

Localized Ethnic Conflict and Genocide

ACCOUNTING FOR DIFFERENCES IN RWANDA AND BURUNDI

RAVI BHAVNANI

DAVID BACKER

Department of Political Science

University of Michigan

Building on Fearon and Laitin, who concede in-group policing could be exploited for genocidal purposes instead of moderating interethnic hostilities, the authors seek to explain variation in the scale of ethnic conflict, using data from Rwanda and Burundi. Their computational model assumes individuals vary in their propensity to engage in violence, form independent beliefs about others, and react to public messages about current levels of ethnic aggression. In addition, the dominant ethnic group is subject to genocidal norms—defectors who fail to participate in ethnic violence face sanctions. Their results demonstrate that (1) the scale of violence varies considerably across episodes; (2) interethnic conflicts are not structurally deterministic but rather reflect endogenous interactions; (3) interethnic trust influences patterns of conflict—communities exhibiting high degrees of trust generally experience intense violence that subsides rapidly, in contrast to the persistent, moderate violence characteristic of less trusting communities; and (4) stronger genocidal norms exacerbate ethnic violence.

Fearon and Laitin (1996) develop a social-matching game model for interethnic relations that occur outside the purview of the state. Within this framework, they consider whether a cooperative outcome can be sustained given distinct information asymmetries. Their central finding is that both in-group policing and the fear of spiraling violence are efficient institutional mechanisms to reduce the incidence of defection—and thus the frequency and intensity of ethnic violence—once play moves off the equilibrium path. In-group policing equilibria are marginally more robust in this regard than are spiral equilibria: whereas the former rapidly confine conflict, violence tends to escalate further under the latter (Fearon and Laitin 1996).

Apart from indicating that small groups are relatively more likely than large groups to develop in-group policing, Fearon and Laitin never establish which ethnic interactions are regulated by each of the mechanisms. This omission is notable given that the

AUTHORS' NOTE: We are grateful to Robert Axelrod, Pradeep Chhibber, Douglas Dion, John Freeman, Ken Kollman, and Steven Mazie for their constructive comments on previous drafts. Earlier versions of this article were presented at the 1998 annual meeting of the American Political Science Association, the 1998 Merriam Junior Masters Class at the University of Illinois, and the April 1999 Comparative Research Group seminar at the University of Michigan. We appreciate the input of those who attended these gatherings. All faults remain our responsibility alone.

JOURNAL OF CONFLICT RESOLUTION, Vol. 44 No. 3, June 2000 283-306
© 2000 Sage Publications, Inc.

TABLE 1
Analytical Frameworks

		Group A			
		<i>Spiral</i>	<i>In-Group Policing</i>	<i>Individual Propensities</i>	<i>In-Group Sanctions</i>
Group B					
Fearon and Laitin					
Spiral	Higher violence				
In-group policing	Lower violence				
Bhavnani and Backer					
Individual propensities with high trust					Extreme violence
Individual propensities with low trust				Limited violence	

relative prevalence of these two institutional solutions implicitly affects the nature of the ethnic violence one would expect to observe. Equally significant, Fearon and Laitin acknowledge—but do not theorize—that in-group policing could backfire if the leaders of one ethnic group decide to eliminate members of a rival group instead of seeking to curb conflict and maintain cooperative relations.

We analyze this situation as a means of addressing the question of why the scale of violence varies substantially across episodes of ethnic conflict. Specifically, we consider the scenario of a state controlled by a politically dominant ethnic group conveying a message to the public that this group is engaged in aggressions against its rivals. We then explore the possibilities of interethnic violence when such messages are transmitted. As part of our effort, we evaluate several parameters to assess the relationship between individual-level attributes and the outcomes of these interethnic encounters. Our analysis highlights the “dark side” of in-group policing—when such norms are exploited for genocidal killing instead of being employed to moderate ethnic hostilities. By contrast, Fearon and Laitin limit their analysis to two ideal types of institutional arrangements; in both settings, individual behavior is driven purely by opportunism. Table 1 compares the respective analytical frameworks.

In Section 2, we present our basic empirical puzzle—variation in the scale of ethnic violence—based on episodes of ethnic conflict in Rwanda and Burundi. A comparison of cases leads us to discount the empirical explanations offered by Prunier (1995) and Lemarchand (1996), highly regarded political historians of Rwanda and Burundi, respectively. In Section 3, therefore, we explore the theoretical framework advanced by Fearon and Laitin (1996) and suggestive recent work on collective violence by Gould (1999). Although their approaches provide a foundation for our analysis, we propose several modifications to develop insights that are at once more general and consistent with our empirical backdrop.

