Creation of Social Order in Ethnic Conflict*

Keisuke Nakao
Boston University
April 11, 2007

Abstract

This paper describes a mechanism of ethnic conflict and cooperation, in which the fear of indiscriminate vengeance induces mutual monitoring within the target group of communal violence. In our peer monitoring equilibrium, in-group policing and out-group conflict coexist, and the former is developed by the latter in order to suppress inter-ethnic transgressions. Our theory is in contrast to Fearon and Laitin’s (1996) theory of inter-ethnic cooperation which shows no theoretical linkage between the two forms of punishments. Using a social matching game with costly monitoring, we predict that the success of inter-ethnic cooperation hinges heavily on each group’s quality of in-group policing and that as a consequence, a group with lower quality of policing tends to have more frequent and longer disputes with other groups. Other comparative-statics analyses will also be discussed.

1 Introduction

In their seminal paper of inter-ethnic cooperation, Fearon and Laitin (1996) argue that one of the difficulties for sustaining cooperation between two ethnic groups lies in lack of information, especially in the identification problem about who misbehaved.\(^1\) According to their spiral equilibrium, because people in a group cannot identify a transgressor in the other group, inter-ethnic cooperation can be enforced only by punishing all the suspects, and an individual transgression leads to large-scale ethnic violence.\(^2\)

\(^{1}\)In their words, "at bottom, the problem is informational."

\(^{2}\)Similarly, Kandori (1992) and Ellison (1994) considered cooperation by community enforcement in an economy where each player does not have any information about interactions in which he is not involved (i.e., information is atomized). In their contagious equilibrium, cooperation is sustained by a series of vicarious punishments, which result in total breakdown of cooperation.
Although such a problem of individual identification should not be ignored, however, evidences from anthropology indicate that an inter-ethnic transgression triggers conflict for another reason. Precisely, it is reported from anarchic or weak-state societies that inter-ethnic conflict is sparked off for inducing each ethnic group to discipline its own transgressors. Under the threat of conflict, because people were scared of group-level violence, they were motivated for controlling their brethren not to misbehave against ethnic strangers. In other words, collective violence creates peer pressure or peer punishment in the target group that contributes to social orders in the larger population.

For example, Wilson (1983) reported that among Nyakyusa in Tanzania, a single across-village wrongdoing led to communal violence.

In a case of adultery the injured husband, together with his kinsmen, pursued and attempted to kill, or torture and kill, the adulterer: self-help was not only permitted but expected in this situation, and a man’s near kinsmen were obliged to assist him. Neighbours were not obliged to assist in executing vengeance, but they might be victims of it, for if the injured husband did not find the adulterer he might kill any village-mate of his enemy. Such an attack commonly led to war between the two villages (pp.149).

This event seems to exactly match Fearon and Laitin’s spiral equilibrium. However, this is not the end of the story. Because a wrongdoer was a potential danger to his neighbours, he was expected to leave his village in order to avoid the war: "thieves and adulterers were liable to be banished from a village just like witches and sorcerers, for they too brought misfortune on their fellows" (pp.150). It implies that communal violence between villages induced social ostracism of the wrongdoer from his village.

In addition, from medieval Iceland, it was reported by Miller (1990) that if a person was wronged by a stranger, the object of revenge did not have to be the actual wrongdoer; he simply had to be, in the avenger’s estimation, someone associated with the wrongdoer. The avenger could kill just any random person who was as distant as first cousins. "From a functional point of view, [the fact that the avenger’s victim need not be the actual wrongdoer] had the effect of inducing people who might be held accountable for each other’s actions to involve themselves in each other’s affairs" (pp.197). As a consequence, "[g]roup liability, it could be argued, thus rendered the feud or fear of 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" (pp.198). In other words, by collectively targeting the wrongdoer and his relatives *ex ante*, groups could create stronger internal controls which would be impossible by targeting just the wrongdoer. Other examples of such indiscriminate vengeance are reported from North America (Reid 1999:92-3), Corsica Island (Gould
These reports suggest that communal violence may happen between groups in order to create in-group punishments on transgressors. People may employ such indiscriminate vengeance, because there are two reasons that "group-level sanctions may be expected to outperform individual-level ones" (Levinson 2003:373). First, collective punishment may induce mutual in-group monitoring which can be much cheaper and more effective than monitoring from outside, and in-group monitoring may help to reduce misbehaviors. Peers are in a better position for monitoring coethnics than ethnic strangers, but such monitoring can be further strengthened by threats from outside. Second, because of the tight social connectedness in an ethnic group, peer punishment induced by external collective punishment can also be cheaper and more effective than outsiders’ individual punishment. Peers can impose various kinds of penalties on those who misbehave. For example, just peers’ social ostracism or boycott of business can be sufficient to discourage opportunistic transgressions. On the other hand, it is likely to be more difficult and costly for outside punishers to effectively threaten individual wrongdoers in a group because of the weak social tie between them. In short, peers are advantageous both in monitoring and in punishing than outside entities.

The above argument was not formally explained by the model of Fearon and Laitin (1996), who described two equilibria of in-group policing and of out-group conflicts with no theoretical linkage between the two forms of punishments. Thus, in this paper, we extend their model of inter-ethnic cooperation in a way that in-group policing and out-group conflict coexist and the former is induced under the threat of the latter. Namely, we provide a model of social matching game in which people within each ethnic group costly and imperfectly monitor peers’ behaviors. Our model is in contrast to Fearon and Laitin’s (1996) model in which full information of peers’ behaviors is assumed within each ethnic group. Moreover, in order to investigate the mechanism of inter-ethnic cooperation without centralized institutional arrangements, we develop the model in the following two dimensions: i) there is cultural disparity between groups which generates a possibility of misinterpreting actions of members in another group; ii) in the presence of such "noise", each group has a degree of tolerance allowing for ethnic stranger’s observed misbehaviors as misinterpretation.

Main results are as follows. 1) Inter-ethnic cooperation may require enforcements both by in-group policing and by out-group conflict, and the former punishment is developed by the latter. 2) The success of inter-ethnic cooperation hinges heavily on each group’s quality of in-group policing. As a consequence, groups with high qualities

---

3 Among them, Colson (1974:41) reported of the Eskimo around Point Barrow, the fear of feud "influenced Eskimo behavior and encouraged the suppression of behavior that could lead to violence."

4 These points are also noted by Hardin (1995:118-9). "First, groups are apt to have better information about their members’ actions than about the actions of people in other groups. Second, groups are apt to have fairly straightforward reasons for imposing order on their own members if they are to be held responsible for their fellow members’ actions."
of in-group policing can successfully maintain long-lasting peace and cooperation, whereas those with low qualities of policing tend to have more frequent and longer disputes with other groups. 3) A dense network between groups may provoke ethnic conflict if two groups are culturally disparate, whereas it can facilitate cooperation between culturally close groups. 4) As a group tolerates observed misbehaviors by members in the other group and refrains from conflict, the risk of ethnic conflict can be reduced, but such tolerance may ignite inter-ethnic transgressions if the target group’s internal punishment is insufficient. In short, inter-ethnic tolerance does not necessarily encourage inter-ethnic cooperation, depending on each group’s quality of internal punishment. 5) In-group policing equilibrium and spiral one presented by Fearon and Laitin are two extreme cases of our ‘peer punishment equilibrium’. The former corresponds to ours with strong in-group punishment, in which inter-ethnic relation is characterized by mutual trust, while the latter corresponds to ours with weak punishment, characterized by mutual mistrust. We will describe evidence consistent with these results.

The rest of the paper proceeds as follows. Section 2 presents the model of inter-ethnic cooperation. Section 3 considers equilibria with and without peer monitoring. Section 4 demonstrates the necessity of peer monitoring in the presence of noise. Section 5 analyzes comparative statics, which is the highlight of this paper. Section 6 concludes.

2 Model of Inter-Ethnic Cooperation

2.1 Social Matching Game

This section provides a model of inter-ethnic cooperation. The model employs a social matching game. Suppose that there are two ethnic groups A and B, each of which consists of n individuals. These groups are represented by sets A = \{1, 2, \ldots, n\} and B = \{n+1, n+2, \ldots, 2n\}, where each individual is indexed by i.\(^5\) All the individuals are homogeneous in actions they take and payoffs they obtain. In successive periods \(t = 1, 2, \ldots\), each individual \(i\) is randomly matched with an opponent and plays the stage game of prisoner’s dilemma, in which each individual simultaneously and independently selects to cooperate or to defect \(a_{i,t} \in \{C, D\}\). One can imagine that people roam around a market, encounter one another and trade goods or services. The action of cooperation means to fulfill duties as they agreed, while the action of defection denotes any sort of opportunistic behaviors such as cheat, fraud, steal, robbery or malfeasance.

\(^5\)A number of scholars (e.g., Hardin 1995; Bowen 1996; Fearon and Laitin 2000) argue that ethnic distinctions are not innate or immanent, rather socially constructed for acquisition of political or economic ends (“constructivism” approach). We ignore this endogeneity of ethnic borders and treat the set of ethnic groups as exogenously given in order to focus on our main concerns.
Table 1: The prisoner’s dilemma stage game. $\alpha > 1$, $\beta > 0$, $(\alpha - \beta)/2 < 1$.

