused their attention overwhelmingly on cases of significant ethnic violence—they "select on the dependent variable." To demonstrate this point, we shall show that in two regions of the world viewed by many scholars as especially subject to violent ethnic strife—the post-Soviet world and Africa in the first generation after independence—cases of actual ethnic violence were vastly fewer than cases of potential ethnic violence.

For the former Soviet Union, although we have not found a database that would allow for replication, an informal survey of the interethnic scene supports this claim. Consider, first, the situation of Russians as an ethnic minority in the non-Russian republics. Of the fourteen former union republics (now independent states) excluding the Russian Federation, in only one (Moldova) has there been violent conflict between the titular nationality (the group after which the republic was named) and Russians. Within the Russian Federation, of sixteen former autonomous republics, in only two (Chechnia and Tuva) has there been sustained violence between titulars and Russians. This means that in these 30 (non-Russian) post-Soviet republics, in only three (10%) has there been Russian/titular ethnic violence.

Now consider ethnic violence involving non-Russians in the fourteen former union republics. These contain forty-five nationality groups that are neither titular nor Russian and that constitute at least 1% of the population of the republic.[5] Yet, there are only two cases of sustained communal violence between non-Russian minorities and titulars (Ossetians in Georgia and Armenians in Azerbaidjan). There are several cases of communal violence involving groups with less than 1% of the republican population (e.g., Abkhazis and Adhzeris in Georgia; Meshketian Turks in a variety of republics), but if we allowed *all* such cases, the percentage in which there was ethnic violence would be quite low. Thus, from this sample of non-Russian minorities, in only 4.4% of the potential cases (2 of 45) was there actual sustained violence with the titular group. And if we consider potential cases of ethnic violence between or among nontitulars, the rate for the 1991–95 period would be lower still. Even in the extremely unsettled post-Soviet world, while some ethnic violence has been horrible, it is far from ubiquitous.

Now consider Africa, a continent that many have described as especially prone to communal conflicts. In Table 1, we compute estimates of the ratio of actual to potential cases for four indicators of communal violence taken from the data in Morrison, Mitchell, and Paden (1989, 129): ethnic violence, "an event of short duration . . . in which two identifiable communal groups are antagonists in violence to secure some short-term goal"; irredentism, "an event in which an identifiable communal group seeks to change its political allegiance from the government of the territorial unit in which it resides to a political system . . . in which the authorities share the communal identification of the irredentist group concerned"; rebellion, "an event in which an identifiable communal group seeks by violence to gain increased autonomy from the national political authorities"; and civil war, "an event in which an identifiable communal group . . . seeks by violence to form a new political system based on boundaries of ethnic community."[6]

[5] Groups are counted in each republic in which they appear. The data on groups are drawn from Bremmer and Taras 1993. The data on violence come from a variety of sources. It should be noted as well that violence that occurred principally during the Soviet period (e.g., against Uzbeks in Kyrgyzstan) is not counted.

[6] See Morrison, Mitchell, and Paden (1989, 129) for their complete definitions of these four types of communal violence; Part II of their book gives country-by-country codings.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

TABLE 1. Estimates of Actual and Potential Communal Violence in Africa, Independence through 1979

Type of Communal Violence	Number of Incidents for All Years and Countries[a]	Country Mean of Incidents per Year[b]	Number of Potential Incidents for All Countries and Years[c]	Country Mean of Potential Incidents per Year[d]	Ratio of All Actual Incidents to All Potential Incidents[e]
Ethnic violence	20	.03	38,383	58.86	.0005
Irredentism	29	.04	18,757	26	.0015
Rebellion	27	.04	18,757	26	.0014
Civil war	52	.10	18,757	26	.0028

Source: Based on data in Morrison, Mitchell, and Paden 1989.

Note: There are 36 countries in the sample, those for which Morrison, Mitchell, and Paden (1989) report data on communal violence. See the text for summary statements of Morrison, Mitchell, and Paden's definitions of the four types of communal violence. In constructing the number of potential incidents of communal violence, we used Morrison, Mitchell, and Paden's estimates of the number of languages as a proxy for the number of ethnic groups in each country. Morrison, Mitchell, and Paden rely on diverse data sets, however, and this allows for more than one way to estimate the number of languages per country (which is inevitably a somewhat arbitrary endeavor). In constructing the table above we used the number of indigenous languages spoken in each country as given in their Table 2.3 (p. 43), which is based on the estimates of the linguist Jan Knappert (1965). For Knappert, the mean number of languages per country is 27. While these estimates have the virtue of deriving from (presumably) consistent coding criteria across countries, critics might argue that Knappert counts some dialectical variations as language differences and thereby his data may overestimate the number of "ethnic groups," if by that we mean the ethnic categories in which people place themselves and others. To be cautious, we therefore recalculated the columns in the table using Morrison, Mitchell, and Paden's lists of language groups in their country profiles. These data may have inconsistent coding criteria, but by often overaggregating ethnic groups (e.g., combining Kikuyu and Kamba in Kenya), they appear to provide a much more conservative estimate of the number of languages in each country. With this enumeration, the country mean is only 4.5 languages. Nonetheless, this underestimate produces fundamentally similar results to those presented above. Using this measure, the last column of the table would be: Ethnic violence, 0.0039; Irredentism, 0.0128; Rebellion, 0.0120; Civil war, 0.0230.

[a]For purposes of coding, actual cases of communal violence (of whatever form) that persisted for three years are counted three times, one for each year; two independent conflicts in the same year are coded as two incidents for that year.

[b]The mean for all countries of all incidents in a country divided by the number of full years since its independence through 1979. (Countries independent before 1960 are counted as if they became independent in 1960.)

[c]For irredentism, rebellion, and civil war, we estimated potential cases per year in each country as the number of ethnic groups in the country less one ($N - 1$), under the assumption that typically one group holds power while potential challengers come from all other groups (Morrison et al. 1989, 43, data on the number of language groups were used as a proxy for the number of ethnic groups in a country). We then summed across countries and years to arrive at the figures in this column. For potential cases of ethnic violence, we summed (across years and countries) a conservative estimate of the number of ethnic groups in each country engaged in regular interactions, namely, the smaller of $2N$ and $N(N - 1)/2$. If there are N groups, then the total number of dyads is $N(N - 1)/2$. We reason that in urban areas the upper bound $N(N - 1)/2$ may be approached, while in rural areas most ethnic groups are in regular contact with two or three other groups. Note that the results would barely change if we chose N as the estimate.

[d]The mean for all countries of potential incidents per year.

[e]Computed by dividing column 1 by column 3.

To avoid selection bias, we compare actual cases per year to indicators of potential cases per year, constructed as follows. For "ethnic violence," potential cases per country and year are a conservative estimate of the number of ethnic dyads in regular interaction in the country. For the other three indicators, potential cases per country and year are the number of ethnic dyads in the state that include the group holding state power (since that group cannot be irredentist, secessionist, or engaged in rebellion, as defined by Morrison, Mitchell, and Paden).

As should be astoundingly clear from Table 1, despite the conventional wisdom that ethnic violence is ubiquitous under conditions of cultural pluralism and weak states, the mean figure of actual violent communal events as a percentage of potential events hovers around zero. From around 1960 to 1979, communal violence, though horrifying, was extremely rare in Africa.

These data support our premise. There is a need in political science for a theory to account not only for occasional outbreaks of ethnic violence but also for the much more common outcome of ethnic tensions that do not lead to sustained intergroup violence.

ETHNIC CONFLICT AND THE PROBLEM OF SOCIAL ORDER

In broad terms, the large literature on ethnic groups and interethnic conflict offers two classes of theoretical arguments to explain the existence of such groups and the sources of conflict between them. Rationalist theories view ethnic groups as coalitions formed to extract material benefits from others or to defend possessions (Bates 1983, Gellner 1983, Olzak 1992). Violence may occur when one group threatens another's access to material benefits, which may take the form of state patronage, education, or control of specific labor markets. Psychological theories perceive ethnic groups as satisfying an inherent need to belong to a group and as allowing group members to maintain or enhance self-esteem. Conflict occurs when these psychological satisfactions are somehow threatened by another group or by members of one's own group (Horowitz 1985, chapter 3; Tajfel 1982).

Both the rational-coalition and psychological theories are intuitively plausible and probably help explain a range of particular cases of interethnic violence. Without wishing to dismiss either argument, we will pursue a new approach that postulates a different rationale for "groupness" and locates the problem of sustaining interethnic peace at the level of individual interactions rather than group-level grievances and animosities.

Our starting point finds the rationale for groups in the problem of social order. The motivating idea is that a great variety of human transactions and interactions involve the possibility of opportunism—self-interested behavior that has socially harmful consequences. Examples include cheating, shirking, malfeasance, fraud, exploitation, embezzlement, extortion, robbery, and rape. If unchecked by formal or informal institutions, the

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

expectation of such opportunism leads individuals to avoid interactions or to take costly actions to protect themselves in them, making for a "society" of fear, poverty, and disorder. Solving this problem of social order means somehow making it possible for individuals to trust one another in everyday interactions that have scope for opportunism.

The solution characteristic of the modern era is Leviathan, a legal apparatus backed by an agency that aims to monopolize the means of violence. But as Hobbes himself recognized, the writ of the state can in practice extend only so far—even in well-ordered states travelers will still fear highwaymen, and people still lock their doors at night (Hobbes 1985, chapter 13). Even in the presence of a well-functioning legal apparatus, problems of opportunism still dog a tremendous range of interactions and transactions. Police do not patrol all corners at all times; the details of business and other transactions are observed only by those party to them; contracts are incomplete by necessity and sometimes by incentive; recourse to courts is costly, and the results are unpredictable; and the powers of legal discovery are limited. In practice, moreover, in many or even most places where ethnic groups intermingle, a well-functioning state and legal apparatus does not exist anyway. Instead, judges and judgments are partial or partisan, and the formal legal apparatus of the state is haphazard, weak, and corrupt. This all means that, whether the state is well developed and powerful or relatively weak, a whole host of everyday transactions and interactions will fall below or outside its ambit and are prone to problems of opportunism.[7]

The limited reach of the state opens up a space for more local and decentralized mechanisms for resolving these dilemmas. Often, these mechanisms developed long before the present-day state apparatus and persist after it appears.[8] For example, a common mechanism for defeating problems of opportunism is reputation in a frequently repeated relationship. Cooperation and trust within families are sustained in part by the expectation of many future interactions, as they are in many business and trading relationships. This type of solution is captured in schematic form by cooperation between two players in a repeated prisoner's dilemma. Here, strategies such as tit-for-tat that threaten to revert to noncooperation if the other player defects can sustain cooperation if the players are patient enough and expect a high enough likelihood of future interactions, despite the absence of any third-party enforcer like the state.[9]

While very important empirically for certain interactions, reputation within a long-term repeated relationship faces a major liability in that there are frequently significant benefits to be had from interacting or trading with numerous others. The more people with whom one interacts, the greater is the possible division of labor and gains from trade, and the greater the number of possible marriage partners, friends, contacts, entertainment opportunities, and so forth. But the greater the number of people with whom one interacts, the lower is the frequency of interaction with each individual on average, and this can undermine the two-party reputation mechanism. Thus, something more than that mechanism may be required to capture various benefits that would result from overcoming problems of opportunism in societies more "anonymous" than a family or a business partnership.[10]

Ethnic groups can provide a partial solution to this problem for their own members.[11] At bottom, the problem is informational. To support trust and cooperation among people who interact or transact relatively irregularly and infrequently, individuals need to be able to condition cooperative behavior on their partners' past behavior or history. That is, it must be possible to identify people who have cheated or exploited in the past if they are to be sanctioned and others deterred from misbehaving in the first place.[12] Thus, any formal or informal institutional practices that allow people to learn at low cost about the prior history of a potential business or social partner may help resolve the problem of social order in groups larger than a family.

Ethnic groups are frequently marked by highly developed systems of social networks that allow for cheap and rapid transmission of information about individuals and their past histories. Even if two Ukrainian émigrés in Canada have never met, they can probably learn something about the other's past in short order by "asking around."[13] In the case of relatively small groups living in

[7] Paradoxically, states themselves may create spaces of anarchy within hierarchy. They may do so intentionally, as when certain sorts of contract are deemed illegal (e.g., oligopolistic collusion), or unintentionally, as when making narcotics illegal creates a black market with a need for enforcement institutions (mafias). On the latter, see Gambetta 1993. In addition, it is crucially important that the state not enforce contracts among individuals to cooperate to overthrow the state; this is a necessary condition for the Leviathan solution to work. See Wagner 1993 for this point.

[8] Putnam (1993) suggests that the quality and nature of these local mechanisms are actually critical determinants of whether state institutions perform well. Greif (1989, 1994) and Milgrom, North, and Weingast (1990) have analyzed pre-state institutional mechanisms for policing trade relations.

[9] Taylor (1976) was the first to examine the implications of this mechanism for standard liberal justifications for the state. See also Axelrod 1984.

[10] This argument can be made more formally. Imagine a population of individuals who can pair off with one another in successive periods in order to play a prisoner's dilemma game. It would make sense for them to choose partners and stick to the same one "forever," since they could then use reputation in a long-term relationship to police cooperation. Yet, if the benefits of mutual cooperation increase with the average number of partners with whom both parties interact over time, then playing repeatedly with the same partner would be inefficient.

[11] Grief's (1989, 1994) work views certain "ethnic" institutions as solutions to problems of opportunism in long-distance trading relations. For an analysis of ethnic interactions that draws on some of the same concepts, although focusing more on issues of property rights, see Landa 1994.

[12] Ellison (1994) shows that in a strict sense this claim is too strong, although his example of how cooperation may be supported despite anonymity has very little if any empirical resonance.