In Section 4, we develop a computational model that simulates the behavior of members of two rival ethnic groups during episodes of conflict. This methodology, which we apply for the first time to the study of large-scale ethnic violence, enables us to introduce variation in attributes over time and across individuals as well as to reflect the uncertainties of their interactions.

In Section 5, we describe our major findings. To begin with, we discover that the scale of violence varies substantially across episodes even with reasonable default parameter values. We subsequently skew three key parameters—levels of interethnic trust, genocidal norms, and noise in the transmission of messages—to assess their association with episodes that exhibit extreme violence. These factors are not novel concepts; however, they have not been used to explain differences in the scale of ethnic violence or used together to construct a model of such conflicts. In Section 6, we provide examples from Rwanda and Burundi that illustrate how the tendencies we detect operate in practice. Finally, in Section 7 we offer some concluding thoughts about the significance of our analysis, potential applications, and likely avenues of future research.

THE EMPIRICAL PUZZLE

In August 1988, following a Hutu uprising in the northeastern communes of Marangara and Ntega, during which armed bands killed hundreds of Tutsi civilians, the Burundian army massacred a reported 20,000 Hutu in those localities.¹ The instinct might be to describe this event as simply the latest round in interethnic warfare that has characterized the Great Lakes region since the postcolonial era began in the late 1950s. On further examination, however, one detail stands out in stark relief: the massacre, although genocidal in nature, remained geographically confined and limited in scale.² The outcome contrasts with a 1972 episode during which the Burundian army, in response to a similar uprising, carried out a nationwide extermination of Hutu elites, with a death toll of between 80,000 and 200,000. Why didn't the 1988 reprisals involve a more comprehensive effort by Tutsi to eliminate their ethnic rivals, as transpired in 1972?

This puzzle is replicated in neighboring Rwanda, albeit with the sequence of episodes inverted. The instance of nationwide genocide has been extensively publicized: the nightmarish April, May, and June of 1994 during which approximately 500,000 to 800,000 Tutsi were killed by civilian militias and the Hutu-dominated army following the assassination of President Juvénal Habyarimana. Less well known are the prior

1. Estimates of deaths from ethnic violence are notoriously uncertain. The accounting process is especially difficult when a segment of the population has been displaced and the infrastructure has been severely damaged, as is characteristic of large-scale episodes. Thus, we provide ranges—based on established sources—whenever possible.

2. By "genocidal," we mean killings committed with an evident intent of eliminating members of an ethnic, racial, or religious group, the standard specified in the United Nations Convention on the Prevention and Punishment of the Crime of Genocide. Localized ethnic massacres, however, do not conform to definitions of genocide that incorporate criteria of large-scale and systematic scope. In making this distinction, we do not intend to diminish the seriousness of these incidents; rather, our goal is merely to differentiate related classes of events.

TABLE 2
Select Episodes of Ethnic Violence in Rwanda and Burundi

<i>Event (classification)</i>	<i>Sequence of Events</i>	<i>Scale and Scope of Violence/Displacement</i>
Episode 1: 1963-1964 Rwanda (localized massacre)	The Rwandan army stymies an attack by Tutsi exiles after the rebel forces near Kigali. The government responds by ordering massacres of Tutsi, targeting politicians.	10,000-13,000 Tutsi killed; 150,000 Tutsi refugees
Episode 2: 1972 Burundi (genocide)	After a failed coup attempt by Hutu politicians, Hutu unrest erupts in Bujumbura (capital), Cankuzo (east), and Nyanza Lac and Rumonge (south), during which rebels kill Tutsi and Hutu detractors. The Tutsi-dominated army and Tutsi civilians eliminate Hutu, targeting educated elites.	80,000-200,000 Hutu killed nationally; 150,000-300,000 Hutu refugees
Episode 3: 1988 Burundi (localized massacre)	Elections that fail to produce a change in political authority prompt Hutu unrest. Tensions are highest in Marangara and Ntega, where army patrols spark Hutu violence against Tutsi. The army responds by massacring Hutu civilians.	20,000 Hutu killed in Marangara and Ntega; 50,000 Hutu refugees
Episode 4: 1994 Rwanda (genocide)	President Habyarimana, returning from signing an agreement to implement the Arusha Accords, is shot down over Kigali; President Ntarayamira of Burundi is also killed. Aided by civilian militias, the army massacres Tutsi and moderate Hutu.	500,000-800,000 Tutsi killed nationally; 105,000 Tutsi refugees

instances of localized violence observed since Rwanda gained its independence in 1961. One notable episode, from December 1963 to January 1964, resulted in the deaths of between 10,000 and 13,000 Tutsi civilians at the hands of the Rwandan army. The juxtaposition of events elicits a similar enigma: why didn't the violence escalate further in 1963-1964, as it would several decades later, and result in a more systematic effort to eliminate the perceived provocateurs, in this case the Tutsi minority?