<table>
<thead>
<tr>
<th></th>
<th>Cooperate</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperate</td>
<td>$1, 1$</td>
<td>$-\beta, \alpha$</td>
</tr>
<tr>
<td>Defect</td>
<td>$\alpha, -\beta$</td>
<td>$0, 0$</td>
</tr>
</tbody>
</table>

In each period, $k$ individuals in each group are paired with members in the other group ("outsiders") while the remaining $n - k$ are paired among themselves ("peers"). $p = k/n$ is a fraction of members in a group who interact across groups in each period, and it denotes the density of the network between groups. It is assumed that $k < n/2$ — thus, $p < 1/2$ — so that interactions within a group ("in-group matches") are more frequent than interactions across groups ("out-group matches").

In the stage game, if both individuals successfully cooperate with each other, they obtain payoff of one. But, one may be tempted to defect against the other for obtaining payoff of $\alpha > 1$, while payoff to a cheated individual is $-\beta < 0$. Payoff from mutual defection is normalized to be zero. As a consequence, the dominant action in the stage game is to defect, and this payoff structure (Table 1) makes the mutual cooperation difficult. Payoffs in the future periods are discounted by a common discount factor $\delta \in (0, 1)$. Each individual $i$ selects $a_{i,t} \in \{C, D\}$ to maximize his own payoff.

### 2.2 Information

An ethnic group often has a dense social network which enables to spread information among its members through rumors or gossips. Thus, information about an individual’s behavior can be easily shared among his peers, whereas such information is less likely to be transmitted beyond the ethnic border or to be shared by outsiders due to the relative infrequency of interactions or the difference in social manners or languages. The structure of information in this two-sided social matching game can be characterized by the difference of difficulties in intra- versus inter-group transmission of information. Therefore, it is plausible to assume that information transmission within a group is easier and quicker than between groups, but in order to make the analysis tractable, we will adopt simpler assumptions as follows.

For in-group interactions, actions and identities of all the members in a group are assumed to be perfectly observed by all of them. Also, an individual’s history of play is perfectly known by his peers. However, information about in-group interactions in a group is totally unobservable and unknown to members in the other group. In other words, information about in-group interactions is local in the sense that what happened within a group is known to everyone in the group, but unknown to others.

For out-group interactions, all pairs of actions are perfectly observed by members.

---

6Following Fearon and Laitin (1996), we assume that both $k$ and $n$ are even, and $n$ is larger than 7.
Table 2: The assumptions on information in the view of members in group A. The assumption on the part with asterisk (*) will be modified later. Information obtained by members in group B follows in the same manner.

<table>
<thead>
<tr>
<th></th>
<th>in-group in A</th>
<th>out-group in A</th>
<th>out-group in B</th>
<th>in-group in B</th>
</tr>
</thead>
<tbody>
<tr>
<td>actions</td>
<td>known</td>
<td>known</td>
<td>known</td>
<td>unknown</td>
</tr>
<tr>
<td>identities</td>
<td>known</td>
<td>unknown*</td>
<td>unknown</td>
<td>unknown</td>
</tr>
</tbody>
</table>

of both groups. However, no one knows the identity about who defected. One can imagine that once a crime happens, it will be reported by newspapers or local media and then quickly become public information, but people cannot tell who committed the crime.

The assumptions on information are summarized in Table 2.

3 Inter-Ethnic Cooperation Enforced by In-Group Policing

3.1 Difficulties of Cooperation in the Large Population

In this social matching game, there are multiple equilibria if the discount factor $\delta$ is sufficiently large, but this section first picks up three representative equilibria. They can be distinguished in terms of population joining cooperation: no cooperation $\sigma_0$, only in-group cooperation $\sigma_1$, and both in- and out-group cooperation $\sigma_2$. These three equilibria will be compared in terms of payoffs and conditions on parameters. We adopt Subgame Perfect Nash Equilibrium (SPNE) as a solution concept.

We consider the following equilibria.

0. *Equilibrium of no cooperation*. All individuals adopt the strategy $\sigma_0$, which is to play D for any pairing.

1. *Equilibrium of in-group cooperation*. All individuals adopt the following strategy $\sigma_1$. For in-group pairings, play C with any individual of the normal phase, and play D against any individual of the punishment phase, regardless of one’s own status. An individual enters (or restarts) the punishment phase of $T^{\text{in}}$ periods if he defects against a peer of the normal phase. An individual who is not in the punishment phase is in the normal phase. For out-group pairings, always play D. The game starts with the normal phase.\footnote{Action pairs in interactions within groups forms a strongly renegotiation-proof in terms of Farrell and Maskin (1989).}
2. Spiral equilibrium (Fearon and Laitin 1996). All individuals take the following strategy $\sigma_2$. For in-group pairings, play in the same way as in $\sigma_1$. For out-group pairings, play C during the peace phase, while play D during the conflict phase. Groups go to the conflict phase of $T^{out}$ periods if any individual defects in an out-group pairing during the peace phase. When groups are not in the conflict phase, it is in the peace phase. The game starts with the normal/peace phase.

In $\sigma_0$, no cooperation occurs in any interaction. In $\sigma_1$, people cooperate only within each group, and those who defected will be punished by their peers. People further cooperate with outsiders in $\sigma_2$, where the ethnic conflict is triggered by a single defection against an outsider. The conditions for equilibria are as follows.

Lemma 1 (i) The strategy profile of $\sigma_1$ forms a SPNE if and only if $p < \frac{1}{1+\beta}$,

$$I (T^{in} > 1) \delta^{1-\delta^{T^{in}-1}} \min\{1+\beta, \alpha\} + \delta T^{in} (1-p) (1+\beta) \geq \alpha - 1,$$

and $\delta T^{in} (1-p) (1+\beta) \geq \beta$.\(^8\) (ii) For the strategy profile of $\sigma_2$ to constitute a SPNE, it is required, in addition to conditions for $\sigma_1$, that

$$\delta \frac{1-\delta^{T^{out}}}{1-\delta} p \geq \alpha - 1. \quad (1)$$

For both $\sigma_1$ and $\sigma_2$ with $\alpha \leq 1+\beta$, if it is a SPNE for given parameters, then it is always possible to take $T^{in} = 1$.

All the proofs appear in the Appendix.

Inequalities in $\sigma_1$ shows the condition for in-group cooperation. Inequality (1) shows that the threat of ethnic conflict, which causes the expected loss of $\delta^{1-\delta^{T^{out}}-p}$, is effective enough to discourage an out-group defection, which brings the immediate gain of $\alpha - 1$. Lemma 1 confirms that cooperation in the larger population will be more difficult to sustain as predicted by theories of the collective action (e.g., Olson 1965; Taylor 1982:53), even though cooperation with more persons is more beneficial in the model. (The per-period expected payoffs from $\sigma_0$, $\sigma_1$ and $\sigma_2$ are 0, 1 - $p$ and 1, respectively.) No constraint is assigned on $\sigma_0$, because every individual takes the dominant action of the stage game. The condition on $\sigma_1$ implies that the network within a group (represented by $1 - p$) must be dense enough that peer punishment is effective to deter transgressions. The equilibrium $\sigma_2$ requires, in addition to the condition on $\sigma_1$, that the social connectedness between groups $p$ be so tight that the threat of group-level conflict effectively discourages out-group transgressions. The difficulty of large cooperation, according to the theory, owes to the infrequency of interactions and the insufficiency of information.

For cases with $\alpha \leq 1+\beta$, the condition for in-group cooperation will be more relaxed if the length of punishment is shorter, because those who are being punished may refuse to conform if the length is so long.

\(^8I (T^{in} > 1)\) denotes the indicator function which gives one if $T^{in} > 1$ and zero otherwise.
3.2 Peer Monitoring and Peer Punishment

Lemma 1 has shown that if the network between groups $p$ is not dense enough, the spiral equilibrium cannot enforce inter-ethnic cooperation. In order to further support cooperation between groups, we consider two forms of peer punishments to supplement the strategy $\sigma_2$ of the spiral equilibrium.

The first form of peer punishment is collective and imposed on all the members who are suspected to have defected. If a defection against an outsider happens in a period, those who interacted with outsiders in that period should be the suspects, and one of them should really have defected. The following strategy profile describes such a punishment.

3. **Equilibrium of collective peer punishment.** All individuals adopt the strategy $\sigma_3$ specified as follows. For in-group pairings, play in the same way as in $\sigma_1$ except that an individual enters the punishment phase if he defects against a peer of the normal phase, or if at least one member in his group defects in out-group pairing when he matches with an outsider during the peace phase. (When a defection against an outsider occurs, all the individuals who matched with outsiders in that period will enter the punishment phase.) For out-group pairings, play in the same way as in $\sigma_2$. (What triggers the conflict phase is also the same as in $\sigma_2$.)

In $\sigma_3$, inter-ethnic cooperation is sustained both by peer punishment and by ethnic conflict.

**Lemma 2** The necessary and sufficient condition for $\sigma_3$ to form a SPNE is the same as in $\sigma_2$ except that Inequality (1) is replaced by

$$I (T^{in} > 1) \delta \frac{1 - \delta^{T^{in}} - 1}{1 - \delta} (1 - p) \frac{k - 1}{n - 1} (-\beta) + \delta T^{in} (1 - p) \left(1 + \frac{n - k}{n - 1} \beta\right) + \delta \frac{1 - \delta^{\gamma_{out}}}{1 - \delta} p \geq \alpha - 1.$$  

(2)

For $\alpha \leq 1 + \beta$, $T^{in}$ can be chosen to be one if the strategy profile is a SPNE.