[13] Julian Zahalak, personal communication. For anthropologists' examples drawn from a range of settings, see Colson 1974 (chapter 3, esp. pp. 54–9), who stresses the role of dense social networks and the process "whereby public opinion becomes mobilized through gossip" in the internal policing of cooperation in groups.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

close proximity, it is often the case that everyone knows everything about everybody, more or less. And insofar as members of an ethnic group tend to interact more frequently with one another than with outsiders, the simple two-party reputation mechanism may be employable within groups as well.[14]

The fact that ethnic groups are often characterized by relatively dense social networks and low-cost access to information about the past history of individuals' behavior has an important consequence for intra- versus intergroup relations. Within groups, people who exploit the trust of others can be identified as individuals and sanctioned with relative ease by the response of the ethnic community. In game-theoretic terms, cooperation and trust can be supported within an ethnic group by punishment strategies that are conditioned on individual behavior, because the cost of obtaining information about an individual's past is low. By contrast, individual identification is harder in interactions across groups. Because social networks are less developed, it is more difficult to get information on a potential trading or social partner from "across the tracks." And if individuals are hard to identify or investigate across ethnic groups, then cooperation and trust across groups cannot be supported by punishment strategies that condition on individual behavior. For example, if you know nothing more than that the person facing you is a Serb, then you cannot condition your behavior on how the person acted in the past, but only on the fact that the other person is a Serb. Moreover, "the Serb" may have no individual reputation to worry about protecting in interactions with non-Serbs.

Relatively dense social networks and interactions within an ethnic group thus give rise to an asymmetry of information (identifying and getting information about individuals from other groups is more difficult). Any institutional regime for maintaining cooperation across groups must somehow address the problem this asymmetry poses. If one cannot be identified as an individual in some social interaction or if one can anticipate that what one does will not affect one's reputation, then one may have an incentive to exploit the other person or people involved.

Below, we use a game-theoretic model to make this informal story more precise and to explore its implications. In the model, we find two types of equilibria that support ethnic cooperation. Either (1) each group may hold all members of the other group liable for the actions of its individual members, or (2) the members of one group may simply ignore violations of trust by members of the other group, relying instead on the other group to identify and sanction the appropriate individual.

Under the first regime, the members of group A indiscriminately punish all members of group B for nasty behavior by an individual B. This is the *spiral regime*, in which individual defections trigger an escalation and complete breakdown of intergroup relations—noncooperation spreads immediately to all interactions between members of the groups. Under the second regime, *in-group policing*, individual defections do not trigger such a spiral and total breakdown. Instead, if a B exploits an A, members of group A continue cooperating with members of group B as though nothing had happened, while members of group B identify and sanction the individual who acted badly. Thus, ethnic groups cooperate to take advantage of each side's superior information about the behavior of individuals within the group, and this leads to the containment of interethnic violence.

FIGURE 1. The Prisoner's Dilemma Stage Game

	Cooperate	Defect
Cooperate	1, 1	$-\beta, \alpha$
Defect	$\alpha, -\beta$	0, 0

Notes: $\alpha > 1$, $\beta > 0$, and $(\alpha - \beta)/2 < 1$.

A MODEL OF DECENTRALIZED COOPERATION BETWEEN ETHNIC GROUPS

In this section we use a social matching game to formalize and expand on the above argument. In such a model, many individuals are randomly paired in successive periods, and each pair plays some game (typically a prisoner's dilemma, since this is a simple model of an exchange relation subject to opportunism). We begin by considering interactions within a single group in order to introduce notation and theoretical issues that arise in this class of models.[15] We then proceed to consider the case of two such groups whose members interact both internally and across groups.

Social Matching Games

Consider a group with n members represented by the set $A = \{1, 2, 3, \ldots, i, \ldots, n\}$, where i is the name of "person i," and n is an even number greater than three. The members of A will be paired randomly in successive periods, $t = 0, 1, 2, \ldots$, to play the prisoner's dilemma shown in Figure 1. Suppose that the individuals have a common per-period discount factor $\delta \in (0, 1)$. As an example of what this set-up is intended to represent, one may imagine a number of traders who go to a market each day, meet other traders, arrange or fail to arrange deals, and return the next day to interact with new partners (probably). Or one may imagine that each day people wander out into the world and have chance social encounters that may be good or bad, depending on how both parties act. "Defection" can be interpreted as an attempt to cheat or rob the other player, or in some

[14] Our notion that ethnic groups are marked by high levels of interaction parallels Karl Deutsch's communications theory of the emergence of political communities. Different from Deutsch (1966, 177), who relates peaceful outcomes to shared meanings, we relate peaceful outcomes to shared information, which permits reputation-building and trust.

[15] For social matching games considered in the game-theoretic literature, see Calvert 1995; Ellison 1994; Kandori 1992; and Milgrom, North, and Weingast 1990. Our analysis owes a great deal to Kandori's work in particular.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

contexts can be taken as avoiding or forgoing the interaction.

Three general points about this class of models are worth making by way of introduction. First, the information the players receive about what happens in other pairs' interactions is important. The polar cases are (1) full information, that is, each individual observes what happens in every interaction in every period, and (2) atomized information, that is, individuals observe what happens only in the interactions in which they participate. Information matters because the less players know about a partner's history, the less they are able to condition how they act on the partner's past behavior.[16]

Second, social matching games normally have many equilibria, a feature they share with simpler repeated games. For example, the following strategy profiles can all form stable patterns of play (equilibria) in the game under the full information condition.

1. Every individual defects in every period, regardless of the past history of play. This is a Nash and subgame perfect equilibrium for any discount factor δ, and remains so even under atomized information.[17]
2. Individuals begin by cooperating, and if any individual is observed to have defected in any period, then all players switch to the "nasty" equilibrium of (1). This is a Nash and subgame perfect Nash equilibrium for large enough δ.
3. Individuals cooperate with any player who is not in "punishment phase" and defect against any player being punished. In one variant (3a), players in punishment phase always defect while being punished. In another (3b), players being punished cooperate with partners not being punished and defect when paired with other players in punishment phase. In both cases, a player enters the punishment phase for $T > 0$ periods if s/he defects against a player who is not in punishment phase. In (3b), if a player defects against a cooperator while being punished, the punishment phase is restarted. Both (3a) and (3b) can form Nash and subgame perfect Nash equilibria for sufficiently large δ and an appropriately chosen T.[18]

Under (1), society is characterized by multilateral nastiness and lack of trust, whereas in (2) and (3) social expectations give people the incentive to act cooperatively with each partner they meet. Note that while the *observed* equilibrium behavior of individuals in societies under (2) and (3) is identical—all cooperate with every partner they meet—this observed behavior is driven by very different social expectations about what would happen if anyone were to defect. Under (3), people expect that only defectors are punished while others continue cooperating; under (2), they expect that one defection leads everyone to start defecting. In game-theoretic terminology, equilibria (2) and (3) are the same on the equilibrium path (everyone cooperates each period), but they differ off the equilibrium path (in the event someone defects). Regarding interethnic relations, the intergroup cooperation we observe in different settings may be based on very different expectations and beliefs about what would happen in the event of a breakdown (off the equilibrium path).

The comparison of (1), (2), and (3) makes it clear that equilibria of social matching games can differ greatly in their normative and/or intuitive appeal. For instance, equilibrium (1) is inefficient since everyone gets zero each period when all would certainly prefer one. By contrast, (2) and (3) are efficient on the equilibrium path, since everyone is cooperating. Given our interest in whether and how ethnic cooperation across groups can be sustained, the analysis below focuses on equilibria that are efficient on the equilibrium path.

But the comparison of (2) and (3) suggests a further criterion as well—robustness. As stated, equilibrium (2) is not robust in the sense that if anyone deviates, then cooperation breaks down forever. This is a normatively unattractive feature if one allows for some "noise" in individuals' choices, that is, if we allow for occasional mistakes, misinterpretations of opponents' moves, or unobserved, random variability in payoffs from encounter to encounter. In such cases, play will eventually leave the path of cooperation, and the social equilibrium in the longer run will be that to which the system ultimately moves as a result of such random deviations. In what follows, we will consider only equilibria that ultimately return to mutual cooperation and thus are robust against such noise.

A final criterion for evaluating different equilibria is efficiency in the presence of noise. If there is no noise in the system, equilibria (2) and (3) are equally good. Everyone gets a payoff of one each period under both sets of expectations, as long as play remains on the equilibrium path. But if someone makes a mistake or a misinterpretation, then (2) is a decidely less attractive arrangement for society than is (3), since under (3) only the defector is punished, while under (2) everyone in society suffers. For just this reason, (2) may seem intuitively unnatural. If one were to accept the common presumption in economic theory that relatively efficient equilibria (and institutional arrangements) are more likely to evolve or to be chosen by planners, then (3) would be judged more likely to occur than (2).[19]

[16] Note that under the full information condition a "group history" is commonly known, whereas under atomized information there is just a collection of privately observed "personal" histories.

[17] Loosely, in a Nash equilibrium everyone is choosing optimally, given what the other players are choosing. Subgame perfection adds the requirement that players do not anticipate others will carry out threats or promises not in their interest if they actually have to choose.

[18] Neither (2) nor (3) can be applied in the atomized information condition because both require players to condition their behavior on the past behavior of each partner. For 3a the condition on δ for subgame perfection may be very restrictive, due to a problem pointed out by Kandori (1992): If many players are in punishment phase, cooperators who meet may be tempted to cheat each other because they are not likely to meet other cooperators again for a long time. Kandori suggests a technical solution for this problem ("forgiving" old defectors when a new defection occurs), and it does not arise at all in the "renegotiation proof" punishment of (3b) (see footnote 19).

[19] There is also a more sophisticated argument for efficiency off the equilibrium path, termed "renegotiation proofness." Equilibria sustained by the threat of punishments that are costly for everyone are held to be incredible, on the ground that after a defection all will have an incentive to renegotiate immediate cooperation. To avoid this problem, in a renegotiation-proof punishment, the defector must make amends (or provide compensation) by letting nondefectors exploit him

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

To summarize, we will confine our attention to robust subgame perfect Nash equilibria that are efficient on the equilibrium path, and we compare equilibria according to how well they perform in the presence of small amounts of noise. This follows from our concern with exploring the formal and informal institutional mechanisms that support and maintain interethnic cooperation on a day-to-day basis.

A Social Matching Game with Two Groups

Suppose that there is a second group represented by the set $B = \{n + 1, n + 2, \ldots, 2n\}$. In each period, k individuals will be drawn at random from each group and paired for cross-group interactions, while the remaining $n - k$ in each group are paired among themselves for in-group interactions. Thus, $p = k/n$ is the probability that an individual interacts across groups in any period. We assume that k is less than $n/2$—thus, $p < 1/2$—so that in-group interactions are more common than cross-group interactions.[20]

As argued above, ethnic groups are typically marked by relatively dense social networks and low-cost access to information about other group members' behavior; it is often much easier to gain information about in-group than out-group members. While it is not literally true even in small ethnic groups that all members observe all interactions (the "full information condition"), rumor, gossip, and inquiry tend to be more developed and efficacious within than between ethnic groups. Furthermore, as we discuss below, in many places ethnic groups are delimited by a variety of institutions (churches, schools, respected elders, etc.) that certify and advertise individuals' reputations within the group.

For these reasons and as a first approximation, we will suppose that players in the model can identify coethnics as individuals in in-group interactions and, furthermore, that they know one another's history of play. By contrast, in cross-group interactions, individuals know only that they are paired with someone from the other group; they observe neither the individual identity nor personal history of an out-group partner. For example, imagine two groups living in nearby villages or city neighborhoods; members observe (or learn about) what goes on within their own community and what happens in occasional cross-group interactions. They do not observe what happens within the other ethnic community, and they do not learn the names or backgrounds of individuals whom they or their coethnics meet in cross-group interactions.

These assumptions are not intended to characterize every type of interethnic interaction, such as a durable, long-term relationship between two traders, for example, or members of different groups who live in the same village or neighborhood and interact very frequently with one another. Rather, these assumptions depict a particular class of interactions that pose a difficult problem for interethnic cooperation due to informational asymmetries and relative infrequency. Nor are we saying that there are ethnic groups whose members observe or learn about everything that takes place within the group. Rather, the assumption of full information within groups is intended as a tractable simplification of a much weaker and more plausible assumption—that in many places for many communities, individuals have *relatively* easier and lower cost access to information about coethnics than about individuals from other groups, whether due to gossip and rumor or to formal or informal in-group institutions.[21]

Formally, then, a history of play up to period t for the individuals in a group—say, A—will be a list of (1) who was paired with whom and what happened in all intra-A pairings up to period t and (2) which members of group A were paired with Bs and what happened in each period up to t. Members of group A observe neither what happens among the B players nor the identity or history of B players paired with As. A complete strategy for a player is a set of maps, each from the set of all such histories to period t into the range {C, D}, with one map for each $t \geq 0$.

The question of interest for this game (call it G) is whether and how cooperation can be supported in the *inter*ethnic interactions, which are essentially anonymous and also relatively infrequent. Anonymity implies that members of the other group cannot hold a cheater individually responsible, thus short-circuiting the mechanism used within groups to support cooperation, as in equilibrium (3) above. Infrequent interactions will tend to undermine the incentive to act nicely even if we were to assume that interethnic interactions are not anonymous.

In order for individuals to be willing to cooperate in the interethnic interactions in an equilibrium, they must face some sort of punishment (that is, loss of future payoffs) for defecting. Logically, there are three possible ways such punishment can be delivered: by members of the other group; by members of the defector's own group; or by members of both groups. We will describe strategies of the first and second types that can support interethnic cooperation as equilibria, and then we will consider how the two equilibria fare in the presence of noise in individual interactions.[22]

At first glance, punishment by members of the other group may seem most natural, since a defector in an interethnic interaction "offends against" the other

or her (as in 3b), which gives them a positive incentive to carry out the punishment.

[20] Assume also that k and n are both even, and n is greater than 7.

[21] We briefly introduce in-group institutions that allow us to weaken this assumption below. An alternative approach, not taken here, would be to replace the assumption of full information with a mechanism of "gossip." For example, one may assume that if i's partner j defects, i reports this to all subsequent partners, they to their partners, and so on; those who have heard about j's defection then sanction j for a specified number of rounds. For any given δ, there will then be a size n^* such that for all group sizes $n \leq n^*$ cooperation can be supported by the threat of gossip "contagiously" spreading the news of a defection. Cf. Kandori (1992) and Ellison (1994), who consider how the fear of a contagious defection process that results in complete social breakdown rather than focused punishments (as in the case of gossip) may be used to support cooperation in the atomized information condition.