Table 2 offers further details on these episodes, which we culled from the numerous incidents of violence in Rwanda and Burundi since 1959.³ We use this empirical background to assist our investigation into the causes of variation in the scale of ethnic violence.

These countries are natural candidates for exploratory analysis because the similarities between them—their ethnic divisions, demographics, economic contexts, and so forth—facilitate comparison. The cases share another key characteristic: the ethnic group in power is reacting to a provocative event attributed to its rivals. Under such circumstances, there is an objective reason to expect reprisals. Despite the similarities, our cases exhibit two distinct outcomes: in each country, one episode concluded with

3. We synthesized the details on the episodes from Prunier (1995) and Lemarchand (1996), as well as Block (1994), Chrétien (1996), Edwards (1997), Keane (1995), Lemarchand (1992), Reyntjens (1995), and Smyth (1994).

localized massacres, whereas the other resulted in systematic genocide. Our primary objective, therefore, is to account for these differences in scale and scope.⁴

We initially returned to the empirical literature in search of a compelling explanation. Prominent histories of Rwanda and Burundi suggest several major reasons for variation in the scale of ethnic violence. First, Lemarchand's (1996) interpretation of the massacres by the Burundian army in 1988 is that comprehensive reprisals were considered unnecessary so long as the threat of Hutu unrest remained local and low grade.⁵ Yet, the 1972 uprisings, which involved only marginally greater organization than the 1988 unrest, prompted far more extensive and considerably more methodical reprisals.⁶ This evidence challenges the premise that extreme violence is necessarily precipitated by provocations that are especially threatening to the establishment, despite the obvious political overtones to these sorts of conflicts.

Second, Lemarchand (1996) also posits that the 1988 episode remained confined because the Tutsi elite expected that more extensive violence would have serious international repercussions, unlike in the case of the 1972 genocide. He attributes the change in their outlook to advances in human rights monitoring and media coverage.⁷ Yet, the 1994 genocide in Rwanda supplies a counterfactual: despite exposure and condemnation of events by foreign representatives, journalists, and activists—including many who were on the scene—the international community evidently was not treated as an obstacle.⁸

Finally, Prunier (1995, 237-48) attributes the massive scale and systematic scope of the 1994 genocide in Rwanda to the elaborate "final solution" developed by high-ranking Hutu extremists. He deduces that this agenda enabled the perpetrators to respond with such speed and efficiency that they overwhelmed efforts at resistance or escape. Two pieces of evidence cast a shadow, however, on this admittedly intuitive account of the preconditions for such extreme violence. One is the apparent absence of a detailed plot prior to the Burundian genocide of 1972.⁹ Thus, advance planning may not be necessary to accomplish systematic, nationwide killings, even if some measure of coordination and synchronization—as well as politicization of ostensible ethnic differences—is implicit in the collective nature of the event. In addition, the genocidal agenda conceived in Rwanda—far from creating an absolute information asymmetry

4. This article is also motivated by a normative concern about events in Rwanda and Burundi. We hope that a better understanding of the dynamics of episodes of ethnic conflict can contribute to efforts to prevent their recurrence.

5. Lemarchand (1996) suggests that Tutsi leaders were satisfied simply to retaliate against Hutu in the two localities where unrest had spontaneously erupted. Chrétien (1996) makes a similar argument.

6. The uprisings in the different regions may have been coordinated, but the movement is characterized as being too fragile and diffuse to pose any conceivable threat to the Micombero regime. See Lemarchand (1996, 93).

7. See Lemarchand (1996, 128-30).

8. Prunier (1995) indicates that the genocide occurred partly because of the lack of intervention by the international community; he highlights the reluctance of the United States to become engaged in the conflict. Yet, foreign military forces were on the scene in Rwanda, and there is no indication that the perpetrators anticipated this sort of muted response.