The difference of $\sigma_3$ from the spiral equilibrium $\sigma_2$ is that in $\sigma_3$, inter-group cooperation relies not only on the threat of conflict, but also on collective peer punishment. As a result, it can sustain cooperation between groups which is impossible in the spiral equilibrium. If people in a group impose peer punishment on all the suspected members, severer punishments than in $\sigma_2$ are possible despite the anonymity of defectors.

The second form of peer punishment is individual rather than collective. However, individual punishment is impossible unless a defector is identified. Therefore, we modify the assumption on information as follows.

Before starting the social matching game (at $t = 0$), each ethnic group $I \in \{A, B\}$ simultaneously determines the effort level $m_I$ for monitoring its members and
assigns per-period cost $m_I$ to every group member. Through this costly investigation, those who defected in out-group interactions will be detected by all the peers with probability $r(m_I)$ at the beginning of the next period, where $r(\cdot)$ is a monotonically increasing function with properties $r(0) = 0$, $r'(m) > 0$ and $r(\cdot) \in (0, 1)$ for $m > 0$. Given a strategy profile of the social matching game, each group $I$ is assumed to maximize the payoff of a representative (homogeneous) member in the group by selecting $m_I$.\(^9\)

For simplicity, let us further set the following assumptions on monitoring. The cost $m_I$ is considered to be a membership tax and cannot be rejected by members. (Those who evade paying taxes will be excluded from the group.)\(^10\) Once $m_I$ is determined, the probability $r(m_I)$ is known to all the members in group $I$ (but, not to anyone in the other group), and $m_I$ cannot be changed in the midst of the game. If a group fails to identify who defected in the period, they can never be identified for all the future periods. If a member who defected is identified, it becomes common knowledge (only) within the group.

By this change in assumption on information, the part with asterisk (*) in Table 2 is modified such that identities in out-group interactions are imperfectly detected with probability $r(m_I)$. The payoff of individual $i$ in group $I$ at a period is equal to the payoff from the prisoner’s dilemma minus the cost for monitoring $m_I$. One may wonder why a group has incentives to monitor its members, although it is costly. Under the threat of ethnic conflict, people are willing not only to cooperate with outsiders, but also to control opportunistic behaviors by their ethnic brethren. Without such in-group control, someone may harm outsiders, and then everyone in the group would be penalized. In order to control others in an effective way, a group may introduce monitoring to induce conformity by peers.

Here, we do not specify the process of determining the effort level of peer monitoring within a group. It may be determined by a benevolent group leader. Or, it may be by majority rule. Neither do we specify how peer monitoring is implemented or how information is correctly shared by peers. For the time being, we simply treat a group as a decisionmaking body for the level of peer monitoring and assume that the

---

\(^9\)Here, we assume a collective decision (costly monitoring) in order to solve for another collective action problem (cooperation in stage games). This is so-called the second-order collective action problem which is often discussed with the provision of common-pool resources. Some scholars (Taylor 1982:65; Singleton and Taylor 1992) are relatively optimistic about resolving the problem without a formal institution, while others (Hechter 1984; Ostrom 1992) are more pessimistic. Since it is not the main focus of this paper, the present model assumes it away to reduce the complexity of analysis and does not specify how the monitoring is implemented. Members in a group may call for specialists of monitoring, or members themselves may be able to effectively monitor each other. Thus, this assumption does not necessarily imply the need of a formal institution.

\(^10\)For example, in Jewish communities of seventeenth-century Lithuania, licenses were issued to taxpayers, and without the license, people were ostracized from communities. Without payment of tax, "no Jew was allowed to settle in a given locality or to enroll his name in the community" (Dubnow 1916:190).
result of monitoring is common knowledge within a group. The issue of monitoring process will be discussed in the section of conclusion.

By using peer monitoring, we introduce a strategy profile with individual peer punishment.

4. Peer monitoring equilibrium. All individuals take the strategy σ₄ specified as follows. For in-group pairings, play in the same way as in σ₁ except that an individual enters the punishment phase by defecting a peer of the normal phase, or by being identified to have defected (with a probability determined by costly monitoring) in out-group pairings during the peace phase. For out-group pairings, play in the same way as in σ₂. The monitoring cost $m_I$ for each group $I$ is determined as the minimum level $m^*$ which satisfies Inequality (3) below.

In contrast to $σ₃$, peer punishment in $σ₄$ is imposed on a particular individual who defected in the previous period. If a group fails to identify who defected, peer punishment will not be imposed on anyone.

**Proposition 1** The strategy profile $σ₄$ forms a SPNE if and only if $p < \frac{1}{1+β}$, $I (T^{in} > 1)$

$$δ^{1-δT^{in}-1} \min\{1 + β, α\} + δ^{T^{in}} (1 - p) (1 + β) ≥ α - 1, \ δ^{T^{in}} (1 - p) (1 + β) ≥ β,$$

and there exists $m ∈ [0, p]$ such that

$$r (m) δ^{T^{in}} (1 - p) (1 + β) + δ \frac{1-δT^{out}}{1-δ} p ≥ α - 1. \quad (3)$$

For $α ≤ 1 + β$, $T^{in}$ can be chosen to be one if the strategy profile is a SPNE.

Proposition 1 gives an important implication; inter-ethnic cooperation of $σ₄$ is possible only if each group can effectively police its members within a limited cost of monitoring. If the cost ($m^*$) exceeds the benefit from inter-ethnic cooperation ($p$), the group refrains from cooperating with outsiders. Thus, an ethnic group which lacks the means of efficiently disciplining its members fails to cooperate with other ethnic groups. Such a case with very inefficient monitoring leads to the strategy profile $σ₁$.

On the other hand, this equilibrium of $σ₄$ appears problematic for two reasons. First, from Inequality (3), it follows that $m^*$ falls in $T^{out}$, implying that as ethnic conflict becomes longer, its threat becomes more effective, so that the required level of monitoring falls. Thus, the optimal equilibrium involves very long periods of conflict ($T^{out} → ∞$). This optimality of very large $T^{out}$ comes from the fact that the ethnic conflict never occurs on the equilibrium path.

Second, in comparison with Lemma 1, Proposition 1 says that the condition for the peer monitoring equilibrium is weaker than that for the spiral equilibrium in the sense that monitoring is employed ($m^* > 0$) as a remedy for recovering incentives for inter-ethnic cooperation. However, in comparison with Lemma 2, it might be possible to sustain cooperation even without costly monitoring if collective peer punishment
is introduced. In addition, if $m^* > 0$, the expected payoff from the peer monitoring equilibrium, which is $1 - m^*$, is strictly lower than that from the equilibrium of collective peer punishment, which is one. It is because in an idealistic world without noise, any sort of punishment is never realized on the equilibrium path; peer monitoring simply causes inefficiency.

To address these two problems, so we consider cases with noisy interactions in the next section. As analyzed in the next section, the result drastically changes once a possibility of misinterpretation in actions (noise) is introduced. In the presence of noise, groups may want to pay costs for monitoring in order to reduce the risk of ethnic conflict and/or to make its length shorter.

4 In-Group Policing Developed by Out-Group Conflict

4.1 The Game with Noisy Interactions

This section develops the model for more plausible cases with noisy interactions, in which, with a small probability $\varepsilon$, a person’s cooperative behavior is mistakenly interpreted as a "defection" by ethnic strangers. As pointed out by sociologists (e.g., Hechter 1987:178), misinterpretation of actions may be caused by cultural difference between groups, and it is more likely if the difference is larger. Thus, the probability $\varepsilon$ shows how culturally different two groups are. Another interpretation of noise would be the reduced version of more plausible assumption that in each group, there are a very small number of players of bad "type" who occasionally damage ethnic strangers regardless of the severity of punishment. We avoid the complexity of incorporating such a type, but this section takes such a possibility into account. As shown later, this small probability of misinterpretation makes inter-ethnic cooperation more difficult. Precisely, the assumption on out-group interactions is modified as follows.

A cooperative action by member $i$ in group $I$ in out-group interaction is occasionally perceived as a defection by members in the other group $J \neq I$ with a small probability $\varepsilon$. Such an accident of misinterpretation happens independently from any other accidents of misinterpretation. As a consequence, given that individual $i$ chooses to cooperate, the opponent’s cooperation gives individual $i$ the payoff of one with probability $1 - \varepsilon$ and of $-\beta$ with probability $\varepsilon$. Similarly, given that individual $i$ defects, the opponent’s cooperation gives individual $i$ the payoff of $\alpha$ with probability $1 - \varepsilon$ and of zero with probability $\varepsilon$. The ex ante payoff of the stage game with noise is summarized in Table 3.

In the case of accident that $i$’s cooperation is misinterpreted by outsiders, it is also misinterpreted by the rest of members in $I$ until it is costly investigated through peer monitoring. By peer monitoring, the identity of who intentionally defected is revealed with probability $r(m_I)$, and this result of peer monitoring is shared by all
the peers, but not by outsiders. (This assumption might be justified if only peers can imperfectly see whether the suspect’s payoff is $\alpha$ or 1.) So, outsiders still cannot tell whether an out-group defection is intentional or accidental. For simplicity, such misinterpretation can happen only in out-group interactions, but no misinterpretation occurs for in-group interactions.