[22] Analysis of the third possible approach to sanctioning cross-group defectors adds little insight; it also proves less efficient off the equilibrium path than does either (1) or (2).

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

group. But this scheme faces a dilemma: If members of the other group cannot identify (or gain access to) the individual defector, then they cannot hold the individual personally responsible. It follows that to support cross-group cooperation solely by threat of punishment by the offended group, punishment must be indiscriminate, targeting either all members of the cheater's group or some random collection of them. While such random reprisals or feuding behavior may seem grossly unfair to innocent members of the defector's group, they follow, in a strategic sense, from the fact that it is harder to identify and check up on members of other ethnic groups than on members of one's own.

The following strategy profile describes a spiral equilibrium in which interethnic cooperation is based on the off-equilibrium path threat of indiscriminate, group-based reprisal for individual defections. In the event of a defection in an interethnic encounter, noncooperation rapidly spirals beyond the original parties to the dispute, involving coethnics in a collective effort to punish the other group.[23]

4. *The spiral equilibrium.* All individuals play the strategy σ_S, defined as follows. In in-group pairings, always play C with any player not in the in-group punishment phase (a "cooperator"), and always play D against any player in the in-group punishment phase (a "defector"), regardless of one's own status. A player enters (or restarts) the in-group punishment phase for T^{in} periods by defecting against a cooperator. In out-group pairings, play C provided that neither group is in the out-group punishment phase, in which case play D. A group enters that phase for T^{out} periods if any member defects in a cross-group pairing when neither group is in the out-group punishment phase.[24]

Thus, the spiral equilibrium rests on a sort of interethnic deterrence—individuals cooperate in interethnic interactions for fear of losing future payoffs should they defect and cause a larger breakdown of intergroup relations. Within groups, by contrast, cooperation is supported by individual punishment, and group members do not punish their coethnics for defecting against members of the other group.[25]

For σ_S to form an equilibrium, three key conditions must obtain. First, for individuals to want to cooperate in interethnic encounters, the immediate gains of defecting ($\alpha - 1$) must be less than the expected payoffs forgone due to the "spiral" breakdown (which are $\delta p + \delta^2 p + \ldots + \delta^{T^{out}} p$, since there is a p chance of being paired outgroup each period, and in this event one forgoes the benefits of cooperation [1] over defection [0]). This immediately implies that interethnic interactions cannot be too infrequent for a spiral equilibrium to be possible. If p is too small, the threat of a breakdown of intergroup relations is not compelling enough to induce individual members to cooperate.[26]

Second, cooperators cannot have an incentive to defect against other cooperators within their own group. Third, those who defect within the group must have an incentive to comply with their punishment (they must prefer to accept $-\beta$ for T^{in} periods to regain cooperation status over continual defection). In contrast to the first, these second and third conditions imply that interethnic interactions cannot be too frequent for intraethnic cooperation to be supportable; the more often individuals interact outside their own group, the less effective is the threat of in-group social sanctions.[27] It has been noted in the anthropological literature that ethnic groups are marked by institutions which maintain and police ethnic boundaries by sanctioning individuals who interact too much or in certain ways with outsiders (Banton 1983, Barth 1969, see also Laitin 1995a). In our model, an incentive to police ethnic boundaries and so delimit ethnic identity arises endogenously.

The second approach to giving individuals an incentive to cooperate in interethnic interactions is to have groups sanction their own members for cross-group violations of the norm. The following strategy profile characterizes an in-group policing equilibrium in which defections in interethnic interactions are completely ignored by the offended group, while the defector is sanctioned within his or her own ethnic group.

5. *The in-group policing equilibrium.* All individuals follow the strategy σ_{igp}, defined as follows. Play C in all out-group pairings. For in-group pairings, always play C with any partner not in punishment phase, and always play D against any player in punishment phase (regardless of one's own status). A player enters (or restarts) the punishment phase for T^{igp} periods by defecting when paired with an out-group member or a coethnic who is not in punishment phase.[28]

Under the in-group policing regime, the two groups in effect make a deal that benefits both sides. By adopting a policy of "you identify and punish your miscreants and we will do the same," they take advantage of the fact that each group has better information about the behavior of its own members than about the other group and so can target individuals rather than whole groups.

The conditions necessary for the in-group policing strategy σ_{igp} to form an equilibrium of the game are simpler and typically less restrictive than those for the

[23] As a referee observed, this equilibrium does not represent a "spiral" in the sense of a dynamic process of action-reaction escalation. While punishment strategies that have this quality could easily be constructed, little insight would be gained in the present model. We intend "spiral" in the sense of rapid escalation beyond the immediate parties to the dispute.

[24] Necessary and sufficient conditions for σ_S to form a subgame perfect equilibrium of G (when it is chosen by all players) are given in proposition 2 in the Appendix.

[25] Note that we adopt the renegotiation-proof punishment strategy of (3b) for policing in-group cooperation here—the defector must accept being "the sucker" for T^{in} periods in order to be returned to the good graces of society. It is worth noting that while we also can adopt such a punishment regime for the interethnic interactions, the parameter conditions necessary to make this supportable as an equilibrium are extremely restrictive, because the infrequency of interethnic interactions greatly reduces the incentive to comply with being the sucker in the punishment phase.

[26] Formally, the necessary condition is that $p \geq (\alpha - 1)(1 - \delta)/\delta$.

[27] As shown in the Appendix, for these conditions to hold, the frequency of interethnic interaction p must be less than $1/(1 + \beta)$.

[28] Necessary and sufficient conditions for σ_{igp} to form a subgame perfect equilibrium of G are given in proposition 1 of the Appendix.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

spiral equilibrium. The reason is that a cooperator faces essentially the same incentives in both intra- and interethnic interactions. In both cases, defecting leads to T^{igp} periods of in-group punishment. Thus, equilibrium requires that cooperators have no incentive to defect in either in-group or out-group pairings (the same condition applies for both), and that defectors are willing to comply with their punishment in intraethnic interactions and have no desire to extend their punishment phase by defecting when paired with an ethnic other.

In contrast to the spiral equilibrium, under in-group policing the infrequency of interethnic interactions never poses an obstacle to intergroup cooperation—there is no lower bound on p below which interethnic cooperation is unsupportable.[29] The reason is that cooperation within the group has been linked to cooperation with out-group members. It does not matter whether one rarely interacts with members of the other group, since defecting in an out-group pairing will lead to sanctioning by in-group members with whom one interacts more frequently.[30]

Just as in the spiral equilibrium, however, maintaining intraethnic cooperation under in-group policing requires that interethnic interactions not be too frequent. Once again, the efficacy of in-group sanctions weakens as individuals interact more often with ethnic others, providing a rationale for institutions that police ethnic boundaries.[31]

Before considering how the spiral and in-group policing equilibria fare in the presence of "noise," we briefly examine a natural objection to the stability of in-group policing: Will groups be reluctant to sanction their own members for transgressions against the out-group? As we have specified σ_{igp}, there is no problem here, because σ_{igp} employs punishment strategies that actually benefit an individual who sanctions a defector; the strategy σ_{igp} requires that defectors make amends by cooperating while others in good standing defect against them. Thus, under σ_{igp}, individuals have a positive incentive to sanction people who have defected, no matter against whom.

Arguably, however, it is not always possible or realistic that sanctioning defectors will be costless to those who sanction. If it is costly—say, if players being punished and their partners both defect (as in equilibrium 3a), each getting zero rather than one—then the in-group policing regime is open to the above objection. Indeed, if we allow for explicit coordination and collusion within groups, then the in-group policing equilibrium may not be sustainable in this case. Consider the following proposal that could be made by a member of group A to coethnics when s/he is paired with a B: "I will defect and share the benefits $\alpha - 1$ with all of you if you agree not to punish me. Because of in-group policing, my defection will not lose us anything in interactions with the other group." If such in-group coordination is feasible, and for relatively small groups with their own internal governing apparatuses it probably would be, then the in-group policing equilibrium will break down in our model when sanctioning is costly.[32]

The problem in this instance is that, given our assumptions, groups have no way to verify that defectors in cross-group pairings are in fact being punished. In the real world, this seems an easy problem to solve—groups can simply make the punishment of defectors against the out-group public and observable to members of the other group, or they can just hand over the offender. Insofar as doing this is costly or difficult for the group harboring the offender, it may be necessary for the other group to provide an incentive in the form of a threat to return to a spiral equilibrium. In other words, failure to publicly sanction one's coethnics for violations against the out-group will lead the out-group to begin a spiral or feud. Thus, one can imagine an in-group policing regime that rests or depends on a deeper threat to resort to widespread feuding (or total avoidance); the threat of feuding induces in-group policing.[33]

Efficiency in the Presence of Accidents and "Noise"

The spiral and in-group policing regimes suggest how ethnic groups may manage to maintain mutually beneficial cooperative relations as the normal state of affairs, despite the obstacles posed by asymmetric information and the relative infrequency of interethnic interactions. And if play is kept to the equilibrium path, the spiral and in-group policing regimes are equally good in this respect.[34] But occasional defections (or perceived defections) are inevitable, due to mistakes, misinterpretations, drunkenness, sudden passions, or unobservable variations in payoffs from encounter to encounter. We need to ask, then, how the two regimes perform in the presence of such "noise."

Intuitively, one might expect fallible individuals to fare better under in-group policing than under the spiral regime. In-group policing focuses punishment on individuals, whereas in a spiral regime the whole group suffers a "collective bad" from the transgression of one member. Furthermore, if the groups are moderately large and there are many interethnic interactions in each period, then the probability that noise will produce a defection in at least one pairing is high, even if the amount of noise is low. This can mean that the groups spend most of the time not cooperating, or that the

[29] See the Appendix for a statement of parameter conditions under which both, neither, or only one of σ_{igp} and σ_S are supportable as equilibria. Briefly, if $\alpha < 1 + \beta$ (which is likely when defection is understood to entail violence), the in-group punishment regime is less restrictive in the sense that, for all other parameter values, if σ_S forms an equilibrium, then so does σ_{igp}, while the converse is not true.

[30] In an entirely different context, Lohmann (1997) shows how linked punishment strategies can support cooperation for lower δ's than would be possible without linked strategies.

[31] The upper bound on p is the smaller of $1/(1 + \beta)$ and $1 - ((\alpha - 1)/(1 + \beta))$.

[32] This problem can arise even with renegotiation-proof in-group punishments like those of σ_{igp}, if the benefits of cheating an ethnic other exceed the benefits of sanctioning a defector.

[33] There is quite a bit of evidence for this in the anthropological literature. See Colson 1974, chapter 3; Miller 1990; Moore 1978.

[34] They also will be observationally equivalent, without soliciting information about peoples' beliefs.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

TABLE 2. Performance of the Two Equilibria with "Noise" in Interethnic Interactions: Average Per-period Expected Equilibrium Payoffs for $\alpha = 1.1$, $\beta = 2$, $\delta = .95$

p	k	ε	V^*_{igp}	V^*_S	T^{out}_{min}
.05	100	.1	.973406	—	
	(n = 2,000)	.01	.997251	—	
		.001	.999724	.976064	6
		.0001	.999972	.997426	3
		0	1	1	3
	1,000	.1	.973407	—	
	(n = 20,000)	.01	.997251	—	
		.001	.999724	—	
		.0001	.999972	.976137	6
		0	1	1	3
.10	100	.1	.948266	—	
	(n = 1,000)	.01	.994641	—	
		.001	.999462	.974636	2
		.0001	.999946	.996420	2
		0	1	1	2
	1,000	.1	.948267	—	
	(n = 10,000)	.01	.994641	—	
		.001	.999462	—	
		.0001	.999946	.974839	2
		0	1	1	2
.20	100	.1	.902293	—	
	(n = 500)	.01	.989836	—	
		.001	.998980	.970113	1
		.0001	.999898	.996250	1
		0	1	1	1
	1,000	.1	.902297	—	
	(n = 5,000)	.01	.989836	—	
		.001	.998980	—	
		.0001	.999898	.970568	1
		0	1	1	1

Notes: V^*_{igp} and V^*_S give time-averaged per-period equilibrium payoffs for the in-group policing and spiral equilibria, respectively. A dash means that σ_S does not form a Nash equilibrium for these parameter values. T^{out}_{min} is the smallest T^{out} such that σ_S is supportable, given ε.

spiral regime breaks down entirely because the anticipation of noise producing a spiral gives individuals an incentive to defect *deliberately* in interethnic encounters.

These intuitions are borne out when we introduce a small amount of noise into the model. For simplicity, we assume that mistakes and misinterpretations sometimes occur in the interethnic pairings, while there is no such noise in intraethnic interactions.[35] Assume that in every interethnic pairing there is an $\varepsilon > 0$ chance that a player who intends to cooperate accidentally defects. For several different sets of parameter values, Table 2 gives an individual's ex ante expected per-period payoff given such noise under both the spiral and in-group policing regimes.[36]

Two main points emerge from these examples. First, in-group policing yields higher expected payoffs for all members of both groups than does the spiral equilibrium.[37] The size of the improvement varies in these examples from 0.5% to 2.5% per period, and improvement up to 10% per period can be generated in examples with a larger α.

Second, the spiral equilibrium is much less robust against the introduction of small amounts of noise in interethnic interactions than is in-group policing. Notice that the spiral regime typically ceases to be an equilibrium if ε exceeds even a tiny threshold, while the payoffs under in-group policing are scarcely affected by variation in ε. In these examples, ε has to be on the order of $1/10pn$ for players to have an incentive to cooperate with ethnic others under the spiral regime (recall n is group size, and p is the interethnic interaction rate). The reason for the fragility of the spiral regime is obvious. The odds that an accidental defection will occur in at least one interethnic encounter increase rapidly with the number of encounters. Thus, even for relatively small communities (e.g., 1,000 interethnic interactions each period), a small amount of noise (e.g., $\varepsilon = .001$) will still be associated with a relatively high risk of breakdown each period (in this case, 86% each period). Thus, there is a large temptation to defect on purpose since a breakdown is likely anyway.