9. A common Hutu myth imputes responsibility for these killings to Burundi's Minister of Foreign Affairs Arthemón Simbananiye, who is alleged to have masterminded efforts to provoke a Hutu uprising to justify reprisals designed to achieve ethnic parity between (minority) Tutsi and (majority) Hutu. See Lemarchand (1996, 27).

that gave perpetrators an insurmountable advantage over their prospective victims—had likely been common knowledge for several months prior to April 1994.¹⁰ For example, a Tutsi refugee described to Keane (1995) how her family

heard the news about the president's plane from some Tutsi neighbors who had heard the story on the radio. Everyone knew it would be bad. The militias had been training in the area for a long time, months and months. The Tutsis knew about list of people that were going around, lists of Tutsis who were to be killed. These names were being circulated among the Interahamwe and the police for weeks ahead of the plane crash.¹¹ (p. 88)

Thus, advance planning is not altogether sufficient: one must account for the Tutsi who opted not to act on the information at their disposal, leaving themselves vulnerable, however unwittingly, to attack.¹²

In sum, we are inclined to discount these empirically based justifications of the variation in the scale of ethnic violence. The nature of the provocation, expectations as to the international response, and the extent of advance planning may contribute in some regard to the outcome of particular episodes. Yet, a careful comparison of relevant cases indicates that none of these factors provides a consistent rationale for the qualitative and quantitative differences between localized ethnic conflict and genocide.

EXISTING THEORETICAL MODELS

Seeking a coherent, comprehensive solution to our empirical puzzle, we turn instead to the innovative theoretical frameworks proposed by Fearon and Laitin (1996) and Gould (1999). Whereas much of the ethnic conflict literature concentrates on the origins of collective identity and the sources of intergroup hostility, these two studies offer useful insights concerning the dynamics of episodes of conflict. We use their work as a foundation for our analysis, which emphasizes factors they consider only tangentially or in a manner that is largely incompatible with our cases.

10. The personal accounts are substantiated by reports from UN observers as early as January 1994 "that Hutu leaders were mobilizing to slaughter minority Tutsis and moderate Hutus." See Metzl (1997, 15).

11. At the very least, radio broadcasts ensured that as the plan was being set into motion, many Rwandan Tutsis would likely have understood that they were targets for extreme reprisals by the Hutu militias:

On 6 April, the day of the plane crash, Radio Mille Collines told its audience that 'Tutsis need to be killed.' The theme was to dominate the station's broadcasts for weeks. The official state radio was only marginally less virulent, constantly calling on the Hutus to rise up and defend Rwanda against the invasion of the *inyenzi*. . . . Several privately owned newspapers and journals were harnessed for the task of disseminating hate propaganda.

See Keane (1995, 10). *Inyenzi* (cockroaches) is a derogatory term among the Hutu for the Tutsi rebels.

12. Based on the ominous signals, many Tutsis did take evasive measures. Thousands relocated as refugees to Burundi, Tanzania, Uganda, and the former Zaire (see Table 1), and numerous others went underground within Rwanda itself. Given the examples—both in this setting and elsewhere—of people taking flight despite the difficulties and losses this entails, socioeconomic status, ties to home, and other common family circumstances provide an inadequate explanation for the failure to escape the apparent threat of genocidal violence.

FEARON AND LAITIN'S SOCIAL-MATCHING GAME MODEL OF ETHNIC CONFLICT

In a significant departure from convention, Fearon and Laitin (1996) argue that "peaceful and cooperative relations between ethnic groups are far more common than is large-scale violence" and thus seek to "account for ethnic violence without overpredicting its occurrence." The key to understanding why relatively few episodes of violence transpire is the "decentralized, nonstate institutions [that] often arise to mitigate problems of opportunism in interactions between individuals of different ethnic groups" (p. 715). Their social-matching game model of ethnic conflict substantiates that these collective institutions can engender two types of nonviolent outcomes. The first type results from *spiral equilibria*, under which tacit cooperation between ethnic groups is induced by their mutual fear of conflict escalating out of control. The second type results from *in-group policing equilibria*, which rely on common expectations that transgressions by members of other ethnic groups will be observed and punished by their compatriots. Both equilibria implicitly moderate the frequency and scale of episodes of ethnic conflict.