### 4.2 Inter-Ethnic Cooperation in Noisy Interactions

Given such misinterpretation of actions, we will modify the strategy profiles with inter-ethnic cooperation $\sigma_2$, $\sigma_3$ and $\sigma_4$ in two ways. First, the length of peer punishment is set to be one ($T^{in} = 1$), which is always possible for $\alpha \leq 1 + \beta$ if a strategy profile is a SPNE. The purpose of this restriction is to make the analysis tractable and to reduce the damage of collective peer punishment.

Second, each group allows for some defections by outsiders in a period as unavoidable "noise." Due to the presence of misinterpretation, each group ignores a certain amount of observed out-group defections in order to reduce the likelihood of ethnic conflict, and ethnic groups refrain from conflict as long as the amount of defections is negligible such that the number of defections (henceforth $\#D_I$) in either group $I$ against outsiders in a period does not exceed a cutoff level $Q$. The parameter $Q$ denotes the propensity to preserve peace or "tolerance" which is larger if groups are more concessive toward each other and smaller if they are more confrontational. For simplicity, we assume that two groups have the same level of $Q$.

The equilibria modified from $\sigma_2$, $\sigma_3$ and $\sigma_4$ follow.

5. **Spiral equilibrium** (modified). All the individuals adopt the following strategy $\sigma_5$. For in-group pairings, play in the same way as in $\sigma_2$. The length of in-group punishment is one ($T^{in} = 1$). For out-group pairings, play as in $\sigma_2$ except that groups enter the conflict phase if at least $Q + 1$ members in either group are perceived, by the other group, to have defected.\(^{12}\)

6. **Equilibrium of collective peer punishment** (modified). All the individuals adopt the following strategy $\sigma_6$. For in-group pairings, play in the same way as in $\sigma_3$

\(^{11}\)Bendor and Mookherjee (1987) considered a similar approach of noisy monitoring with setting a cutoff level of mistakes in $N$-player collective action problems.

\(^{12}\)Although ethnic conflict in $\sigma_5$, $\sigma_6$ and $\sigma_7$ is conditioned only on events in the previous period, a more realistic condition would be the accumulation of all the past events. Moore (1978:105), for example, argues that disputes between Tallensi and Nyakyusa in eastern Africa were enlarged by "the prior or nascent structural oppositions and competition between groups."
except that an individual enters the punishment phase if he defects against a peer of the normal phase, or if at least $Q + 1$ members in his group defect in out-group pairings when he matches with an outsider during the peace phase. For out-group pairings, play as in $\sigma_5$.

For simplicity, we use the same degree of tolerance $Q$ as a cutoff level of collective peer punishment in $\sigma_6$.

7. Peer monitoring equilibrium (modified). All the individuals adopt the following strategy $\sigma_7$. For in-group pairings, play as in $\sigma_4$ except that an individual enters the punishment phase by defecting a peer of the normal phase, or by being revealed to have intentionally defected against outsider during the peace phase. For out-group pairings, play as in $\sigma_5$. The monitoring cost $m_I$ for each group $I$ is determined as the minimum level $m^*$ which satisfies Inequality (4) below.

Before presenting equilibrium conditions, it may be convenient to show some properties of the random variable $D_I$. Because there are $k$ out-group interactions in a period and a misinterpretation happens with probability $\varepsilon$, the random variable $D_I$ follows a binomial distribution during the peace phase; i.e., $D_I \sim B(k, \varepsilon)$. Let $F_k(Q)$ be the cumulative distribution function of $D_I$ less than or equal to $Q$ observed defections among $k$ interactions.

$$F_k(Q) = \Pr (D_I \leq Q | D_I \sim B(k, \varepsilon)) = \sum_{j=0}^{Q} C_j^k (1 - \varepsilon)^{k-j} (\varepsilon)^j,$$

where $C_j^k = \frac{k!}{j!(k-j)!}$. Then, $F_k(Q)$ equals the probability that a group does not trigger ethnic conflict in a period, and therefore the square of $F_k(Q)$ equals the probability that the peace phase continues to the next period. If a player intentionally deviates, the probability to preserve peace declines to $F_{k-1}(Q-1) F_k(Q)$.

Equilibrium condition for $\sigma_7$ is the following. (Conditions for $\sigma_5$ and $\sigma_6$ appear in the Appendix.)

**Proposition 2** The strategy profile $\sigma_7$ is a SPNE if and only if $p < \frac{1}{1 + \beta}$, $\delta (1 - p) (1 + \beta) \geq \max \{\alpha - 1, \beta\}$, and there exists $m \leq V^{out}$ such that

$$r(m) \delta (1 - p) (1 + \beta) + (F_k(Q) - F_{k-1}(Q - 1)) F_k(Q) \delta \frac{1 - \delta^{out}}{1 - \delta} V^{out} \geq (1 - \varepsilon)(\alpha - 1) - \varepsilon \beta,$$

$^{13}$Due to the presence of noise, the strategy profile for in-group pairings is no longer renegotiation-proof (both action profiles of (C,C) and (D,D) coexist), but we continue to employ it.
where \( V^{\text{out}} = \frac{\left(1 - \delta\right)p\left(1 - \varepsilon\right) - \varepsilon \beta}{1 - \delta\left(1 - \delta \tau^{\text{out}}\right)F_k(Q)^2 + \delta^2 \delta^{\text{out}}} \) denotes the per-period ex ante expected payoff from out-group interactions.

The interpretation of the above conditions is immediate. While the term \((1 - \varepsilon) (\alpha - 1) - \varepsilon \beta\) in the right-hand side (RHS) of Inequality (4) denotes the gain from intentional deviation, the term in the left-hand side (LHS) shows the change in the expected loss caused by it. So, the condition says that the expected losses by punishment must be large enough to discourage opportunistic deviations.

According to this equilibrium, minor misconducts between ethnic groups could unexpectedly destroy the inter-ethnic norm, and group-based reprisals may follow, involving a large portion of people who seemed to have kept peaceful relations for a long time. This sudden fall of inter-ethnic norm may explain ferocious events such as the communal violence between Christian Copts and Muslims in southern Egypt in 2000.\(^{14}\)

### 4.3 Superiority of Peer Monitoring Equilibrium

In the next step, we will investigate how the presence of small noise influences these equilibrium outcomes. For three different levels of noises, Table 4 gives per-period ex ante expected payoffs \( V_5, V_6 \) and \( V_7 \) of three equilibria \( \sigma_5, \sigma_6 \) and \( \sigma_7 \), respectively. (Formulas for \( V_5, V_6 \) and \( V_7 \) appear in the Appendix.) For the peer monitoring equilibrium \( \sigma_7 \), three cases with different efficiencies of peer monitoring are analyzed in the table.

Table 4 shows that although all three strategy profiles can be sustained as SPNE in cases without noise (\( \varepsilon = 0 \)) with the given parameter values, results drastically change once noise (\( \varepsilon > 0 \)) is introduced. As noise rises to \( \varepsilon = .01 \), the spiral equilibrium, which has the weakest punishment device among three strategy profiles, first becomes impossible to sustain regardless of \( Q \).

For the equilibrium of collective peer punishment \( \sigma_6 \), although it can achieve SPNE for any levels of noise (\( \varepsilon = .001, .01 \) or .05) it is not possible to form SPNE with large degrees of tolerance \( Q \). (For example, with \( \varepsilon = .01, Q > 3 \) destroys the equilibrium condition of \( \sigma_6 \).) It is because a sufficiently large degree of tolerance makes both peer and external punishments less likely and less effective. Knowing that an amount of

\(^{14}\)I provide two events in which conflict seems to be caused by noisy interactions between ethnic groups. First, an event of rural Egypt in 2000 might be a case that miscommunication between groups led to inter-ethnic turmoils. The communal violence of southern Egypt in 2000 was triggered by a Muslim customer’s murder of a Christian shopkeeper who refused to apologize about his insult on the Muslim. For details, see *The Economist*, 8 January 2000; *The Financial Times*, 8 February 2000. Second, Dumont (1982:222) reported from the Ottoman Empire that conflicts were caused simply by mistrust between ethnic groups. "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."
defections are recognized as mistakes and ignored, an individual correctly expects that the probability that his deliberate defection triggers punishment is negligible, and his willingness to conform will be eroded. Moreover, low ex ante payoffs $V_6$ with a large amount of noise makes inter-ethnic cooperation by collective peer punishment ($\sigma_6$) incredible. For example, in a case with $\varepsilon = .05$, due to the presence of noise, groups may suffer from frequent peer punishment. If ex ante payoff is below $1 - p$, inter-ethnic cooperation is not worthwhile since in-group cooperation of $\sigma_2$ guarantees payoff of $1 - p$. In short, spiral equilibrium and equilibrium of collective peer punishment tend to be destroyed as the noise grows.