One might try to render the spiral regime more robust by raising the threshold that triggers a spiral. Consider, for example, an equilibrium in which defection must occur in at least $j > 1$ interethnic encounters in a period for a breakdown to follow. Then, for an appropriately chosen threshold j, individuals might be deterred from defecting against ethnic others by the risk that they will tip relations over the edge. Perhaps we could then maintain the spiral regime by raising the threshold as group size increases.[38]

Though we lack the space to give the details, the scope for improvement here turns out to be almost nil. Briefly, when the number of interethnic interactions is moderately large, the *marginal* probability that any one person's defection will tip the groups into conflict is infinitesimal for all but the lowest thresholds (i.e., one or two), so we are back to the problem that a small amount of noise can imply a large risk of breakdown. In essence, with groups of moderate or large size, players face a collective action problem under the spiral regime. Opportunistic behavior with ethnic others is individually beneficial but socially costly due to indiscriminate punishment. In-group policing may be seen as an arrange-

[35] This conforms with the notion that cultural (and perhaps language) differences make misinterpretation relatively more likely in interethnic interactions.

[36] The table gives time-averaged payoffs, which are the players' ex ante expected value for the game multiplied by $1 - \delta$. See the Appendix for the expressions that yield these numbers. Note also that the expected per-period payoff would be 1 without noise, and without any interethnic cooperation it would be $1 - p$.

[37] We are unable to find any parameter values such that the reverse holds for σ_{igp} and σ_S. If in-group punishment strategies are changed to make sanctioning costly, then for very small groups it can be possible for the spiral equilibrium to be more efficient with small amounts of noise.

[38] The origins of this idea are in Green and Porter (1984), who showed how oligopolists may police collusion despite imperfect monitoring by appropriately setting a "trigger price" below which the firms begin a price war. For applications in political science, see Downs and Rocke 1990, 1995.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

ment which, if effective, dramatically lessens this collective action problem. Indeed, a spiral regime may have a natural tendency to develop features of in-group policing, as it makes sense for individuals to try to prevent others in their group from provoking conflict that hurts all.[39]

SOME ALTERNATIVE ASSUMPTIONS AND ADDITIONAL IMPLICATIONS

In this section we consider the effects of weakening and varying some of the assumptions that underlie the game studied above. New insights emerge, and the core insight proves robust, namely, that two quite different types of social institutions can enable interethnic cooperation in domains in which the state is weak or absent.

Interethnic Interactions that Are Not Anonymous

For the game G, we assumed that individuals cannot subsequently recognize or reliably identify ethnic others who cheat or exploit them in interethnic encounters (or that they cannot reliably communicate this information to their ethnic brethren). Suppose instead that interethnic encounters are not anonymous in this sense, so that if a B cheats an A, the other members of the A group observe the identity of the B player (or can learn it reliably). In that case, interethnic cooperation may be supported by focused and individual rather than indiscriminate punishment; i may be deterred from cheating an ethnic other j by the prospect that i (and i alone) will be "boycotted" or otherwise sanctioned by the members of j's group. Thus, no spiral of defection beyond i and j need occur. There are two problems with this regime, however.

First, insofar as interethnic interactions are still relatively infrequent, the force of the threat by j's group to punish i may be undermined, just as in the spiral regime case. Even if members of j's group recognize i, they may rarely encounter or have great difficulty finding i. An in-group policing scheme thus again appears as the natural alternative solution (short of more complex state-based solutions). Second, if all players on both sides do not observe the interethnic interaction in which i cheats j, then the potential for a new kind of spiral or feud emerges. Namely, if members of j's group sanction i, then members of i's group may be unsure (or misled) as to whether this is retaliatory punishment or a transgression that in turn requires sanctioning. Thus, imperfect observability of interethnic encounters may imply that even focused cross-group punishment schemes will be accompanied by periodic tit-for-tat spirals and feuds.

[39] As Miller (1990, 198) argues with respect to Icelandic bloodfeuds, group liability "had the effect of inducing people who might be held accountable for each other's actions to involve themselves in each other's affairs" and thus made the feud "much more effective as an instrument of social control than it would otherwise have been if only the actual wrongdoer suffered the consequences of his actions." On this point see also Colson 1974, 40–1; Gellner 1988; Hardin 1995, 118–9; and Moore 1978.

Intraethnic Institutions for Maintaining In-Group Cooperation

For G, we assumed that individuals observe (or learn about) all intraethnic interactions each period. This strong assumption was intended as a tractable stand-in for more plausible institutional mechanisms that groups may use to support cooperation, such as gossip and quasi-formal institutions like Milgrom, North, and Weingast's (1990) law merchant (see also Calvert 1995). Here we briefly consider the implications of introducing such institutions.

Assume atomized information, so that individuals observe what happens only in their own interactions. Add an additional player to each group, called the mediator. The mediator engages in no transactions, but every other individual in the group has the opportunity to appeal to the mediator after each interaction at a small personal cost, $c > 0$. For in-group interactions, if i appeals to the mediator of her group in period t, the mediator ascertains what happened in the interaction between i and j and awards a judgment of size $\gamma > 0$ (in utility) to the aggrieved party if one of the two cooperated while the other defected.[40] If the award is paid, then the mediator does nothing further; if the award is not paid, then the mediator communicates the name of the guilty party (say, j) to all other members of the group, and j is then socially sanctioned for T^{igp} periods. (The mediator may announce j's status in a public forum or spread the word via other notables.) Cooperation within the group will then be supportable, despite atomized information, if there is an award level γ large enough to deter defections but small enough such that it is worthwhile for defectors to pay the compensation, given δ and the punishment regime.[41]

What effect would such in-group mediators have on interethnic cooperation? On the plausible assumption that each group's mediator lacks detailed knowledge of and social connections in the other group, the same set of two cooperative options, spiral and in-group policing, will emerge. Suppose an A player, i, is cheated or exploited by a B player, j. Appealing to the A group's mediator is of no use if s/he cannot identify or locate j, or if s/he lacks the connections and authority to disseminate a negative ruling among group B. So, to support interethnic cooperation, either (1) the two groups coordinate on an in-group policing arrangement, wherein i appeals directly to the mediator of group B for compensation or goes to the A mediator, who then brings the

[40] The mediator makes no award if either CC or DD was played. The assumption that the mediator can discover what happened in the interaction (which is also made by Milgrom, North, and Weingast 1990) can be weakened so that the mediator may err.

[41] This brief statement does not address the issues of paying the mediator, which can be arranged in a variety of plausible ways; ensuring that mediators diffuse information about violations, which may require implicit threats to cease using the mediator if she does not do this; and ensuring that the mediator does not turn to extortion, which may be addressed by a competition among potential mediators, such that they have an incentive to cultivate a reputation for honesty or else lose respect and business (cf. Milgrom, North, and Weingast 1990; Calvert 1995). Note that this intraethnic institution is only one short step from a state-like policing apparatus; the only difference is that enforcement remains decentralized.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

case to the B mediator, or (2), failing this, the groups threaten a breakdown of relations or feud due to interethnic transgressions, to be announced and coordinated within A and B by the in-group mediators (that is, a spiral regime).

Distinguishing Types

The game G modeled interactions in which all individuals have an incentive opportunistically to exploit their partners. For many interactions, however, such as routine encounters in the street, it is more realistic to suppose that people vary in their disposition to behave opportunistically (or aggressively). Furthermore, it is reasonable to suppose that people are better at assessing the likely disposition of coethnics they have never met than of ethnic others they have never met, due to a lifetime of training in the cultural habits and signs of their own ethnic group. The strategic consequences of this differential ability to sort among types can be usefully explored in a variant on the game G.

Suppose there are two types in both groups—"good" types, who gain $\alpha_g < 1$ for defecting when their partner cooperates, and "bad" or opportunistic types, whose value for defecting on a cooperator is $\alpha_b > 1$. Otherwise, both types' payoffs are as in Figure 1. Assume further that in in-group interactions players observe the type of their partner in each period prior to choosing strategies, while in interethnic interactions they do not observe their partner's type[42]; that the proportion of bad types is the same in both groups, a small fraction $q > 0$; and that individuals observe only what happens in their own interactions (atomized information).

Supporting cooperation among the good types within ethnic groups is then a simple matter, even under atomized information. The good types cooperate if matched with a good type and defect if matched with a bad type, yielding them an average in-group payoff of $1 - q$. (Bad types defect against all partners.) The problem arises with interethnic interactions. If the proportion of bad types is believed to be high enough (if $q > (1 - \alpha_g)/(1 - \alpha_g + \beta)$), then good types will prefer to defect in interethnic interactions to protect themselves against the risk that they are facing a bad type, making the payoff for all interethnic interactions zero in equilibrium. Thus, we have a very different argument for why cooperation may be more difficult to sustain in interthan in intraethnic interactions, one based not on asymmetric information resulting from different social networks, but on differential abilities to distinguish types inside versus outside the ethnic group.[43]

There appear to be two basic institutional approaches the groups may take to overcome this problem, short of constructing an effective, overarching police apparatus. To gain interethnic trust and cooperation, opportunists need to be induced to cooperate despite their inclination, which may be achieved—once again—either by threatening indiscriminate retaliation (the spiral regime) or by arranging for the in-group sanctioning of opportunists.

Groups of Unequal Size

If one group (say, A) is larger than the other, then the rate of interethnic interaction decreases for the larger group and increases for the smaller group. For example, if A has 10,000 members, B has 1,000, and there are 200 interethnic encounters per period, then the probability for an A of an interethnic encounter is $p_A = .02$, while it is $p_B = .20$ for a B (under uniform matching).

This implies that the threat of spiral punishment cannot be used effectively by small groups against large ones, since As who defect against Bs interact so rarely with Bs that the Bs threat of a breakdown of relations is not compelling (as in the case of equal-sized groups with small p, discussed above). By contrast, the threat of spiral punishment is highly credible and effective as an instrument by which large groups may induce cooperation by small groups, since a breakdown of relations is very costly for members of the small group who interact often with the members of the larger group. Thus, we would predict that small groups will be more likely to evolve in-group policing strategies to try to avoid the costs of group punishment, while the threat of indiscriminate punishments will be more typical of how large groups give members of small groups an incentive to cooperate. In addition, note that the efficiency gains of setting up an in-group policing regime are small for members of the large group since their rate of interethnic interaction is small. Thus, it may be difficult to provide incentives for the members of the larger group to cooperate with the smaller group, since the costs of setting up in-group policing may increase with group size, and members of the larger group lose little from conflict with the smaller group.

In addition, if the members of small groups tend to have relatively high interethnic interaction rates, then such groups may need to develop strong boundary policing institutions in order to maintain themselves as distinct groups. As seen in the previous section, high interethnic interaction rates tend to undermine a group's ability to maintain cooperation and harmony within the group by reducing the force of in-group sanctions. Small groups, our model suggests, will seek either to assimilate into the larger group's social control mechanisms or to develop strong boundary policing institutions and practices, some of which may intentionally limit the amount of interethnic interaction in which their members may engage.[44]

[42] We could add a (more realistic) stochastic element to peoples' observation of type, but this gains little insight.

[43] Cornell (1995) suggests essentially this argument informally. Note that this mechanism may help explain why individuals of the same ethnic group may be more inclined to trust one another even if the group is very large and they have no common network connections.

[44] Of course, the larger group may also attempt to police its boundaries by resisting assimilation, either for the reason identified in the model or (perhaps more likely) to protect low-status members of the group who are threatened by competition from the minority. See Laitin 1995a for an analysis of boundary policing and marginal groups.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

EMPIRICAL ILLUSTRATIONS AND ASSESSMENT

The model and theoretical analysis presented above readily explain two important empirical aspects of interethnic relations that appear in a very broad range of cultural contexts. First, many have observed that interactions between individuals from different ethnic groups are marked by a distinct tension and lack of trust, which we explain as a strategic consequence of problems of asymmetric information due to the lower density of social networks across groups and to the differential ability to distinguish types inside versus outside the group. Second, as we argued earlier, despite greater tensions, peaceful interethnic relations are by far more common than violence or complete avoidance, which we explain as a result of institutional mechanisms that moderate cross-group problems of opportunism and so allow people to avoid the costs of violence and capture the benefits of peace. Existing rationalist and psychological theories of ethnic conflict, which argue that violence results from shifts in relative power, status, or access to resources, cannot so easily explain these facts. For example, it is not clear on either account why relations among *individuals* from different ethnic groups would be marked by any unusual tension or lack of trust. Furthermore, these accounts cannot help but be somewhat ad hoc in explaining how interethnic relations can be simultaneously tense and peaceful almost all the time.

Nonetheless, while our theory is consistent with these empirical regularities, there surely are plausible alternative factors involved as well. For instance, simple prejudice may account for much of the tension observed in interethnic interactions, and the activities of (centralized) state institutions may be crucial for explaining why interethnic peace is the modal outcome in many places. Our claim is not that the mechanisms we have identified are the only ones that matter, but that they have not been clearly identified before and do explain a part of the empirical puzzle.

If the latter part of this claim is correct, then we should be able to find empirical examples in which the causal mechanisms we have identified appear and matter as predicted. Specifically, we should be able to find evidence that asymmetric information creates problems of opportunism in interethnic interactions and that institutional mechanisms mimicking or paralleling in-group policing and spiral equilibria have evolved to deal with them. We consider a range of such examples in this section. While the examples support our core theoretical claims, at the same time they suggest that the diversity of institutional responses to the problem we identify is greater than that captured by our model.