We have several reservations about Fearon and Laitin's (1996) framework, particularly in light of our empirical reference points. The first concerns their assumption that members of an ethnic group have full information about fellow members but no information about ethnic outsiders.¹³ They offer two justifications for assuming this complete asymmetry: (1) within each group, there are numerous mechanisms for transmitting information, ranging from formal institutions to informal rumor and gossip; and (2) interaction is frequent within, but not across, groups. Both presuppositions are unusual given their concern that "existing rationalist and psychological theories of ethnic conflict are premised on assumptions about group-level demands, grievances, and animosities [and] . . . tend to treat groups as actors or implicitly equate group motivation with that of representative members" (p. 731). Moreover, Fearon and Laitin explore the consequences of encounters that are not anonymous: in their model, individuals can reliably identify those who cheat or exploit them as ethnic outsiders. The fact that members of one ethnic group can distinguish members of other ethnic groups requires some familiarity and thus frequency of interaction. Yet, Fearon and Laitin still assume that interethnic interactions are rare and population segments are homogeneous, premises that ultimately support their two moderate equilibria. Even leaving aside issues of theoretical consistency, their assumptions are not compatible with the cases of Rwanda and Burundi, where ethnic heterogeneity is the norm.

Second, Fearon and Laitin acknowledge that self-policing could have brutal consequences: a group might employ sanctions to induce its members to participate in attacks against its rivals rather than punish them for doing so. In-group policing equilibria could thereby be associated with high levels of violence, potentially well in excess of those observed under spiral equilibria. Fearon and Laitin fail, however, to describe the circumstances under which such reversals of function take place and to account for their impact on the outcomes of ethnic conflicts.

13. Fearon and Laitin allow that this is a simplification for the weaker assumption that information about fellow members is available at a lower cost than is information about ethnic outsiders.

Our third reservation concerns Fearon and Laitin's (1996) conclusion that moderating institutions inevitably develop because of the costs associated with violence. We question this assessment given that introducing a small amount of noise into their game leads to a breakdown of both types of equilibria. The sources of noise Fearon and Laitin allude to—"mistakes, misinterpretations, drunkenness, sudden passions, or unobservable variations in payoffs from encounter to encounter" (p. 723)—are hardly the exception in multiethnic societies such as Rwanda and Burundi. This observation suggests that the conditions required to foster peaceful ethnic coexistence may not prevail as universally as they contend.

Finally, we have qualms about Fearon and Laitin's answer to a fundamental question: why in some cases do interethnic relations remain cooperative for a long time but periodically lapse into spiraling violence, whereas in other cases peace is quickly restored after violence erupts? They argue that whereas in-group policing equilibria rapidly confine ethnic conflict, such violence tends to escalate further under spiral equilibria. Yet, they never establish why incidents spiral, as opposed to being regulated by in-group policing (apart from stating that small groups are more capable of policing). This omission is notable given that the relative prevalence of these two institutional solutions implicitly affects the frequency and scale of ethnic violence.

GOULD'S MODEL OF INTERGROUP FEUDS

Like Fearon and Laitin, Gould (1999) contends that existing studies of ethnic conflict typically overestimate the probability of group violence. He attributes the exaggeration to neglect of the inherent collective action problem: "Intergroup struggle is implicitly regarded as a sufficient condition for group participation in violent conflict" (p. 356). Instead, he observes, individuals engaged in a dispute may be unable to depend on their compatriots for support; as a result, the conflict cannot escalate into collective violence. More important, such moderate outcomes are feasible even when groups are involved at the outset of a dispute. In fact, Gould finds compelling evidence from nineteenth-century criminal trials in Corsica to support Fearon and Laitin's basic deterrence hypothesis: group contentions escalated to lethal violence less frequently than did the one-on-one disputes. The implication is that community involvement signals a threat of collective violence and thereby discourages intensification of conflict.

Gould (1999) argues, however, that deterrence is conditional on impressions of group unity: "expressions of group solidarity in the midst of a dispute may avert conflict escalation . . . only if people are convinced that the expressions are sincere" (p. 359). In other words, the viability of this mechanism is a function of reputation, that is, whether a group's members are perceived to be steadfast or unreliable. The latter circumstance is associated with the escalation of disputes, because any pledge to act collectively lacks credibility and must be backed up if the group hopes to employ expressions of solidarity to forestall future conflicts. Thus, "collective violence occurs when group action fails to convince an adversary to back down" (p. 356).