In contrast, if each group can efficiently monitor its members (e.g., $c = .2$), the peer monitoring equilibrium $\sigma_7$ can form a SPNE even with large $Q$ and large $\varepsilon$ as shown in Table 4. Moreover, it can achieve higher ex ante payoffs than the equilibrium of collective peer punishment $\sigma_6$. These results indicate that the selection of collective
punishment and individual one depends on the efficiency of peer monitoring. If a
group can monitor its members in a cheap way, the peer monitoring equilibrium is
more payoff-enhancing and more robust to noise than the equilibrium with collective
peer punishment. These results suggest that collective punishment by peers is unlikely
to occur in the presence of noise.

5 Comparative Statics

This section analyzes comparative statics. Because it is already discussed that equi-
libria without peer monitoring are unlikely in cases with noise, we confine attention to
the peer monitoring equilibrium $\sigma_7$. So far, we have treated $Q$ as exogenously given
and paid attention to the Pareto optimal equilibrium among equilibria with various
levels of $Q$. Trivially, such an equilibrium survives even if each group endogenously
chooses its tolerance toward the other group simultaneously and independently after
observing $m_A$ and $m_B$.\footnote{Suppose that groups are in the Pareto optimal equilibrium with $m^* > 0$. Given that each group $I$ simultaneously and independently chooses its tolerance to the other group $Q_I$ after observing $m_A$ and $m_B$, group $I$ has no incentive to increase in $Q_I$, because it destroys the incentive constraint (Inequality 4) in the other group. Also, group $I$ does not want to decrease $Q_I$, because makes conflict more likely.} Therefore, we continue to treat $Q$ as exogenous and regard
the Pareto optimal equilibrium as the most plausible outcome.

5.1 Efficiency of Peer Monitoring and Tolerance Toward Eth-
nic Strangers

The first observation is that tolerance can either support or harm inter-ethnic coop-
eration, depending on the efficiency of monitoring. It is trivial that the cumulative
distribution function $F_k(Q)$ increases in the cutoff level of conflict $Q$. In other words,
as long as equilibrium condition in Proposition 2 is satisfied, ethnic conflict is less
likely between groups with higher tolerances. The following corollary is immediate.

**Corollary 1** There exists a SPNE $\sigma_7$ with large $Q < k$ if, in addition to condition
in Proposition 2, there exists $m \leq V^{out}$ such that

$$r(m) \delta (1 - p) (1 + \beta) \geq (1 - \varepsilon) (\alpha - 1) - \varepsilon \beta.$$    (5)

All the proofs of Corollary 1 and after are straightforward and will be omitted.

Inequality (5) implies that groups capable of effectively monitoring and punish-
ing its own members can successfully cooperate by expecting each other group to
effectively police its own transgressors to the extent that in-group policing does not
require external threat. Such an inter-ethnic relation between tolerant groups is char-
acterized by mutual trust: ethnic groups refrain from group-based reprisals as means
for sustaining inter-ethnic social order. The equilibrium with efficient monitoring corresponds to a case with \( c = .2 \) in Table 4. However, even in cases with effective in-group policing and large tolerance, threat of ethnic conflict still needs to exist for urging costly monitoring in the other group. It is worth pointing out that threat of conflict is intended not to directly discourage individual transgressions from the target group, rather to give an incentive for the group to costly monitor its members. In the absence of such threat (i.e., \( Q = k \)), there will be no reason to make costly effort for monitoring, and as a result, inter-ethnic cooperation cannot be sustained. Our peer monitoring equilibrium with efficient monitoring and high tolerance corresponds to Fearon and Laitin’s (1996) in-group policing equilibrium, in which "individuals ignore transgressions by members of the other ethnic group, correctly expecting that the culprits will be identified and sanctioned by their own ethnic brethren."\(^{16}\)

In contrast, the same argument cannot apply for groups incapable of effective punishments (e.g., the case with \( c = .08 \) in Table 4). If groups are very tolerant, conflict would be very unlikely regardless of a player’s action. Knowing that, an individual might be more tempted to take a deviant behavior. If a group kindly tolerates the other group’s wrongdoings as mistakes or misinterpretations, external threat of conflict becomes impotent (\( Q \geq 5 \) in Figure 1), and the tolerance may cause more deviant actions by destroying the incentive to conform. Thus, severe in-group punishment is required in order to maintain the incentive for out-group cooperation. However, the in-group policing is costly, and if the required cost for peer monitoring is higher than the benefit from inter-ethnic cooperation, it is impossible to discipline its members (\( Q \geq 4 \) with \( c = .08 \) in Figure 2), and inter-ethnic cooperation breaks down. Therefore, for groups with ineffective in-group policing, tolerance may rather harm inter-ethnic cooperation, and occasional ethnic conflict is unavoidable. These groups cannot ignore observed transgressions as misinterpretation since ignoring them may call for further transgressions, and therefore the groups must provoke conflict for demonstrating that threats are in reality.\(^{17}\) Such an equilibrium mainly sustained by external threats, corresponding to Fearon and Laitin’s (1996) spiral equilibrium, is characterized by mutual mistrust: they are threatening each other in order to directly suppress transgressions from the target group.

From historical observation, Jewish communities of the Ottoman Empire in the nineteenth century successfully reduced the socioeconomic antagonisms from non-Jewish ones and alleviated their economic and social backwardness by reinforcing the social control on management of communal business and by promoting alliance among

---

\(^{16}\)In his autobiography, well-known ‘Lawrence of Arabia’ reported a lonely Arab man who was ostracized and cut off from any friendly intercourse with his tribe, being penalized for murdering a Christian (Lawrence 1935:77-8). This example seems to exactly fit the punishment of in-group policing equilibrium. I thank Dilip Mookherjee for pointing out this story.

\(^{17}\)A similar observation of competition and collusion in oligopolistic markets is theorized by Green and Porter (1984).
Figure 1: The vertical axis shows the expected loss from ethnic conflict by a deliberate defection (the second term of the LHS in Inequality 4) in a case with $\varepsilon = .01$ of Table 4. The figure shows that if the tolerance is too low or too high, then it is unlikely that an intentional defection becomes pivotal to trigger ethnic conflict, so that the external threat is not very effective to induce out-group cooperation.

Figure 2: $V^{out}$ and $m^*$ with $\varepsilon = .01$ for each of $c = .1$ and $c = .08$ in Table 4 are shown. For $c = .08$, inter-ethnic cooperation is not sustainable with $Q \geq 4$, because the monitoring cost exceeds the benefit from it.
Jewish subgroups (Dumont 1982:229-30). This might be an example showing that strong in-group policing can enhance inter-ethnic cooperation.

Also, from a relatively recent event, although it is not about monitoring cost, the difficulty of in-group policing was often reported during Arafat’s period of the Palestinian Authority. Under pressure from the United States, Russia and European countries, Arafat was faced with the dilemma of cracking down on some members of Hamas and Islamic Jihad, which was deeply unpopular among most Palestinians. From the opposite side, the Netanyahu administration conditioned the progress of the Road Map for peace on the Palestinian Authority’s fulfillment of obligations, including the cessation of terrorist attacks. The Sharon administration’s demand includes the removal of the Palestine leader, Yasser Arafat, and strict limits on Palestinian security forces. These policies look very much like the threat of external sanction for deterring transgressions in the target group. These examples suggest that the quality of in-group policing matters for inter-ethnic peaceful order and that a group without strong internal control may fail to avoid violent interactions with other groups.

5.2 Length of Ethnic Conflict

The second observation is about the length of ethnic conflict. Obviously, if ethnic conflict is shorter, the damage caused by occasional conflict also will be smaller in the presence with noise. On the other hand, shorter ethnic conflict reduces the threat on the target group and requires more cost for in-group policing. Therefore, groups with high qualities of in-group policing can achieve inter-ethnic cooperation with shorter conflict, whereas those with lower qualities may suffer from longer struggles.

Figure 3 shows per-period ex ante expected payoffs of high-quality groups and of low-quality ones with the same parameter values as in Table 4 except giving various lengths of conflict. For high-quality groups, the expected payoff is larger for higher tolerance $Q$ and for shorter length of conflict $T_{\text{out}}$. Because in-group policing can be achieved very efficiently, it does not have to rely on the threat from outside.

In contrast, for low-quality groups, the payoff is maximized at a medium range of tolerance $Q$, implying that inter-ethnic cooperation cannot be sustained either for too tolerant or too intolerant groups with low qualities of policing. Also, it can be seen that the longer period of conflict $T_{\text{out}}$ achieves the larger expected payoff, since it

---

18 According to Dumont (1982:229-30), "[t]hese manifestations of brotherhood caused not only an improvement in the plight of the most disadvantaged strata. They also helped reduce socioeconomic antagonisms which, since the sixties, troubled the life of some communities", and "the various programs of the alliance in order to alleviate the economic and social backwardness of the Jewish communities of Turkey proved, on the whole, very effective."

19 For example, see J. Bennet, "Palestinian Authority Arrests Jihad Leader, Causing a Riot," New York Times, 15 November 2001, sect. A.