Asymmetric Information and Tension

A number of observers of interethnic relations have noticed that the lack of information about ethnic others as individuals is connected to interethnic tension. Boehm (1994), in his ethno-history of Montenegrin feuds, distinguishes between Christian-Muslim disputes, which were nearly impossible to control, and feuds among Christian clans. He reports that headhunting, the most dreadful insult possible in Montenegrin culture, was normal in wartime. It was most likely done "against Moslems who lived at a considerable distance, where the individuals would not know exactly who had killed whom in the engagement" (p. 91). This tactic was not ordinarily used in blood feuds, since families knew each other. Not knowing the other as an individual, Boehm implies, loosens the constraints on "defection" and also allows for more indiscriminate retaliation. Similarly, in her study of Albanian blood feuds, Hasluck (1954, 226) notes: "If an Albanian was killed by a Slav, any Albanian would kill any Slav in revenge," whereas within Albanian settlements, more focused retaliation was both possible and enjoined by norms.

Particularly insightful on this issue is Rieder's (1985) study of the Jewish-Italian alliance against African Americans in Canarsie. When Jews and Italians could not recognize African Americans as individuals, or even as members of a social class, it was very difficult to develop cooperative relations with them.

> The ambiguous position of the new blacks deprived white residents of adequate strategies for promoting their own well-being, deciphering the environment, and appraising risk. More than a tiresome slight, the notion that one could not tell blacks apart bespeaks the intricate way urban dwellers scan their environment, sift visual and other clues for forecasts of danger, and make guesses about the intentions of strangers.... However imperfect or superstitious their basis, predictive clues reduce the feeling of risk, and they are often more than consoling fictions. The signs that elicited alarm included race, but not indiscriminately. Young black males, especially in groups, coiffed with Afros or sporting sneakers ("felony footware") triggered apprehension (Rieder 1985, 86, 177).

In this example as in several others given below, the ethnographer links the specific tensions that mark interethnic relations to an inability to distinguish among individuals and types of people in the other group.

The Evolution of Institutional Mechanisms

When it is not possible to identify the guilty party as an individual, one feasible strategy for policing cooperation is collective liability, or group punishment. All members of the other ethnic group are held equally responsible for the actions of any one of them. If a member of one group is murdered, for example, the other group seeks retribution by killing some random person of the other group. To a considerable extent, this is the way the Israeli Defense Forces handled Palestinian raids inside the Green Line for many years, and it also is the way that Palestinian nationalists targeted Israelis after what were perceived as land grabs. It was not necessary for either party to identify the person on the other side who was guilty of the transgression, only to strike back at some set of members of the other group with equal or greater damage. While this mechanism addresses the problem of low information, it fosters spirals of violence, as we discussed more formally above, and it also may systematically enrage a wider circle of innocent victims. Indeed, the solution of Israeli Labor Party leaders, which con-

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

tinues as the declared policy of Prime Minister Netanyahu, looks very much like in-group policing[45]—the Israeli government grants some autonomy to the Palestinian Authority on the condition that Arafat develop effective in-group policing.

Since indiscriminate retribution is not efficient in the presence of noise, we may find evidence of the evolution of other, more efficient institutional arrangements. One problem with searching for such mechanisms is that anthropologists are in the best position to study localized conflict management, but few of them have focused on inter- as opposed to intraethnic conflict resolution. Thus, our examples mainly were found at the margins rather than at the center of anthropological studies. Another problem is that many of the examples of in-group policing developed under imperial authority, so we cannot rule out the possibility that the inducements of a central state are necessary for their smooth operation. Nonetheless, these examples demonstrate that in-group policing plays an important role in cauterizing conflict when members of different groups have low information about one another as individuals.

In some cases, in-group policing is provided not by formal arrangement between groups but as a private strategy of individuals to facilitate trade across ethnic boundaries. Craig R. Whitney reported an excellent example in the *New York Times* (June 16, 1995). In Cocinelle, France, Miro Rizvic, manager of a grocery store, adjusted as follows to the problem of low information under conditions of ethnic heterogeneity that plagued other store owners: "Most of the crimes are committed by young North African kids. . . . [So] I hire local people, a Moroccan for the vegetables, a Tunisian for the canned goods, a Turk for the cash register. That way they know just about everybody who comes in here. It's better than hiring security guards."[46] (Note also how this example supports the claim that information asymmetries create problems of opportunism in interethnic relations.)

Clifford Geertz (1973, 7–9) gives a related example of a merchant who made private arrangements to overcome his inability to recognize individuals within a different ethnic group. In this case, the merchant contracted with agents specializing in information and purchased insurance from them in order to identify predators. Geertz relates that in early colonial Morocco, a marauding band of Berbers attacked the home of a Jewish trader in the Maghrib named Cohen. He survived but his guests were killed and his goods stolen. Cohen could get no help from the French authorities, but he belonged to a *mezrag*, or trade-pact system, and he went to his insurance broker, a tribal sheikh, to demand the assistance due. The sheikh knew precisely who had Cohen's merchandise, accompanied him in a climb up the Atlas directly to the shepherd of the thief's tribe, and took control of the entire herd. The tribal warriors soon returned, saw what had transpired, and prepared to attack. But then they saw Cohen and his insurance agent, a palaver began, and Cohen peacefully regained his goods at the precise insured value. Note that "on the equilibrium path" this institutional innovation of tribal "information brokers" would make mutually beneficial trade relationships between Jews and Berbers possible, despite problems of opportunism due to a low density of social network relations. And, in the case Geertz relates, the institution also prevented spiraling, here understood as a total breakdown of trading and relations between Jews and Berbers.

Somewhat more institutionalized is the arrangement that operated until recently among Armenians and Azeris living in (what were) mixed regions in Armenia and Azerbaijan. There was a formal relationship of "friendship" (*kirva* in Armenian), often initiated by inviting the potential "friend" from the other group to a child's circumcision ceremony; an Armenian in one village and an Azeri in another agreed to aid the other in any business or other dealings with members of their own ethnic group. This help might involve settling disputes, providing information about people with whom to trade, or facilitating connections in the "foreign" community.[47] Presumably, an outsider with such a friend was less likely to be cheated because that friend might be able to harm a defector's reputation within the group.

Examples of institutionalized in-group policing are abundant under conditions of imperial rule. In the millet system in the Ottoman Empire, each religious/nationality group (the core groups were Orthodox Christians, Armenians, and Jews) received a high degree of autonomy as long as their leaders successfully policed their own communities within their millet and monitored members of their own community in interactions with outsiders. Consider anti-Semitic incidents in Turkey in the second half of the 19th century. According to Paul Dumont, there was intense antagonism. Despite intermingling of Jews, Muslims, and Christians in "apparent harmony," Dumont states (1982, 222–3),

> the slightest spark sufficed to ignite the fuse. Whenever a young Christian disappeared at the approach of Passover, Jews were immediately accused of having kidnapped him to obtain blood necessary for the manufacture of unleavened bread. Threats and violence followed close behind the suspicions and generally things ended with a boycott of Jewish shops and peddlers.

This is an example of indiscriminate group punishment, and Dumont recounts many instances in which all Jews suffered for any supposed transgression against Christians. A more nuanced picture appears, however, in his discussion of a riot in 1874 in Urla, involving the disappearance of a young Greek girl. Although she "eventually reappeared, Jews were persecuted and boycotted for several weeks. Here the governor of the province was obliged to order Orthodox priests to preach peace and goodwill in their churches" (pp. 221–2). The message from the Ottoman authorities was clear: Monitor the defections of your own group or you will be replaced by someone else as millet leader. Thus,

[45] See, for example, N. R. Kleinfield, "Netanyahu Promotes Views on Peace and Investments," *New York Times*, 12 July 1996, sect. A.

[46] Whitney adds: "Mr. Rizvic said that he was now a French citizen, but that he, too, emigrated 20 year ago—from Bosnia."

[47] This information comes from personal communications with Sergei Arutiunov.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

the normal functioning of the system, according to which the leaders of each millet curbed excesses in violence against members of other millets, helped ensure that "while in Russia, Rumania, and most of the Balkan states, Jewish communities suffered from constant persecution (pogroms, anti-Jewish laws, and other vexations), Jews established on Turkish territory enjoyed an altogether remarkable atmosphere of tolerance and justice" (pp. 221–2).

In much of premodern Europe, state authorities relied heavily on in-group policing in Jewish affairs, institutionalized in kahals, or self-governing communities (Dubnow 1916, 105–6; Goldscheider and Zuckerman 1984, 18–24). Based upon a 1551 charter of Sigismund Augustus, for example, Jews in Poland were given the right to elect their own rabbis and "lawful judges." In these kahals, Jews received local autonomy in education, religious affairs, and internal conflict management. In return, the kahal leaders took full responsibility for the collection of government taxes. We have found fragmentary evidence that the kahals engaged in forms of in-group policing that went beyond identifying tax evaders. In 1672, to cite one example, the Polish king "bestowed upon the Kahal elders of Lithuania the right of excluding from the community or of punishing by other measures those recalcitrant members of their Kahals who by their acts were likely to arouse the resentment of the Christian population against the Jews" (Dubnow 1916, 190). This suggests that kahal leaders identified Jews whose acts hurt non-Jews—one example being the passing into general circulation of clipped or counterfeit coinage (Lederhendler 1989, 29)—and punished them.

After the abolition of the kahal in Russia in the mid-19th century, the czarist regime began to attach Russian-speaking Jewish advisers to various offices of the bureaucracy, to "impose greater order and direct supervision on the Jewish population" (Lederhendler 1989, 92), and they presumably played a key role in identifying lawbreakers to state authorities. One quasi-institutionalized role that transcended the kahal and post-kahal periods is that of *shtadlan*, a Hebrew term designating "public-spirited Jews who defend the interest of their coreligionists before the Government" (Dubnow 1916, 111). These *shtadlanim* regularly informed on members of their own community to state officials. In fact, Lederhendler emphasizes that "informing" on the treasonous behavior of fellow Jews was common practice from the kahal period onward (1989, 12–4, 28–9, 88–92). Reciprocally, state authorities policed the non-Jewish communities to prevent attacks on Jews, which largely contained pogroms for more than three centuries, despite persistent local anti-Semitism. It was not until 1881 and the chaos following the assasination of Czar Alexander II that hostilities against Jews went unchecked. In-group policing on both sides (though the Russian and Polish states were not motivated by a desire to prevent indiscriminate reprisals by Jews so much as by a desire to maintain the tax base) may help explain the long period of relative peace.

Another example of in-group policing comes from Yorubaland under British rule. When Hausa traders began dominating the kola and cattle trade in Ibadan, many interethnic conflicts arose that were quite difficult to manage and occasionally spiraled. The solution for the Yoruba government (under a watchful British imperial eye) was Hausa autonomy within carefully circumscribed neighborhoods (called *sabos*). As Cohen (1969, 21) describes the process,

> a number of Hausa commission agents would lodge temporarily with Yoruba house-owners, and then, as their number increased, and as they were joined by men from related occupations, they constituted a social problem within the polity of the settlement. The Yoruba population would complain that the Hausa were harbouring burglars, pickpockets and other categories of "undesirables" who endangered law and order in the community. To meet this danger, and also to meet the simultaneous clamour of the Hausa to live on their own . . . the local chiefs would decide to allot to the Hausa a special Quarter within the settlement and would recognize one of the Hausa men of influence . . . as chief. He would then be held responsible to them for the conduct of his people, inform the authorities about the undesirables in his community, and help in the collection of taxes.

It may be asked why Hausa chiefs in the sabos (or Jews in the millet and in the pale) would agree to inform higher authorities about the identity of their "defectors." Our theoretical analysis suggests an answer. Smaller ethnic groups within a larger society have a great percentage of interactions with outsiders—nearly all Hausa transactions in the kola and cattle trades were with Yorubas—and therefore suffer greatly from a breakdown of relations. Thus, they face strong incentives to set up effective in-group policing institutions.

As our theoretical analysis suggested, in-group policing may be accomplished in many instances through professional mediators who know one or both groups well and who specialize in extracting precise information from disputing parties to design finely calibrated compensation packages that prevent spiraling violence. Parallel to what Milgrom, North, and Weingast (1990) call law merchants, we find evidence that "mediation merchants" with a reputation for objectivity in interethnic disputes are hired by feuding parties.

An example is given by Samatar (1982, 36–54), who describes a war between a nomadic and a farming clan in eastern Ethiopia, while the state was quite content to see the clans decimating each other. The source of the problem was famine, which brought the nomads into areas traditionally rangelands but recently become farmland. One evening the herds of the nomads wandered into a maize field and damaged the crop. A skirmish followed, in which a nomad and a farmer were killed, and for three months young toughs of both clans roamed the regional city with rifles, shooting indiscriminately at members of the other group. This spiraling and indiscriminate punishment reached a climax with the murder of a nomadic elder. The nomads began to recruit an army from their wider kin, seeking to teach this "worthless scum of saucy slaves" (the nomads saw themselves as nobles and the farmers as former slaves) a lesson.

On the night the nomad force planned to attack, a group of nomad elders who had trade relations with the

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

farmers stood in front of the attackers and demanded a council to determine whether there should be war. The leader of the warriors gave a powerful speech, calling upon Islam and wounded pride and demanding the destruction of the farming community. But the leader of the peace group used a plea of common Islam and then appealed to a common Somali interest (shared by the farmers and nomads) against the Ethiopians, who were confiscating the farmers' cattle and the nomads' sheep and camels. This leader called for a delegation to work out a proper compensation scheme, identifying precisely who required redress. The peace platform won, and the conflict was, through negotiation and compensation, toned down. The settlement along the lines of in-group punishment—targeted compensation packages and the aid of coethnics in identifying who is liable within the group—was made possible by a mediation merchant, who was Samatar's father.

One way mediation merchants can matter, as suggested by Morrow (1992) but not by our model, is that they can present a solution which is agreeable to both sides but which the disputants may be unwilling to propose for fear of appearing "weak." In Montenegrin conflicts, Boehm (1994, 122) notes that "Courts of Good Men" were mediators whose members had sufficient prestige to settle feuds. They could be asked to intervene only after a third successful truce had been called. Customarily, the clan with the higher score (i.e., had caused more bloodshed than it suffered) initiated a truce, but this was seen as a sign of weakness, and a great deal of outside pressure or war-weariness usually was necessary before a first truce was achieved. In other words, mediators can propose solutions that participants themselves want but are reluctant to raise themselves, and for this very reason the combatants may be unwilling to propose calling upon mediators. Nonetheless, to the extent that mediation merchants are available for hire, heated interethnic tensions are more likely to cool. As in the case of insurance schemes, mediation merchants can fill in the information gaps in ways that address the problem identified in the in-group policing equilibrium.