Our major reservation about Gould's (1999) framework is prompted by his observation that there is an inherent tension between group and individual interests, especially when loyalty entails risking one's life. At such a juncture, he notes, "even mod-

erately sensible members hesitate before joining a possibly fatal fray" (p. 359). Yet, in theory, solidarity must be sustained at some level—or even extended, on Gould's account—for mere group contention to progress to collective violence. (This requirement may be especially problematic if a group's members are perceived to be unreliable, which potentially has a corrosive effect on group solidarity.) Otherwise, any subsequent escalation of the dispute will tend to dissipate into scattered interpersonal violence. Although Gould explicitly represents this outcome in his model of intergroup feuds, he ultimately pays little attention to the potential for solidarity to unravel during these conflicts. Despite his repeated references to the collective action problem, the model lacks a clear mechanism—he alludes to compulsion based on communal pride or concerns about credibility—by which to counteract defections and nonparticipation, which we maintain are inevitable among large groups living in societies that are not strictly polarized. Consequently, Gould's conclusion that collective violence is conditional on collective contention is inadequate as an explanation for why certain group conflicts ultimately result in more extreme forms of violence than do others.

A NEW MODEL OF ETHNIC CONFLICTS

For our purposes, the crucial step is to identify factors that affect an individual's propensity to engage in ethnic violence. We start by assuming that individuals vary in their level of extremism and thus in the extent to which they have antipathy for nominal rivals or believe they pose a threat. In this regard, we follow Lohmann (1993), who establishes that the trajectory of violence during incidents of unrest depends on the distribution of nonparticipants, conditional participants, and willing participants within an assembled crowd.¹⁴ The underlying logic clearly applies to Rwanda and Burundi—where conflict does not strictly follow ethnic lines but instead has personal and political dimensions—and is a reasonably generic assumption.

We further assume that individuals have at least some information about the members of other ethnic groups. Rwanda and Burundi are both densely populated, heterogeneous societies, characterized by relatively frequent interactions (and even some intermarriage) between members of the two major ethnic groups. Consequently, there is no reason to expect dramatic information asymmetries, although the level of information will vary across individuals and be subject to periodic updates based on ongoing interactions. This knowledge, in turn, permits individuals to form independent beliefs about nominal ethnic rivals and the way they are likely to behave. Here, we follow Kandori (1992), who observes that individuals exhibit multiple behavioral attributes: for example, reputation can manifest independently from mere group affiliation.

14. Absent such an approach, one faces the selection bias that undermines many theories of ethnic violence: groups that constantly interact will exhibit more violence irrespective of how prone individual members are to attack those outside the group. See Dion (1997), especially p. 641. Given stable patterns of contact and a static assumption of antagonism between two groups, one cannot account for variation over time in the rates or levels of intergroup conflict. See Olzak (1992). For a similar assessment of Burundi, see Prunier (1995, 140).

Because these attributes inevitably provide a range of cues to others, the interaction between any two individuals—even ostensible ethnic rivals—is rarely deterministic.

We also assume that a catalyst is required for individuals to engage in ethnic violence. Following our empirical examples, the catalyst consists of a message that conveys to the entire population the severity of the ethnic violence transpiring in the country. We allow for the possibility that some individuals misinterpret the message—reflecting our stated reservations about the fragility of Fearon and Laitin's model in the face of noise. Such errors could result from a number of circumstances, such as poor lines of communication between the central government and local officials and suggestive appearances by national officials.

Finally, based on extensive anecdotal evidence, we entertain the prospect that the perverse form of self-policing we described earlier plays a key role in inducing more widespread participation in genocidal killings. We term this a genocidal norm, that is, an arrangement whereby participation in ethnic violence is sustained under threat of sanctions for noncompliance. Our use of this concept builds on the work on meta-norms by Axelrod (1986). He finds that the stronger the metanorm—because enforcement is more likely and/or involves more severe punishment—the greater the likelihood of sustaining a first-order norm.¹⁵

Thus, rather than take group solidarity for granted, we extend Gould's (1999) basic logic and accept that ethnic groups face a collective action problem even after a conflict is in progress. We then offer means by which groups might overcome the obstacles to maximizing the intensity of violence, which we assume supplants deterrence as the primary strategic objective in the midst of a conflict. The inclination to engage in violence initially depends on an individual's political extremism and perceptions of solidarity both within the group and among its rivals. These attributes are linked to the prospect of collective behavior by genocidal norms, which intervene as the formal control that may be necessary—as indicated by Hechter (1987)—for an ethnic group to overcome the inherent obstacles to collective action.¹⁶ Although we do not explicitly represent political leaders or other social elites in our model, the strength of genocidal norms serves as a proxy for the effectiveness of socialization processes that induce and enforce group solidarity.¹⁷

We use these assumptions to develop a computational model that addresses the basic puzzle of variation in the scale of violence across episodes of ethnic conflict.¹⁸ The next section offers a narrative of the sequence of events; this is followed by a for-

15. Axelrod's (1986) analysis focuses on the behavior of witnesses to crimes, but we envision that his findings extend to a setting where the first-order norm entails eliminating ethnic rivals. For another theorization of self-policing potentially engendering high-violence equilibria, see Coleman (1989). For further work on second-order norms, see Heckathorn (1990).