Figure 3: It shows ex ante payoffs from equilibria with two different monitoring efficiencies ($c = .2$ and $c = .08$) and three different lengths of ethnic conflict ($T^{out} = 1, 10, 100$). Other parameter values are the same as in Table 4. With efficient monitoring ($c = .2$), the equilibrium of shorter conflict ($T^{out} = 1$) gives a larger payoff, whereas with inefficient monitoring, the payoff is higher with longer conflict ($T^{out} = 100$). Especially, equilibrium cannot be sustained with $c = .08$ and $T^{out} = 1$. (So, it does not appear in the figure.)
helps to save the monitoring cost. Especially, equilibrium cannot be sustained when the length is too short ($T^{\text{out}} = 1$). These results predict that *groups high qualities of in-group policing can successfully cooperate with each other by expecting autonomous sanctioning of culprits in the other group, whereas those with low qualities cannot eliminate the risk of collective violence and may suffer from more frequent and longer disputes*. Furthermore, the theory predicts that groups with even less qualities cannot construct social order due to the lack of policing regime.

### 5.3 Density of Network Between Groups

A change in the density of network between groups $p = \frac{k}{n}$ has various effects on inter-ethnic cooperation and conflict. First, if noise is present ($\varepsilon > 0$), more frequent interactions between groups generate more defections by first-order stochastic domination of the random variable $\#D_I$, and ethnic conflict may become more likely.

**Proposition 3** *Given other parameter values being equal and condition in Proposition 2 being satisfied, an increase in frequency of out-group interactions relative to frequency of in-group interactions raises the probability of ethnic conflict. This effect is more severe if two groups are culturally more disparate.*

Second, an increase in out-group network relative to in-group network reduces the power of in-group punishment. This makes the incentive for out-group cooperation harder to sustain. On the other hand, frequent interaction between groups may make inter-ethnic cooperation more attractive, making costly monitoring worthwhile. (Mathematically, it alleviates the condition $m \leq V^{\text{out}}$ in Proposition 2.) In total, the impact of a change in the density of out-group network on inter-ethnic cooperation is ambiguous, depending on each group’s quality of policing. But given all else equal, Proposition 3 implies that a dense network tends to harm cooperation if groups are far disparate. Both directions of effects by the change in the density of out-group network were reported from historical evidences and empirical findings.

A striking example was reported from Turkey. From a study of inter-ethnic interactions of Turkish-Kurdish case and Turkish-Laz case, Oztalas (2006) claimed that modernization increases the probability of ethnic conflict and war only if it combined with other factors such as strong ethnic prejudice, extremist elite mobilization, exclusionary system or the security dilemma. Taking into account that modernization here represents migration and urbanization which are likely to facilitate inter-ethnic interactions, this report would support the prediction of our theory that frequent inter-ethnic interaction may increase the likelihood of ethnic conflict if ethnic groups are culturally disparate or intolerant toward each other.

In addition, it was reported from the nineteenth-century Ottoman Empire that due to isolation or separation from other local habitants, Jewish and Gypsy merchants were often collectively accused of local incidents. It implies that the lack of strong network between groups harms inter-ethnic cooperation. According to Dumont
(1982:223), "[w]hen some incident occurs in a locality, the scapegoat was always the same: the accusations were directed at a band of Gypsies or a Jewish ragpicker who had wandered through sometime earlier." Moreover, "numerous anti-Jewish riots were accompanied by boycott. As soon as some trouble occurred, Christians forbade Jews access to their quarters and stopped trading with Jewish bazaar merchants."\(^{21}\)

However, it can also be interpreted that the lack of strong network makes communication between groups more difficult and misinterpretation more likely. If so, the possibility of misinterpretations \(\varepsilon\) might be negatively associated with the density of out-group network \(p\). This is out of the scope of our theory.

6 Conclusion

In this paper, we provided a new cause of ethnic conflict which is missing in existing theories: lack of in-group policing. Inter-ethnic transgression may provoke indiscriminate vengeance for asking collective liability in the target group. Under the threat of such communal violence, people are induced to deter coethnics’ deviant behaviors against outsiders through mutual monitoring and controlling. In our model, inter-ethnic social order is enforced by in-group policing and out-group conflict, but the former is induced by threat of the latter. Each group relies on the other group’s internal policing, because peers are in a good position for monitoring and punishing potential wrongdoers. This pattern of ethnic conflict is widely observed in anthropological and historical literature. Our theory is in contrast to the existing one (Fearon and Laitin 1996) which considers that in-group policing and conflict are mutually independent and that the main cause of ethnic conflict is the anonymity of culprits.

As a logical consequence of our mechanism of inter-ethnic cooperation, each group’s quality of in-group policing matters for the likelihood and the duration of ethnic conflict. To be precise, groups with high qualities of policing could successfully maintain inter-ethnic cooperation, whereas those with low qualities have to bear more frequent and longer disputes. If a group cannot effectively suppress its members’ wrongdoings, ethnic conflict might be inevitable.

Our model gives two other predictions from comparative statics. First, the dense network between groups may help or harm inter-ethnic cooperation, depending on how culturally disparate two groups are. Frequent interactions between groups may provoke ethnic conflict if groups are far disparate, whereas it can help for inter-ethnic social order between groups with less cultural disparity.

Second, although inter-ethnic tolerance facilitates cooperation between groups if

\(^{21}\)In contrast, two recent studies suggest that what Putnam (2000) labeled "bridging" social capital may enhance mutual trust and reduce antagonisms toward each other. First, using data from two minority regions of Russia, Bahry et al. (2005) found that trusts on ethnic strangers decline with social and physical distance. Second, with data from the 1992-1994 Multi-City Study of Urban Inequality and from 1990 Census (both of which are of the U.S.), tests by Oliver and Wong (2003) suggest that ethnic spatial and social isolation bolster negative out-group perceptions.
they can effectively police their own members, tolerance may lead to inter-ethnic transgressions if they are not very capable of policing members. It is because the incentive to conform will be eroded if groups are too tolerant toward each other.

Finally, we discuss an agenda for future research. Although we do not assume any centralized formal institution in the model, it may have a local formal institution or authority in each ethnic group. Especially if costly monitoring involves free-rider problem, the need for formal institution is more significant for larger ethnic groups. For such a local authority, the political decision making process of in-group policing and of out-group tolerance may matter for inter-ethnic peace. As implied by the dilemma Arafat faced, it is not straightforward to make these policies if people in a group have different propensities to peace and conflict. All of these ideas are abstracted from the model. On the other hand, local authority may not be required if the size of ethnic group is relatively small. However, it still matters in a small group how information about peers’ actions is truthfully revealed and correctly transmitted to all the members. We assumed that such information becomes common knowledge among peers once it is revealed, but the information transmission is not immediate since anyone who committed a crime always has incentives to oppose the revealed fact, and in addition, the information is not necessarily verifiable for third parties. It is not clear how informal institution gives enough incentives for sharing the information among members. For both formal and informal institutions, the detailed mechanism of in-group policing is not yet theoretically explained.

APPENDIX

Proof of Lemma 1. Because of the homogeneity of individuals and the symmetry of groups, it suffices, without loss of generality, to show that incentive constraints hold for individual $i$ of group $I$ after any history. Also, by the optimality principle of dynamic programming, it is sufficient to check that one-shot deviations are unprofitable in any state (Fudenberg and Tirole 1991:108-10).

Before proceeding the proof, define the system of states as $s_t = (t_0, t_1, t_2, \ldots, t_n)$, where $t_0$ denotes the number of periods remaining in the conflict phase, and $t_i$ for $i = 1, \ldots, n$ denotes the number of periods remaining in individual $i$’s punishment phase. Individual $i$ is in the normal phase when $t_i = 0$, and the state is in the peace phase when $t_0 = 0$. We call an individual in the normal phase "cooperator" and one in the punishment phase "defector." Also, let $n_{t+i}$ be the number of cooperators except $i$ in $I$ in period $t+l$, and $q_{t+l} = \frac{n_{t+i}}{n-1}$ be the probability that player $i$ is paired with a cooperator in period $t+l$ if $i$ is paired in-group.

(i) For $\sigma_1$ to form a SPNE, all the incentive constraints must be satisfied in all the states $s_t$: a cooperator $i$ ($t_i = 0$) has no incentive (1-1a) to cooperate with an in-group defector, (1-1b) to defect against an in-group cooperator, and (1-1c) to cooperate with an out-group player; as well as a defector $i$ ($t_i > 0$) has no incentive.

\[22\] The proof of Lemma 1 is fundamentally from the Appendix of Fearon and Laitin (1996).
(1-2a) to cooperate with an in-group defector, (1-2b) to defect against an in-group cooperator, (1-2c) to cooperate with an out-group player. Since it is immediate that cases 1-1a, -1c, -2a and -2c are satisfied for any $s_t$, we confine attention to conditions for 1-1b and -2b.