We stipulated in our model that one reason interethnic conflict has a high propensity to spiral is that members of one group have low information about past behavior of members of out-groups. By lumping all outsiders together, it is more difficult in prisoner's dilemma matching games to restore cooperation when it breaks down. In this section we have shown that the problem has real-world analogues. We have also shown that various institutional mechanisms have evolved to ameliorate the potential consequences of this problem, mechanisms that mimic or parallel those found in our model.

CONCLUSION

To conclude, we will summarize how we have addressed the three explanatory puzzles that began the paper, and we then consider several ways that our model may be profitably extended.

Why are interethnic relations frequently characterized by a tension that is relatively absent in intraethnic relations, giving interethnic relations, even when peaceful, an ominous quality? Because social networks are better developed and interactions more frequent within ethnic groups, individuals have easier access to information about their coethnics than they do about ethnic others. Better information and more frequent interactions allow coethnics to develop and maintain *individual* reputations for cooperative behavior that are more difficult to sustain in interethnic interactions. In addition, due to cultural familiarity, people are better able to distinguish opportunists among coethnics, which facilitates peaceful interaction within and hampers it across ethnic groups.

Why in interethnic relations is cooperation the more likely outcome than escalating violence, despite the greater tensions in day-to-day affairs? Due to the costs of persistent violence and the various benefits of peaceful interethnic relations, decentralized institutional arrangements are likely to arise to moderate problems of interethnic opportunism.

Why in some cases do interethnic relations often remain cooperative for a long time yet periodically break down in the form of spiraling violence, while in other cases peace is quickly restored after violence breaks out? We have identified two basic institutional approaches to resolving problems of opportunism in interethnic interactions. In both, interethnic cooperation is sustained by the expectations people have about what will happen if they cheat or otherwise injure someone from another group. The content of these expectations differs markedly, however.[48] In one type of equilibrium, people expect that if cooperation breaks down due to an accident, mistake, or misinterpretation, then members of each group will punish each other indiscriminately for some length of time (long enough to deter defections in typical interactions). These expectations are self-fulfilling. We call this a spiral equilibrium because violence or noncooperation quickly expands beyond the initial parties to a dispute. It is indistinguishable from the other cooperative equilibrium when play remains on the equilibrium path, but it can be quite unpleasant off the equilibrium path.

We also identified an institutional arrangement that is more efficient than the spiral equilibrium in the presence of noise. Under in-group policing, people expect that someone who defects against an ethnic other will be identified and sanctioned by members of his or her own group. Again, these expectations are self-fulfilling, although in practice institutional arrangements that make the punishment visible to the offended group may be necessary.[49] Examples of the evolution of institutions that resemble in-group policing mechanisms demonstrate the feasibility of this sort of institutional arrangement.

How can the model presented here be developed and extended? We have focused upon decentralized mechanisms that delimit interethnic violence and have claimed that these can work in the absence of state authority.

[48] Greif (1994) argues more generally that "culture" may be productively understood as shared beliefs about what happens off the equilibrium path.

[49] The establishment of a formal office to see to this—such as the *shtadlan* and the chief of the Hausa quarter—may be interpreted in these terms.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

Nonetheless, the central state appears in the background in some of the empirical examples we gave, and its role may be significant. In fact, in our vignettes of interethnic relations in three empires (the Ottoman, Russian, and British), systems of in-group policing were certainly related to indirect rule, which also may have made them possible. Further development of our informational approach would require fuller consideration of the state's role in both cauterizing and fostering interethnic violence.

One approach in the context of the present model is to think of the state as a player who will intervene only if violence reaches a certain level. African Americans in Cabrini Green in Chicago know that gang wars within their housing project can escalate up to some specific level before the city authorities intervene. Of course, if a gang were to terrorize a nearby wealthy neighborhood in Lincoln Park, the threshold would be lower. In the race riots in Los Angeles after the Rodney King verdict, police quickly moved into areas in which the state often *seems* absent. The threat of state intervention was explicit in the case of the Jewish near-pogrom in the Ottoman empire. Members of conflicting groups, then, make an assessment of the level of violence necessary to bring in central authorities and may keep below this level. In these cases the state, apparently absent, sets an upper bound on ethnic violence at a level that may prevent spirals and conceivably encourages in-group policing to avoid state intervention. In these cases, however, the state is only apparently absent because its interventions occur off the equilibrium path.

A second extension of the argument is to consider more fully the institutional implications of in-group policing. There are a number of points to explore. First, while we have argued that the quality of in-group policing arrangements may crucially affect the amount of interethnic violence, we have not elaborated a theory stating when effective in-group policing is most likely to arise (beyond the argument that relatively small groups are more likely to develop such arrangements). For that, we need to examine issues of collective action and political competition within ethnic groups. We then may better describe the conditions under which interethnic relations are most likely to remain peaceful, even off the equilibrium path. Second, we need to consider more fully the social and political consequences of in-group policing equilibria. We know that over time the millet and kahal systems broke down, with awful consequences for all groups but especially for minorities. *It is crucial that both spiral equilibria and in-group policing will tend to reproduce and maintain the sense of ethnic difference through time.* In addition, in-group policing may have the added liability that the same in-group institutions that prevent spiraling may be captured by ethnic entrepreneurs with an interest in fomenting ethnic violence and used by them to mobilize ethnic groups for conflict. The same in-group institutional structures used to identify and sanction members who offend against the other group can just as well be used to sanction members who seek to establish ties with ethnic others or who refuse to fight against them. While our analysis suggests that in-group policing is generally preferable to spiraling equilibria for maintaining interethnic cooperation, we are not recommending in-group policing as the best of all possible worlds. Finally, the risk of breakdowns in ethnic relations due to "noise" in individual interactions may be greater under in-group policing than under spiral equilibria, since the consequences can be much worse under the latter (i.e., all hell can break loose). Each of these considerations requires us to examine more fully in-group policing as a formal institution of governance.

How may an empirical research program develop along the lines indicated by the model? Our empirical section illustrated the plausibility of our assumptions and the existence of two distinct equilibria to address the problem of cauterizing interethnic violence once cooperation breaks down. With our model now available, it should be possible to devise more rigorous tests of its observable implications. First, our theoretical results suggest that the in-group policing equilibrium will have relatively greater cauterizing capacities compared to the spiral equilibrium when interethnic relations move off the equilibrium path. Second, our extension of the model in the case of uneven group size suggests that smaller groups are more likely to develop and more carefully monitor in-group policing schemes. Third, there is a rationale and interest for leaders of ethnic groups to limit interaction between coethnics and other groups, that is, to construct boundaries. Fourth, the problem of asymmetric information between groups opens up a market for local mediators who have specialized knowledge of the members (and personality types) in their group. More rigorous tests of these hypotheses remain on the agenda for future research.

Existing rationalist and psychological theories of ethnic conflict are premised on assumptions about group-level demands, grievances, and animosities. They tend to treat groups as actors or implicitly equate group motivation with that of representative members. In sharp contrast, we have attempted to develop a theory of interethnic conflict *and* cooperation that does neither. Instead, we have taken individual interactions subject to opportunism as our theoretical point of departure, showing how these can (1) provide a rationale for intraethnic institutions, some of which may play a role in maintaining "ethnicity" itself, and (2) create frictions and tension in interethnic interactions that nonetheless may be contained most of the time by two ideal types of equilibrium institutional arrangements. As stated in the introduction, we wish to stress that we do not view the approach developed here as a complete theory of ethnic cooperation and conflict. For example, we believe the more standard group-level analyses capture important aspects of the subject that our analysis, in its present form, does not. Furthermore, it is obvious that interethnic relations frequently involve powerful emotions which interact in complex ways with rational calculation. The rationalist analysis pursued here cannot easily address the role of affect in interethnic affairs. We offer the individual-interactions theory as a novel and potentially productive approach to understanding interethnic peace and violence, but not as the last word on the subject.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

APPENDIX: EQUILIBRIUM CONDITIONS, PROOFS, AND EXPECTED PAYOFFS WITH NOISY INTERACTIONS

PROPOSITION 1: *Let* $\underline{\delta}_1 = \dfrac{\beta}{(1-p)(1+\beta)}$, $\underline{\delta}_2 = \dfrac{\alpha-1}{(1-p)(1+\beta)}$, *and* $\underline{\delta}_{igp} = \max\{\underline{\delta}_1, \underline{\delta}_2\}$. *Then,* σ_{igp} *forms a subgame perfect equilibrium of* G *if and only if* $p < \min\left\{\dfrac{1}{1+\beta}, 1 - \dfrac{\alpha-1}{1+\beta}\right\}$, *and* $\delta^{T^{igp}} \geq \underline{\delta}_{igp}$. *These conditions further imply both that* $\underline{\delta}_{igp}$ *is the smallest discount factor that can support* σ_{igp} *as a subgame perfect equilibrium and that if* σ_{igp} *is subgame perfect for given* δ, p, *and other parameters, then it is always possible to choose* $T^{igp} = 1$.

Proof. Consider i, a member of group A (which is sufficient since all the same arguments will apply to members of group B as well). Let the vector $s_t = (t_1, t_2, \ldots, t_n)$ summarize the state of the system at the beginning of period t, where t_i gives the number of periods remaining in player i's punishment phase, with the convention that $t_i = 0$ means that i is not in punishment phase. We will say that a player not in punishment phase is a "cooperator" and a player being punished is a "defector." Since σ_{igp} conditions only on the status of a player and her partner in each round (i.e., in-group/out-group, cooperator or defector), to establish subgame perfection (SGP) it will be sufficient to show that under the conditions in the proposition no player has an incentive to deviate from σ_{igp} in a single period for any state s_t (the optimality principle of dynamic programming is invoked here; see Fudenberg and Tirole 1991, 108–10, for a statement of the theorem and Milgrom, North, and Weingast 1990 for an application in a social matching game).

In particular, this means that we need to establish that for any s_t (1) a cooperator i has no incentive (a) to defect against an out-group player, (b) to defect against an in-group cooperator, and (c) to cooperate with an in-group defector; and (2) a defector i has no incentive (a) to defect against an out-group player, (b) to defect against an in-group cooperator, or (c) to cooperate against an in-group defector.

It is immediately evident that 1c and 2c will be satisfied for any s_t, any T^{igp}, and any δ, since in each case the proposed deviation strictly lowers the player's payoff in round t and then has no subsequent effects on any other player's strategy (these deviations are ignored). Furthermore, notice that a player has no incentive to deviate in case 1a if and only if she has no incentive to deviate in case 1b, since she faces exactly the same payoffs in period t and the same (equilibrium) profile of subsequent punishment in either case. So there are three conditions to check, 1a/1b, 2a, and 2b (for all states s_t).

Fix an s_t, and for integers $l \geq 0$ let n_{t+l} be the number of A-group players $j \neq i$ with $t_j \leq l$ (that is, the number of A players other than i who will not be in punishment phase in period $t + l$, if all follow σ_{igp}). Let $q_{t+l} = n_{t+l}/(n-1)$, which is the probability that player i is paired with a cooperator in period $t + l$ if i is paired in-group.

To check condition 1a/1b, suppose a cooperator is paired with a cooperator or an out-group player in period t, when the state is s_t. Then, by defecting, player i gains $\alpha - 1$ in t while forgoing, in expected terms,

$$\sum_{l=1}^{T^{igp}} \delta^l(1-p)(q_{t+l}(1+\beta) + (1-q_{t+l})\alpha). \qquad \textbf{(A-1)}$$

Note that $1 + \beta = 1 - (-\beta)$ is what i has forgone if i is paired with a cooperator, and $\alpha = \alpha - 0$ is what i has forgone if i is paired with a defector.

SGP requires that $\alpha - 1$ be less than or equal to expression A-1 for *any* feasible state s_t. This means that SGP will hold only if $\alpha - 1$ is weakly less than this expression for the feasible s_t that implies the sequence $(q_{t+1}, q_{t+2}, \ldots)$ that minimizes expression A-1. Clearly, if $\alpha > 1 + \beta$ then expression A-1 is minimized by $q_{t+l} = 1$ for $1 \leq l \leq T^{igp}$, which is feasible since it occurs when all A players are cooperators in period $t - 1$. Thus, if $\alpha > 1 + \beta$, then SGP requires

$$\delta\frac{1-\delta^{T^{igp}}}{1-\delta} \geq \frac{\alpha-1}{(1-p)(1+\beta)},$$

or

$$\delta^{T^{igp}} \leq 1 - \underline{\delta}_2\frac{1-\delta}{\delta}. \qquad \textbf{(A-2)}$$

If $\alpha \leq 1 + \beta$, however, then expression A-1 is minimized by $q_{t+l} = 0$ for $1 \leq l \leq T^{igp} - 1$ and $q_{t+T^{igp}} = 1$. The following argument justifies this claim: For all states s_t, it must be that $q_{t+T^{igp}} = 1$, since if all players $j \neq i$ play according to σ_{igp} from period t forward (as i assumes when deciding whether to deviate in period t), then no $j \neq i$ can be in punishment phase in period $t + T^{igp}$. Thus, the minimum value for expression A-1 occurs in the (feasible) state s_t, where all A players were cooperators in period $t - 2$, and all except i defected in period $t - 1$, which renders $q_{t+l} = 0$ for $1 \leq l \leq T^{igp} - 1$.[50] Thus, in this case SGP requires that

$$\alpha - 1 \leq \left(\sum_{l=1}^{T^{igp}-1} \delta^l(1-p)\alpha\right) + \delta^{T^{igp}}(1-p)(1+\beta). \qquad \textbf{(A-3)}$$

To check 2a, suppose a defector i (thus, $0 < t_i \leq T^{igp}$) is paired with an out-group player in period t. By defecting, he gains $\alpha - 1$ while forgoing

$$\sum_{l=1}^{t_i-1} \delta^l(1-p)(q_{t+l}0 + (1-q_{t+l})0) + \sum_{l=t_i}^{T^{igp}} \delta^l(1-p)(q_{t+l}(1+\beta) + (1-q_{t+l})\alpha).$$

For any s_t (less the component t_i), this expression is minimized by $t_i = T^{igp}$, leaving the requirement that for all s_t

$$\alpha - 1 \leq (1-p)\delta^{T^{igp}}(q_{t+T^{igp}}(1+\beta) + (1-q_{t+T^{igp}})\alpha).$$

But for any s_t, $q_{t+T^{igp}} = 1$, since if all play according to σ_{igp}, then all players $j \neq i$ are cooperators T^{igp} periods hence. Thus, SGP requires that $\delta^{T^{igp}} \geq \dfrac{\alpha-1}{(1-p)(1+\beta)}$.