16. In particular, "formal controls are necessary for the attainment of solidarity in large groups" (Hechter 1987, 12). Furthermore, if the anticipated costs of ensuring compliance are low—such as when the members of a group are confident that others will act similarly—they may be able to coerce one another into participation. See Hardin (1995).

17. As we suggested in Section 2, the existence of a national-level agenda is not sufficient to ensure effective participation at the grassroots level. A potential avenue for future work is to consider alternative mechanisms—in addition to norms—for making this link.

18. Information about the computational model, which was programmed using Matlab v5.0 for Windows software, is available directly from the authors.

mal description of the behavior of the agents. We use this particular methodology, as opposed to the game-theoretic approach employed by Fearon and Laitin (1996), to introduce variation in individual characteristics such as the propensity to engage in violence, the expectation that nominal ethnic rivals will engage in violence, and the interpretation of messages. As a result of this variation, the effects of in-group sanctions are not uniform across agents. Moreover, our approach permits changes over time, whereas in Fearon and Laitin's model, the individual agents retain the same attributes throughout the course of play. Finally, in our model, individual actions reflect the uncertainty of current engagement rather than being based on full knowledge of the outcomes of prior cross-group pairings as in Fearon and Laitin's social-matching game.

SEQUENCE OF EVENTS

The model consists of individuals from two ethnic groups, A and B.¹⁹ An episode begins with a message, which is transmitted to all members of both groups, that ethnic violence has erupted on either a local or national level; each type has an equal probability of being sent out at this initial stage. A "local" message implies that violence is confined to particular regions of the country, whereas a "national" message implies that violence has spread throughout most regions.²⁰ Individuals then react to the message by engaging in or abstaining from violence.²¹

In particular, each member of group A either follows the ethnic cue (i.e., attempts to kill an ethnic rival) or refrains from doing so (i.e., spares ethnic rivals). Their actions are based on the type of message, their level of extremism (i.e., their propensity to engage in ethnic violence), and the strength of genocidal norms. As a result of these norms, defectors—members of group A who fail to attack members of group B when they are expected to—risk being punished by members of their own ethnic group. Each member of group B has three possible behaviors: follow the ethnic cue (i.e., attempt to kill an ethnic rival), refrain (i.e., trust nominal ethnic rivals as a function of their perceived reputation and thus stay put), or flee.²² Their actions are similarly based on the content of the message and their level of extremism, as well as on their beliefs about how the members of group A are likely to behave in response to the message.

19. One could conceivably include moderates from group A as a subset of group B, given that a number of genocidal regimes kill ethnic compatriots who collaborate with ethnic rivals. To retain the simplicity of our model, we have not made this distinction.

20. Our computational model does not explicitly incorporate a spatial dimension. Instead, we use the proportion of agents that are casualties of ethnic violence to gauge whether the violence is local or national in scope. In essence, the relative scale of violence serves also as a proxy for the territorial scope of the violence.

21. To keep things simple, our model of ethnic violence does not involve collective coordination (although we allow for the prospect of group-influenced behavior). Instead, we base mobilization on agents' independent responses—as a function of relevant individual attributes—to the public messages.

22. Flight does not guarantee survival, although (as we describe later) it reduces the probability of an individual being killed by members of the ethnic group in power. Moreover, this response implicitly reduces the extent of killing by the rival ethnic group. In subsequent work, we intend to address in greater detail the issue of flight from ethnic and political violence.