For 1-1b, if an individual defects, he will gain the additional payoff of $\alpha - 1$ in the current period, while the expected loss will be

$$\sum_{t=1}^{T_{in}} \delta^t (1 - p) (q_t + (1 + \beta) + (1 - q_t) \alpha),$$

where $1 + \beta = 1 - (-\beta)$ is the loss if he is paired with a cooperator and $\alpha$ is the loss if $i$ is paired with a defector. Thus, the incentive constraint for 1-1b is

$$\sum_{t=1}^{T_{in}} \delta^t (1 - p) (q_t + (1 + \beta) + (1 - q_t) \alpha) \geq \alpha - 1.$$

This constraint must hold for any possible $q_t$, and it is most restrictive when the loss is minimized. If $\alpha > 1 + \beta$, the loss is minimized by $q_t = 1$ for $1 \leq t \leq T_{in}$. On the other hand, if $\alpha \leq 1 + \beta$, the loss is minimized by $q_t = 0$ for $1 \leq t \leq T_{in} - 1$ and $q_{T_{in}} = 1$. (It is because if all players $j \neq i$ follow $\sigma_1$, no one is in the punishment phase at states with $t_i = 1$.) Therefore, Constraint 1-1b can be shown as

$$I \left(T_{in} > 1 \right) \frac{1 - \delta^{T_{in}-1}}{1 - \delta} (1 - p) \min\{1 + \beta, \alpha\} + \delta^{T_{in}} (1 - p) (1 + \beta) \geq \alpha - 1.$$

For 1-2b, the gain from the deviation is $\beta$ while $i$ loses the payoff of $\sum_{t=1}^{T_{in}} \delta^t (1 - p) (q_t + (1 + \beta) + (1 - q_t) \alpha)$ by the deviation at states with $t_i$. The incentive constraint for 1-2b is

$$\sum_{t=1}^{T_{in}} \delta^t (1 - p) (q_t + (1 + \beta) + (1 - q_t) \alpha) \geq \beta,$$

which must be satisfied for any $t_i > 0$. The loss is minimized when $i$ is in the first period of the punishment phase ($t_i = T_{in}$), and for the same reason as above, it must be that $q_{T_{in}} = 1$. Thus, Constraint 1-2b can be reduced to

$$\delta^{T_{in}} (1 - p) (1 + \beta) \geq \beta.$$

The condition $p < \frac{1}{1 + \beta}$ comes from Constraint 1-2b with $\delta < 1$.

For $\alpha \leq 1 + \beta$, Constraint 1-2b is sufficient for 1-1b, and it is most relaxed when $T_{in} = 1$.

(ii) In addition to the constraints in $\sigma_1$, $\sigma_2$ requires the constraint for out-group cooperation. I.e., a player $i$, regardless of $t_i$, has no incentive (2-1d,2d) to defect against an outsider when $t_0 = 0$. The benefit from the deviation is $\alpha - 1$, and the loss is $p \sum_{t=1}^{T_{out}} \delta^t$. Constraint 2-1d,2d is

$$\delta^{1 - \delta^{T_{out}}} \frac{1 - \delta^{T_{out}}}{1 - \delta} \geq \alpha - 1.$$

Q.E.D.
Proof of Lemma 2. For the strategy profile $\sigma_3$ to constitute a SPNE, in addition to constraints for in-group pairings in $\sigma_1$, the following constraints must hold for out-group pairings: a cooperator $i$ ($t_i = 0$) has no incentive (3-1c) to cooperate if $t_0 > 0$, (3-1d) to defect if $t_0 = 0$; and a defector $i$ ($t_i > 0$) has no incentive (3-2c) to cooperate if $t_0 > 0$, (3-2d) to defect if $t_0 = 0$. Since 3-1c and -2c are trivially satisfied, let us focus on 3-1d and -2d.

For 3-1d, the benefit from the deviation is $\alpha - 1$, and the loss comes from ethnic conflict $\sum_{t=1}^{T_{out}} \delta^t p$ and from collective peer punishment. A player $i$’s defection against an outsider sends into the punishment phase not only $i$, but also other $k - 1$ peers who matched with outsiders when $i$ defected. This change in other peers’ status also affects $i$’s expected payoff when $i$ defected. The effect of $i$’s defection on his peers will be at least (a) that the number of those in the punishment phase increases by $k - 1$ only in the last period of the punishment phase ($t_i = 1$), which happens if all $n - 1$ others entered the punishment phase just one period before $i$ had defected; and at most (b) that the number of the punished rises by $k - 1$ for every period of $i$’s punishment phase ($1 \leq t_i \leq T_{in}$), which happens if all $n - 1$ others are in the normal phase when $i$ defected.

In case 3-1d (a), if $i$ deflects, $i$’s expected payoffs from each of future in-group matches will be zero for the first $T_{in} - 1$ periods of the punishment phase ($2 \leq t_i \leq T_{in}$) and $-\frac{n-1-(k-1)\beta}{n-1}$ at the last period ($t_i = 1$) ($\frac{n-1-(k-1)\beta}{n-1}$ shows the probability to match with a peer in the normal phase), whereas he could enjoy payoffs of $\alpha$ for $2 \leq t_i \leq T_{in}$ and of one at $t_i = 1$ from in-group matches if he did not defect. The incentive constraint for 3-1d (a) is

$$\sum_{t=1}^{T_{in}} \delta^t (1-p) \alpha + \delta^{T_{in}} (1-p) \left( 1 + \frac{n-k}{n-1} \beta \right) + \delta \frac{1 - \delta^{T_{out}}}{1 - \delta} p \geq \alpha - 1,$$

where the first term in the LHS shows the loss by peer punishment for $2 \leq t_i \leq T_{in}$, the second the loss by peer punishment at $t_i = 1$, and the third the loss by ethnic conflict.

In case 3-1d (b), $i$ will have to bear expected in-group payoff of $-\frac{n-1-(k-1)\beta}{n-1}$ for each period during $1 \leq t_i \leq T_{in}$, whereas he would gain the payoff of one in each period if not defected. Therefore, the constraint for 3-1d (b) is

$$\sum_{t=1}^{T_{in}} \delta^t (1-p) \left( 1 + \frac{n-k}{n-1} \beta \right) + \delta \frac{1 - \delta^{T_{out}}}{1 - \delta} p \geq \alpha - 1.$$

Constraints 3-1d of (a) and (b) can be summarized as the following constraint 3-1d:

$$I(T_{in} > 1) \delta \frac{1 - \delta^{T_{in}-1}}{1 - \delta} (1-p) \min\{1+\frac{n-k}{n-1} \beta, \alpha\} + \delta^{T_{in}} (1-p) \left( 1 + \frac{n-k}{n-1} \beta \right) + \delta \frac{1 - \delta^{T_{out}}}{1 - \delta} p \geq \alpha - 1.$$
For 3-2d, the benefit from the deviation is \( \alpha - 1 \), and the loss comes both from collective peer punishment and ethnic conflict. As 1-2b in Lemma 1, the loss is minimized when \( t_i = T^{in} \). The effect of \( i \)'s defection on his peers is at least (a) and at most (b) shown above. In case (a), the defection does not change \( i \)'s payoffs for \( 2 \leq t_i \leq T^{in} \), and it creates the loss of \( 1 + \frac{n-k}{n-1} \beta \) at \( t_i = 1 \). In case (b), the defection increases \( i \)'s payoffs by \( \frac{k-1}{n-1} \beta \) for each period during \( 2 \leq t_i \leq T^{in} \) (since the defection sends other \( k-1 \) peers into the punishment phase, weakening the punishment on \( i \)) and decreases by \( 1 + \frac{n-k}{n-1} \beta \) at \( t_i = 1 \). So, the constraint in case (b) is more restrictive than in case (a). Constraint 3-2d will be

\[
I \left( T^{in} > 1 \right) \delta \frac{1 - \delta^{T^{in} - 1}}{1 - \delta} (1 - p) \frac{k - 1}{n - 1} (-\beta) + \delta^{T^{in}} (1 - p) \left( 1 + \frac{n-k}{n-1} \right) \beta + \delta \frac{1 - \delta^{T^{out}}}{1 - \delta} p \geq \alpha - 1,
\]

which suffices Constraint 3-1d since the first term of the LHS is non-positive. Constraint 3-2d is most relaxed when \( T^{in} = 1 \). So, as in Lemma 1, it is possible to take \( T^{in} = 1 \) for \( \alpha \leq 1 + \beta \). Q.E.D.

**Proof of Proposition 1.** For the strategy profile \( \sigma_4 \) to constitute a SPNE, the following two constraints 4-1d and 4-2d must hold, instead of 3-1d and 3-2d in \( \sigma_3 \).

The constraint for a cooperator in out-group matches (4-1d) requires

\[
\left( r \left( m \right) \sum_{t=1}^{T^{in}} \delta^t (1 - p) \left( q_{t+t} \left( 1 + \beta \right) + (1 - q_{t+t}) \alpha \right) + \delta \frac{1 - \delta^{T^{out}}}{1 - \delta} p \right) \geq \alpha - 1.
\]

For a defector to cooperate in out-group matches, because the loss of \( r \left( m \right) \sum_{t=1}^{T^{in}} \delta^t (1 - p) \left( q_{t+t} \left( 1 + \beta \right) + (1 - q_{t+t}) \alpha \right) + \sum_{t=1}^{T^{out}} \delta^t p \) is minimized when \( t_i = T^{in} \) and \( q_{t+T^{in}} = 1 \) by the same way as in 1-2b, the constraint for 4-2d is

\[
\left( r \left( m \right) \delta^{T^{in}} (1 - p) \left( 1 + \beta \right) + \delta \frac{1 - \delta^{T^{out}}}{1 - \delta} p \right) \geq \alpha - 1,
\]

which is sufficient for 4-1d.

For out-group cooperation to be beneficial to each group, it must be that \( m^* \leq p \). Q.E.D.