To check 2b, suppose a defector i is paired with an in-group cooperator in period t. She gains β by defecting, while forgoing

$$\sum_{l=1}^{t_i-1} \delta^l 0 + \sum_{l=t_i}^{T^{igp}} \delta^l(1-p)(q_{t+l}(1+\beta) + (1-q_{t+l})\alpha).$$

For any s_t (less t_i) this expression is minimized by $t_i = T^{igp}$, which yields

$$\beta \leq \delta^{T^{igp}}(1-p)(q_{t+T^{igp}}(1+\beta) + (1-q_{t+T^{igp}})\alpha).$$

And since $q_{t+T^{igp}} = 1$, we have the condition $\delta^{T^{igp}} \geq \dfrac{\beta}{(1-p)(1+\beta)}$. Combining this with the condition for 2a yields

$$\delta^{T^{igp}} \geq \max\left\{\frac{\beta}{(1-p)(1+\beta)}, \frac{\alpha-1}{(1-p)(1+\beta)}\right\} = \max\{\underline{\delta}_1, \underline{\delta}_2\}. \qquad \textbf{(A-4)}$$

Considering expressions A-2, A-3, and A-4, there are now two cases to consider. In the first, if $\alpha \leq 1 + \beta$, then SGP requires that $\delta^{T^{igp}} \geq \underline{\delta}_1$ and expression A-3 hold. Since $\delta^{T^{igp}}$ is decreasing in T^{igp}, $\delta \geq \underline{\delta}_1$ is necessary for the former to be satisfied. It also proves to be sufficient to guarantee expression A-3 as well, since if we let $T^{igp} = 1$, then we can rearrange expression A-3 to yield $\delta \geq \dfrac{\alpha-1}{(1-p)(1+\beta)} = \underline{\delta}_2$, which is assured by $\delta \geq \underline{\delta}_1$, since $\alpha \leq 1 + \beta$ implies $\underline{\delta}_1 \geq \underline{\delta}_2$.

In the second case, where $\alpha > 1 + \beta$, SGP requires that

$$\underline{\delta}_2 \leq \delta^{T^{igp}} \leq 1 - \underline{\delta}_2\frac{1-\delta}{\delta}.$$

An analogous argument applies, and the conditions given in the proposition follow immediately. (The condition on p is necessary in order for the larger of $\underline{\delta}_1$ and $\underline{\delta}_2$ to be less than one.) *Q.E.D.*

[50] Note that under σ_{igp}, simultaneous defections are not ignored.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

PROPOSITION 2: *(i) For cases in which* $\alpha \leq 1 + \beta$, *let* $\underline{\delta}_S = \max\left\{\underline{\delta}_1, \frac{\alpha - 1}{\alpha - 1 + p}\right\}$. *Then* σ_S *is subgame perfect if and only if* $\frac{(\alpha - 1)(1 - \delta)}{\delta} \leq p < \frac{1}{1 + \beta}$, $\delta \geq \underline{\delta}_S$, $\delta^{T^{in}} \geq \underline{\delta}_1$, *and* $T^{out} \geq \ln\left(1 - \frac{\alpha - 1}{p}\frac{1 - \delta}{\delta}\right)/\ln \delta$. *(ii) For cases in which* $\alpha > 1 + \beta$, *let* T^{in}_{min} *be the smallest integer greater than or equal to*

$$\frac{\ln \underline{\delta}_1}{\ln\left(\frac{\underline{\delta}_2}{\underline{\delta}_2 + 1 - \underline{\delta}_1}\right)},$$

and let $\hat{\underline{\delta}} = \underline{\delta}_1^{1/T^{in}}$. *Then* σ_S *forms a subgame perfect equilibrium of* G *if and only if*

$$\frac{(\alpha - 1)(1 - \delta)}{\delta} \leq p < \frac{1}{1 + \beta}, \quad \delta \geq \max\left\{\hat{\underline{\delta}}, \frac{\alpha - 1}{\alpha - 1 + p}\right\},$$

$$T^{in} \geq T^{in}_{min}, \text{ and } T^{out} \geq \ln\left(1 - \frac{\alpha - 1}{p}\frac{1 - \delta}{\delta}\right)/\ln \delta.$$

In case (i), T^{in} *can be chosen to equal one, while* T^{in}_{min} *is greater than one in case (ii).*

Proof. Consider a member i of group A (again, the same arguments apply for members of B). Let $s_t = (t_0, t_1, t_2, \ldots, t_n)$ represent the state of the system as before, except that t_0 is the number of periods remaining in the out-group punishment phase at the end of period $t - 1$, where $t_0 = 0$ means that neither group is being punished. Parallel to the in-group policing case above, to establish subgame perfection we need to show that for any state s_t, (1) a cooperator i has no incentive (a) to defect in an out-group pairing if $t_0 = 0$, (b) to cooperate in an out-group pairing if $t_0 > 0$, (c) to defect in an in-group pairing with a cooperator if $t_0 = 0$ and if $t_0 > 0$, and (d) to cooperate with an in-group defector when $t_0 = 0$ and $t_0 > 0$; and (2) a defector i has no incentive (a) to defect in an out-group pairing if $t_0 = 0$, (b) to cooperate in an out-group pairing if $t_0 > 0$, (c) to defect in an in-group pairing with a cooperator when $t_0 = 0$ and when $t_0 > 0$, and (d) to cooperate with an in-group defector when $t_0 = 0$ and when $t_0 > 0$.

In cases 1b, 1d, 2b, and 2d, the proposed deviation from σ_S lowers the deviator's payoff in period t but has no subsequent effects, so these are clearly not optimal for any s_t, T^{in}, T^{out}, and δ. We proceed to find conditions on δ, T^{in}, and T^{out} such that 1a, 1c, 2a, and 2c hold.

Define q_{t+l} exactly as before. For the case 1a, suppose a cooperator i is paired with a B player in a period t when $t_0 = 0$. Defecting yields a gain of $\alpha - 1$ in period t while i forgoes $\delta p(1 - 0) + \delta^2 p(1 - 0) + \ldots + \delta^{T^{out}} p(1 - 0)$, since under σ_S a deviation in an out-group pairing has no effect on a player's payoff in in-group pairings. This yields the condition

$$\delta \frac{1 - \delta^{T^{out}}}{1 - \delta} \geq \frac{\alpha - 1}{p}. \tag{A-5}$$

For case 1c, suppose a cooperator is paired in-group with another cooperator and with s_t such that $t_0 = 0$. Then he faces exactly the same expected payoffs for C versus D as in case 1b of the in-group policing proof above. Similarly, if $t_0 > 0$, then what is gained by defecting versus what is lost is the same as in case 1b of the in-group policing case. Thus, the conditions for SGP are given by inequality A-2 when $\alpha > 1 + \beta$, and inequality A-3 when $\alpha \leq 1 + \beta$, substituting T^{in} for T^{igp}.

For case 2a, where a defector i is paired with an out-group player when $t_0 = 0$, i faces exactly the same trade-offs as in case 1a (of the spiral equilibrium case), so we have expression A-5 again. Finally, for case 2c (a defector paired in-group with a cooperator when $t_0 = 0$ and when $t_0 > 0$), defecting yields a gain of β while i forgoes $\sum_{l=t_i}^{t_0 = n} \delta^l(1 - p)(q_{t+l}(1 + \beta) + (1 - q_{t+l})\alpha)$ whether $t_0 = 0$ or not. Thus, the logic of case 2b for σ_{igp} applies, yielding the condition

$$\delta^{T^{in}} \geq \frac{\beta}{(1 - p)(1 + \beta)}. \tag{A-6}$$

Putting these conditions together, there are now two cases. When $\alpha \leq 1 + \beta$, the conditions for δ and T^{in} are the same as for δ and T^{igp} in the parallel case at the end of the proof of proposition 1, with the additional restriction that $\frac{\delta}{1 - \delta} \geq \frac{\alpha - 1}{p}$, or $\delta \geq \frac{\alpha - 1}{\alpha - 1 + p}$, so that it is possible to deter defection in out-group pairings under the strongest possible punishment ($T^{out} = \infty$).

In the second case, when $\alpha > 1 + \beta$, SGP requires that

$$\underline{\delta}_1 \leq \delta^{T^{in}} \leq 1 - \underline{\delta}_2 \frac{1 - \delta}{\delta}. \tag{A-7}$$

Note that if $p < 1/(1 + \beta)$ (so that $\underline{\delta}_1 < 1$), we can guarantee that this will hold by choosing δ close enough to one and T^{in} large enough. The problem is to find the minimum δ such that it holds for some choice of T^{in}. Observe that $\delta \geq \underline{\delta}_1$ is necessary because $\delta^{T^{in}}$ decreases with T^{in}. But it is not sufficient, since $\delta = \underline{\delta}_1$ implies $\underline{\delta}_1 \leq 1 - \underline{\delta}_2 \frac{1 - \underline{\delta}_1}{\underline{\delta}_1}$, which reduces to $\underline{\delta}_2 \leq \underline{\delta}_1$, a false statement when $\alpha > 1 + \beta$.

The minimum δ in this case, call it $\hat{\underline{\delta}}$, must satisfy $\underline{\delta}_1 = \hat{\underline{\delta}}^{T^{in}}$ for some T^{in}, since otherwise it would be possible to lower $\hat{\underline{\delta}}$ without violating either inequality of expression A-7. Thus, $\hat{\underline{\delta}} = \underline{\delta}_1^{1/T^{in}}$, and SGP requires that we choose the minimum T^{in} such that

$$\underline{\delta}_1 \leq 1 - \underline{\delta}_2 \frac{1 - \underline{\delta}_1^{1/T^{in}}}{\underline{\delta}_1^{1/T^{in}}}. \tag{A-8}$$

Algebra can then be used to extract the conditions for δ, T^{in}, and T^{in}_{min} given in the proposition. The stated requirement on T^{out} derives from expression A-5 and is needed so that there is no incentive to defect in an out-group pairing, and the condition that $\delta \geq \frac{\alpha - 1}{\alpha - 1 + p}$ (equivalently, $p \geq \frac{(\alpha - 1)(1 - \delta)}{\delta}$) is necessary for there to exist such a T^{out}. The condition $p < \frac{1}{1 + \beta}$ is necessary for $\underline{\delta}_1 < 1$, which is in turn necessary for expression A-7 to be possible. *Q.E.D.*

Claim. When $\alpha \leq 1 + \beta$, σ_{igp} is less restrictive than σ_S in the sense that whenever σ_S forms a SGP equilibrium then so does σ_{igp}, while σ_{igp} may be SGP for parameter values such that σ_S is not. When $\alpha > 1 + \beta$, for given parameter values it may be possible to support both, neither, or only one of σ_{igp} and σ_S as SGP equilibria.

Proof. This follows immediately from propositions 1 and 2 in cases where $\alpha \leq 1 + \beta$, since the conditions for SGP are the same except that σ_S also requires that $\delta \geq \frac{\alpha - 1}{\alpha - 1 + p}$. For cases with $\alpha > 1 + \beta$, the claim is directly implied by the fact that the upper bound on p is larger under σ_S than under σ_{igp}, while the lower bound on p is smaller under σ_{igp} than under σ_S (see propositions 1 and 2). *Q.E.D.*

Expected Payoffs with Noise in Interethnic Interactions

We assume that the noise takes the following form. Players intending to cooperate "accidentally" defect with probability $\varepsilon > 0$ in interethnic pairings. Defection always produces defection. Note that under σ_{igp} and σ_S mutual defections in interethnic pairings are not ignored, so that DD triggers a spiral under σ_S and in-group punishment of both players under σ_{igp}. These assumptions are made simply to reduce the complexity of the calculations and are not crucial for the results reported in the text.

Noise in the spiral equilibrium. If σ_S forms an equilibrium given a level of noise $\varepsilon > 0$, then a player's ex ante expected equilibrium payoff, call it V^*_S, can be computed as follows. The expected payoff in period t if paired with an ethnic other is $B = (1 - \varepsilon)^2 + (\alpha - \beta)\varepsilon(1 - \varepsilon)$. V^*_S is then defined by the recursive equation

$$V^*_S = pB + (1 - p) + \delta\Big((1 - \varepsilon)^{2k} V^*_S + (1 - (1 - \varepsilon)^{2k}) \cdot \left[(1 - p)\frac{1 - \delta^{T^{out}}}{1 - \delta} + \delta^{T^{out}} V^*_S\right]\Big),$$

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

where $pB + (1 - p)$ is expected payoff in the current round, $(1 - \varepsilon)^{2k}$ is the probability that no one defects in an interethnic pairing, and the expression in brackets is expected payoff if cooperation with the out-group breaks down. This solves to

$$V_S^* = \frac{pB + (1 - p) + \delta(1 - p)\dfrac{1 - \delta^{T^{out}}}{1 - \delta}(1 - (1 - \varepsilon)^{2k})}{1 - \delta(\delta^{T^{out}} + (1 - \delta^{T^{out}})(1 - \varepsilon)^{2k})}.$$

For σ_S to remain an equilibrium with ε noise, it is necessary that a player's expected payoff for cooperating when paired with an ethnic other (call it $V_S|C$) be greater than or equal to the payoff for defecting against an ethnic other ($V_S|D$). $V_S|C$ and $V_S|D$ are computed as follows:

$$V_S|C = B + \delta\Bigg((1 - \varepsilon)^{2k}V_S^* + (1 - (1 - \varepsilon)^{2k}) \cdot \left[(1 - p)\frac{1 - \delta^{T^{out}}}{1 - \delta} + \delta^{T^{out}}V_S^*\right]\Bigg),$$

and

$$V_S|D = (1 - \varepsilon)\alpha + \delta\left((1 - p)\frac{1 - \delta^{T^{out}}}{1 - \delta} + \delta^{T^{out}}V_S^*\right)$$

Using a computer, one can then calculate the minimum T^{out} such that σ_S can form a Nash equilibrium given ε noise and other parameter values, and also the ex ante expected value of the equilibrium for a player, V_S^*.

For cases in which $\alpha - \beta < 1$, it is straightforward to show that Nash equilibrium and subgame perfection with $\varepsilon > 0$ require no conditions additional to the one just given and to those stated in proposition 2(i). Consider a player i paired in-group in period t. The trade-offs she faces for defecting versus cooperating are exactly those given for case 1c in the proof of proposition 2, since whether she defects in-group has no effect on what happens in subsequent (noise-affected) outgroup relations. That is, expected payoffs from outgroup interactions will be the same regardless of one's behavior within the group, so the in-group conditions for the case in which $\varepsilon = 0$ apply. Next, concerning out-group pairings and off-equilibrium-path states s_t, note that when $\alpha - \beta < 1$ we can always satisfy the conditions for in-group propriety by setting $T^{in} = 1$. Thus, if one is paired outgroup in period t, one expects that all coethnics will be in cooperation status in $t + 1$. This implies that regardless of the state s_t (but with $t_0 = 0$), $V_S|C$ and $V_S|D$ are exactly as given above. This demonstrates subgame perfection, since the remaining cases (1b, 1d, 2b, and 2d in the proof of proposition 2) are again trivially satisfied.

For the case of $1 < \alpha - \beta < 2$ (which is less empirically relevant when defection means violence, since one then would expect β to be much greater than α), T^{in}_{min} is greater than one, and the expressions for the expected utility of cooperating versus defecting with $\varepsilon > 0$ become extremely complicated. We have analyzed the case of $T^{in} = 2$, which proves to reduce (for any s_t with $t_0 = 0$) to the same condition as that given immediately above. We conjecture that the same would be true for any T^{in}, but we have not shown this.

Noise under in-group policing. Let $B = (1 - \varepsilon)^2 + (\alpha - \beta)\varepsilon(1 - \varepsilon)$ as before. For notational convenience, let $K = \dfrac{k}{n - 1}$ be the fraction of A-group players other than i who were paired out-group in the last period if i was paired in-group in that period, and let $K_{-1} = \dfrac{k - 1}{n - 1}$ be the fraction of A-group players other than i who were paired out-group in the last period if i was paired out-group. Suppose provisionally that $T^{igp} = 1$, which is always possible when σ_{igp} forms an equilibrium with $\varepsilon = 0$ (see proposition 1). Then player i's ex ante expected equilibrium payoff is $V^*_{igp} = pB + (1 - p) + \delta\dfrac{\hat{V}^*_{igp}}{1 - \delta}$, where $\hat{V}^*_{igp}$ is defined as

$$\hat{V}^*_{igp} = pB + (1 - p)[p\varepsilon(K_{-1}\varepsilon 0 + (1 - K_{-1}\varepsilon)(-\beta)) + p(1 - \varepsilon)(K_{-1}\varepsilon\alpha + 1 - K_{-1}\varepsilon) + (1 - p)(K\varepsilon\alpha + 1 - K\varepsilon)].$$

The expression within brackets gives probabilities and expected payoffs for three events that may have occurred in the prior period. There is a $p\varepsilon$ chance that i paired out-group and defected, a $p(1 - \varepsilon)$ chance that i paired out-group and cooperated, and a $1 - p$ chance that i paired in-group. Payoffs in each case depend on the probability of being paired with someone who defected in an out-group pairing in the prior period—for example, $K_{-1}\varepsilon$ is the probability that i meets a defector if i paired out-group in the last period.

Related expressions may then be derived for $V_{igp}|C$, i's expected payoff if i chooses to cooperate in an out-group pairing, and $V_{igp}|D$, i's expected payoff if i deliberately defects in an out-group pairing. These are

$$V_{igp}|C = B + \delta[pB + (1 - p)(-\varepsilon\beta(1 - K_{-1}\varepsilon) + (1 - \varepsilon)(K_{-1}\varepsilon\alpha + 1 - K_{-1}\varepsilon))] + \frac{\delta^2\hat{V}^*_{igp}}{1 - \delta}$$

and

$$V_{igp}|D = (1 - \varepsilon)\alpha + \delta[pB + (1 - p)(K_{-1}\varepsilon 0 + (1 - K_{-1}\varepsilon)(-\beta))] + \frac{\delta^2\hat{V}^*_{igp}}{1 - \delta}.$$

V^*_{igp} can then be computed as above, subject to the condition that $V_{igp}|C \geq V_{igp}|D$. In contrast to the spiral equilibrium with noise, however, in this case we need to revise the equilibrium conditions for in-group pairings, because under in-group policing noise in interethnic pairings affects the expected number of players to be punished *within* the group. Consider first the case of a cooperator paired in-group with another cooperator. Provisionally assuming that we can support an SGP equilibrium with $T^{igp} = 1$, cooperating yields

$$1 + \delta(pB + (1 - p)[K\varepsilon\alpha + 1 - K\varepsilon]) + \delta^2\frac{\hat{V}^*_{igp}}{1 - \delta},$$

while defecting yields

$$\alpha + \delta(pB + (1 - p)[K\varepsilon 0 + (1 - K\varepsilon)(-\beta)]) + \delta^2\frac{\hat{V}^*_{igp}}{1 - \delta}.$$

The former is greater than the latter provided that

$$\delta \geq \frac{\alpha - 1}{(1 - p)(K\varepsilon\alpha + (1 - K\varepsilon)(1 + \beta))},$$

which is virtually the same as $\underline{\delta}_2$ for small ε. A parallel argument establishes that a defector paired in-group with a cooperator has no incentive to defect provided that

$$\delta \geq \frac{\beta}{(1 - p)(K\varepsilon\alpha + (1 - K\varepsilon)(1 + \beta))},$$

which is approximately equal to $\underline{\delta}_1$ for small ε. Note that since $T^{igp} = 1$, these expressions apply regardless of the state of the system s_t, since if all follow σ_{igp} in period t, then all except those who accidentally defect in cross-group pairings will be in cooperation status in period $t + 1$. Thus, if these conditions and $V_{igp}|C \geq V_{igp}|D$ are satisfied, then σ_{igp} chosen by all players can form an SGP with $\varepsilon > 0$ noise in interethnic interactions. (The analogue to condition 2a in the proof of proposition 1 is satisfied when T^{igp} is set equal to one, since a defector in period $t - 1$ paired out-group in period t is then in the same situation as a cooperator paired out-group in period t. The analogues to conditions 1c and 2c are again trivially satisfied.)

REFERENCES

Axelrod, Robert. 1984. *Evolution of Cooperation*. New York: Basic Books.

Banton, Michael. 1983. *Racial and Ethnic Competition*. Cambridge: Cambridge University Press.

Barth, Fredrik, ed. 1969. *Ethnic Groups and Boundaries*. Boston: Little Brown.

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

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms

Boehm, Christopher. 1994. *Blood Revenge: The Anthropology of Feuding in Montenegro and Other Tribal Societies*. Lawrence: University Press of Kansas.
Bremmer, Ian, and Ray Taras. 1993. *Nation and Politics in the Soviet Successor States*. Cambridge: Cambridge University Press.
Calvert, Randall. 1995. "Rational Actors, Equilibrium, and Social Institutions." In *Explaining Social Institutions*, ed. Jack Knight and Itai Sened. Ann Arbor: University of Michigan Press.
Cohen, Abner. 1969. *Custom and Politics in Urban Africa*. Berkeley: University of California Press.
Colson, Elizabeth. 1974. *Tradition and Contract*. Chicago: Aldine.
Cornell, Bradford. 1995. "A Hypothesis on the Origins of Ethnic Discrimination." *Rationality and Society* 7 (January):4–30.
Deutsch, Karl. 1966. *The Nerves of Government*. New York: The Free Press.
Downs, George, and David Rocke. 1990. *Tacit Bargaining, Arms Races, and Arms Control*. Ann Arbor: University of Michigan.
Downs, George, and David Rocke. 1995. *Optimal Imperfection?* Princeton: Princeton University Press.
Dumont, Paul. 1982. "Jewish Communities in Turkey during the Last Decades of the 19th Century." In *Christians and Jews in the Ottoman Empire*, vol. 1, ed. Benjamin Braude and Bernard Lewis. New York: Holmes and Meier.
Dubnow, S. M. 1916. *History of the Jews in Russia and Poland*, vol. 1. Trans. I. Friedlaender. Philadelphia: The Jewish Publication Society of America.
Ellison, Glenn. 1994. "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching." *Review of Economic Studies* 61 (July):567–88.
Fearon, James D. 1994. "Ethnic War as a Commitment Problem." Paper presented at the 1994 Annual Meeting of the American Political Science Association, New York.
Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge, MA: MIT Press.
Gagnon, V. P. 1994/95. "Ethnic Nationalism and International Conflict: The Case of Serbia." *International Security* 19 (Winter):130–66.
Gambetta, Diego. 1993. *The Sicilian Mafia: The Business of Private Protection*. Cambridge, MA: Harvard University Press.
Geertz, Clifford. 1973. *The Interpretation of Cultures*. New York: Basic Books.
Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca, NY: Cornell University Press.
Gellner, Ernest. 1988. "Trust, Cohesion, and the Social Order." In *Trust: Making and Breaking Cooperative Relations*, ed. Diego Gambetta. London: Blackwell.
Gluckman, Max. 1955. *Custom and Conflict in Africa*. Glencoe, IL: Free Press.
Goldscheider, Calvin, and Alan S. Zuckerman. 1984. *The Transformation of the Jews*. Chicago: University of Chicago Press.
Green, Edward, and Rob Porter. 1984. "Noncooperative Collusion Under Imperfect Price Information." *Econometrica* 52 (January): 87–100.
Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *Journal of Economic History* 49 (December):857–82.
Greif, Avner. 1994. "Cultural Beliefs and the Organization of Society." *Journal of Political Economy* 102 (October):912–50.
Hardin, Russell. 1995. *One for All*. Princeton: Princeton University Press.
Hasluck, Margaret. 1954. *The Unwritten Law in Albania*. Cambridge: Cambridge University Press.
Hobbes, Thomas. 1985 [1651]. *Leviathan*. New York: Penguin Books.
Horowitz, Donald. 1985. *Ethnic Groups in Conflict*. Berkeley: University of California Press.
Kandori, Michihiro. 1992. "Social Norms and Community Enforcement." *Review of Economic Studies* 59 (January):63–80.
Kaplan, Robert D. 1993. *Balkan Ghosts*. New York: Vintage.
Knappert, Jan. 1965. "Language Problems of the New Nations of Africa." *African Quarterly* 5 (1):95–105.
Laitin, David. 1995a. "Marginality: A Microperspective." *Rationality and Society* 7 (January):31–57.
Laitin, David. 1995b. "National Revivals and Violence." *Archives Européennes de Sociologie* 36 (1):3–43.
Landa, Janet Tai. 1994. *Trust, Ethnicity, and Identity*. Ann Arbor: University of Michigan Press, 1994.
Lederhendler, Eli. 1989. *The Road to Modern Jewish Politics*. New York: Oxford University Press.
Lewis, I. M. 1961. *A Pastoral Democracy*. London: Oxford University Press.
Lohmann, Susanne. 1997. "A Theory of Domestic-International Linkage." *Journal of Conflict Resolution*. Forthcoming.
Milgrom, Paul, Douglass North, and Barry Weingast. 1990. "The Role of Institutions in the Revival of Trade." *Economics and Politics* 2 (March):1–23.
Miller, Willian Ian. 1990. *Bloodtaking and Peacemaking*. Chicago: University of Chicago Press.
Moore, Sally Falk. 1978. *Law as Process: An Anthropological Approach*. London: Routledge.
Morrison, Donald, Robert Mitchell, and John Paden. 1989. *Black Africa: A Comparative Handbook*. New York: Paragon House.
Morrow, James D. 1992. "Signaling Difficulties with Linkage in Crisis Bargaining." *International Studies Quarterly* 36 (June):153–72.
Moynihan, Daniel P. 1993. *Pandaemonium: Ethnicity in International Politics*. New York: Oxford University Press.
Olzak, Susan. 1992. *The Dynamics of Ethnic Competition and Conflict*. Stanford, CA: Stanford University Press.
Posen, Barry. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35 (Spring):27–47.
Putnam, Robert, with Robert Leonardi and Raffaella Y. Nanetti. 1993. *Making Democracy Work*. Princeton: Princeton University Press.
Rieder, Jonathan. 1985. *Canarsie*. Cambridge, MA: Harvard University Press.
Samatar, Said S. 1982. *Oral Poetry and Somali Nationalism*. Cambridge: Cambridge University Press.
Tajfel, Henri, ed. 1982. *Social Identity and Intergroup Relations*. Cambridge: Cambridge University Press.
Taylor, Michael. 1976. *Anarchy and Cooperation*. London: Wiley.
Wagner, R. Harrison. 1993. "The Causes of Peace." In *Stopping the Killing: How Civil Wars End*, ed. Roy Licklider. New York: New York University Press.

This content downloaded from 139.18.242.114 on Tue, 11 Oct 2016 10:22:52 UTC
All use subject to http://about.jstor.org/terms