**Ex ante expected payoffs.** For \( \sigma_5 \), let the per-period ex ante expected payoff \( V_5 \) be divided into the one from the one in-group matches \( 1 - p \) and the other from out-group matches \( V^{out} \left( V_5 = (1 - p) + V^{out} \right) \). The ex ante out-group payoff (not per-period averaged) can be shown as \( V^{out} = p \left( (1 - \varepsilon) - \varepsilon \beta \right) + \delta \left( F_k (Q) \right)^2 V^{out} \)
\[ + (1 - F_k(Q)^2) \delta^{T_{out}} \hat{V}_{out}. \]

So, the per-period out-group payoff is

\[ V_{out} = (1 - \delta) \hat{V}_{out} \frac{(1 - \delta) p [(1 - \varepsilon) - \varepsilon \beta]}{1 - \delta \left(F_k(Q)^2 + (1 - F_k(Q)^2) \delta^{T_{out}} \right)}. \]

For \( \sigma_6 \), let the per-period ex ante expected payoff be \( V_6 = V_{6i}^{in} + V_{out} \), and let \( \hat{V}_{6i}^{in} \) denote the ex ante in-group payoff (not per-period averaged). We consider three cases of preserving peace and entering conflict: (x) ethnic conflict is triggered by group \( I \) (\#\( D_I > Q \)), (y) conflict is triggered only by \( J \) (\#\( D_J > Q \) and \#\( D_I \leq Q \)), (z) the peace phase is preserved (max\{\#\( D_I, \#D_J \} \leq Q \).

In case (x), \( i \)’s continuation payoff from the next period and after will be \( \delta (1 - p) \left\{ (1 - p) \left( \frac{k}{n-1} \alpha + \frac{n-k-1}{n-1} \right) + p \frac{n-k}{n-1} + \sum_{t=2}^{T_{out}} \delta^t (1 - p) \right\} + \delta^{T_{out}+1} \hat{V}_{6i}^{in} \), in which he receives per-period in-group payoff of \( \frac{k}{n-1} \alpha + \frac{n-k-1}{n-1} \) if he matched with a peer in the last period (that happens with probability \( 1 - p \)), while his per-period in-group payoff is \( \frac{n-k}{n-1} \) if he matched with an outsider in the last period (that happens with probability \( p \)). In case (y), \( i \)’s continuation payoff from the next period and after will be \( \sum_{t=1}^{T_{out}} \delta^t (1 - p) + \delta^{T_{out}+1} \hat{V}_{6i}^{in} \). In case (z), his continuation payoff from the next and after will be \( \delta \hat{V}_{6i}^{in} \).

By using facts that cases (x), (y) and (z) happen with probabilities \( 1 - F_k(Q) \), \( F_k(Q) \) (\( 1 - F_k(Q) \)), and \( F_k(Q)^2 \), respectively, the ex ante in-group payoff can be shown as

\[ V_{6i}^{in} = (1 - \delta) \hat{V}_{6i}^{in} \]

\[ = (1 - \delta) (1 - p) \left[ (1 + (1 - F_k(Q)) \right] \]

\[ \left\{ \delta \left\{ (1 - p) \left( \frac{k}{n-1} \alpha + \frac{n-k-1}{n-1} \right) + p \frac{n-k}{n-1} \right\} + \sum_{t=2}^{T_{out}} \delta^t (1 - p) \right\} + \delta^{T_{out}+1} \hat{V}_{6i}^{in} \]

\[ + F_k(Q) (1 - F_k(Q)) \frac{1 - \delta^{T_{out}}}{1 - \delta} \}. \]

For \( \sigma_7 \), the per-period ex ante expected payoff is \( V_7 = 1 - p + V_{out} - m_I \).

**Lemma 3** (i) The strategy profile \( \sigma_5 \) is a SPNE if and only if \( p < \frac{1}{1+\beta} \), \( \delta (1 - p) (1 + \beta) \geq \max\{\alpha - 1, \beta\} \), and

\[ (F_k(Q) - F_{k-1}(Q - 1)) F_k(Q) \delta \frac{1 - \delta^{T_{out}}}{1 - \delta} \geq (1 - \varepsilon) (\alpha - 1) - \varepsilon \beta. \quad (A1) \]
(ii) For \( \sigma_6 \) to be a SPNE, Inequality (A1) is replaced by

\[
(F_k (Q) - F_{k-1} (Q - 1)) \delta \left[ (1 - p) \left( 1 + \frac{n - k}{n - 1} \beta \right) + F_k (Q) \frac{1 - \delta^{T_{\text{out}}}}{1 - \delta} (V_{6}^{\text{in}} - (1 - p) + V_{\text{out}}) \right] \geq (1 - \varepsilon) (\alpha - 1) - \varepsilon \beta.
\]

(A2)

**Proof of Lemma 3.** Because the incentive constraints for in-group cooperation are the same as in cases without noise, we will focus on constraints for out-group interactions.

(i) For \( \sigma_5 \), the per-period ex ante payoff from out-group interactions is \( V_{\text{out}} \) if a player chooses to cooperate with an outsider, whereas it is zero during the conflict phase. Thus, Constraints 2-1d,2d of \( \sigma_2 \) is replaced by the following constraint 5-1d,2d

\[
(F_k (Q) - F_{k-1} (Q - 1)) F_k (Q) \delta \frac{1 - \delta^{T_{\text{out}}}}{1 - \delta} [V_{\text{out}} - 0] \geq (1 - \varepsilon) (\alpha - 1) - \varepsilon \beta,
\]

where \((F_k (Q) - F_{k-1} (Q - 1)) F_k (Q)\) shows the change in probability of ethnic conflict caused by an intentional out-group defection.

(ii) For \( \sigma_6 \), we consider incentive constraints for out-group cooperation if player \( i \) is in the normal phase (6-1d) and if \( i \) is in the punishment phase (6-2d). For 6-1d, (x’) if group \( I \) triggers conflict, \( i \)’s continuation payoff from the next period and after will be \( \delta (1 - p) \frac{n - k}{n - 1} (-\beta) + \sum_{t=2}^{T_{\text{out}}} \delta^t (1 - p) + \delta^{T_{\text{out}} + 1} \hat{V}_{6}^{\text{in}} \), which differs from the continuation payoff in (x), because unlike in (x), it is already revealed that \( i \) matched with an outsider. The continuation payoffs for cases that conflict is triggered only by the other group \( J \neq I \) and that the peace phase is preserved are the same as those of (y) and of (z) above. If \( i \) deviates, probabilities of cases (x’), (y) and (z) change to \( 1 - F_{k-1} (Q - 1), F_{k-1} (Q - 1) (1 - F_k (Q)) \) and \( F_{k-1} (Q - 1) F_k (Q) \), respectively.
Therefore, the incentive constraint for 6-1d is \( V_{6|C^{out}} \geq V_{6|D^{out}} \), where

\[
V_{6|C^{out}} = (1 - \varepsilon) - \varepsilon \beta + (1 - F_k(Q)) \left[ \delta (1 - p) \left\{ \frac{n - k}{n - 1} \right\} \right. \\
+ \sum_{t=2}^{T_{out}} \delta^t (1 - p) + \delta^{T_{out} + 1} \left( \hat{V}_6^{in} + \hat{V}_6^{out} \right) \right] + F_k(Q) (1 - F_k(Q)) \\
\left. \sum_{t=1}^{T_{out}} \delta^t (1 - p) + \delta^{T_{out} + 1} \left( \hat{V}_6^{in} + \hat{V}_6^{out} \right) \right] + F_k(Q)^2 \left[ \delta \left( \hat{V}_6^{in} + \hat{V}_6^{out} \right) \right],
\]

\[
V_{6|D^{out}} = (1 - \varepsilon) \alpha + (1 - F_{k-1}(Q - 1)) \left[ \delta (1 - p) \left\{ \frac{n - k}{n - 1} \right\} \right. \\
+ \sum_{t=2}^{T_{out}} \delta^t (1 - p) + \delta^{T_{out} + 1} \left( \hat{V}_6^{in} + \hat{V}_6^{out} \right) \right] + F_{k-1}(Q - 1) (1 - F_k(Q)) \\
\left. \sum_{t=1}^{T_{out}} \delta^t (1 - p) + \delta^{T_{out} + 1} \left( \hat{V}_6^{in} + \hat{V}_6^{out} \right) \right] + F_{k-1}(Q - 1) F_k(Q) \left[ \delta \left( \hat{V}_6^{in} + \hat{V}_6^{out} \right) \right].
\]

It leads to Inequality (A2). Because \( T^{in} = 1 \), statuses of \( i \) and his peers do not affect \( i \)'s payoff (when \( i \) goes to the phase, peers' punishment phases will have expired), and Constraint 6-1d applies to 6-2d, too. Q.E.D.

**Proof of Proposition 2.** For \( \sigma_7 \), Constraint 4-2d of \( \sigma_4 \) is replaced by the following constraint for out-group cooperation (7-1d,2d)

\[
r(m) \delta (1 - p) (1 + \beta) + (F_k(Q) - F_{k-1}(Q - 1)) F_k(Q) \delta \frac{1 - \delta^{T_{out}}}{1 - \delta} V^{out} \geq (1 - \varepsilon) (\alpha - 1) - \varepsilon \beta.
\]

Because a group minimizes the monitoring cost, the smallest \( m \) that satisfies 7-2d is chosen as the equilibrium level of monitoring \( m^* \). Q.E.D.

**References**