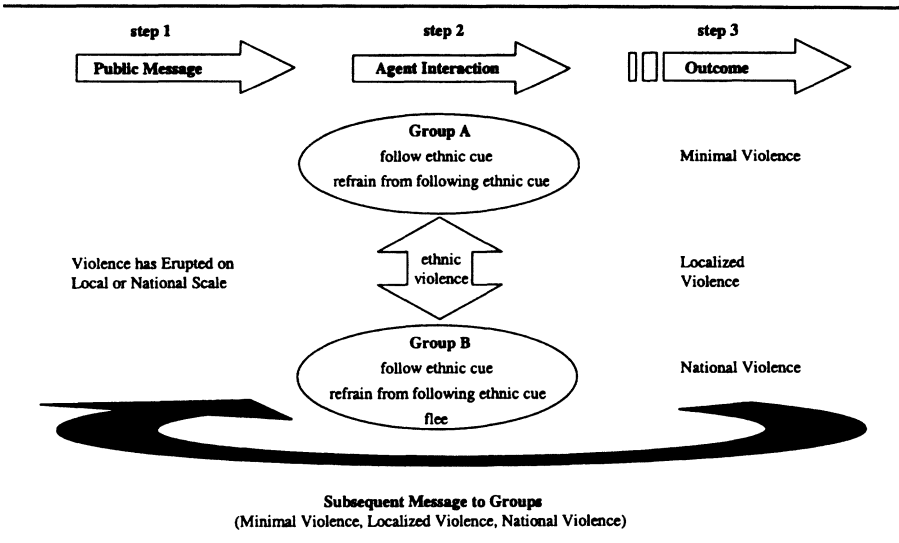


Figure 1: Sequence of Events in an Episode

Episodes—each composed of a maximum of 20 “events”—are structured so as to reflect the sequence of interactions that occur during incidents of ethnic conflict (see Figure 1). The participants in a given event are effectively a sample of the overall population. During an event, each participant acts just once—attacking, refraining, or fleeing. The aggregation of their behavior ultimately yields a measure of ethnic violence: the total number of deaths for an event.²³ The progression of an episode is dictated by the subsequent messages that are transmitted to new segments of the population, which in turn depend on the scale of violence that was observed during the prior events. If violence remains stable from event to event, the same message generally persists. Rising violence could lead the message to be upgraded from local to national, whereas declining violence could result in a downgrade of the message or in an end to hostilities.

AGENT BEHAVIOR

To reiterate, agents belong to one of two ethnic groups: group A or group B. As a matter of convenience, we let group A represent the ethnic group in power and group B the rival ethnic group. Depending on the circumstances, either group could be a minority. We define θ as the proportion of group A in the overall population.

At time t , every member of group A has a probability of following the ethnic cue (E) and of refraining from following the ethnic cue (R). The balance of probabilities depends on the type of message that is transmitted, the individual’s level of extremism

23. We use this measure of ethnic violence partly because we formulate our empirical puzzle in terms of unexplained variance in the number of deaths. The single measure also simplifies the model and facilitates comparisons.

$x_{ai} \sim U[0,1]$, and the strength of the genocidal norm. The norm dictates that any member a_i who fails to follow the ethnic cue when the message is national is killed by another member a_j with a probability of m . We assume norms do not apply in the case of a local message because there is no prevailing mandate that members of the ethnic group in power kill members of the rival ethnic group. In the base model, the strength of the norm is randomly drawn from a uniform distribution $m \sim U[0,1]$ at the start of each episode. In subsequent analysis, we narrow the range over which the norm may vary for a given episode by setting a lower bound ϵ , such that $m \sim U[\epsilon,1]$ where $\epsilon > 0$.

The effect of the genocidal norm—an increase in the probability of an individual following the ethnic cue—is inversely proportional to the individual's level of extremism. A strong norm will substantially increase a moderate individual's propensity to follow the ethnic cue but have a relatively minor effect on an extreme individual. This relationship is captured by making the propensity to follow the ethnic cue a convex combination of individual extremism and the effect of the norm: a_i follows the ethnic cue with a probability of $(x_{ai} + m - mx_{ai})$ and refrains with a probability of $(1 - x_{ai} - m + mx_{ai})$. If a_i plays E , he then kills a member of group B with a probability k_a .²⁴ If a_i plays R , then he does nothing. If, instead, the message received by a_i is local, then he plays E with a probability of x_{ai} and R with a probability of $(1 - x_{ai})$.

Each member of B follows the ethnic cue (E), refrains (R), or flees (F). Their actions reflect the type of message, their level of extremism $x_{bi} \sim U[0,1]$, and their belief y_i that members of group A will play E . Individual beliefs, which are randomly drawn from a uniform distribution $y_i \sim U[0,1]$ at the start of each episode, are uncorrelated with individual levels of extremism. If the message is national and $y_i > \beta$, then b_i always plays F . If $y_i \leq \beta$, then b_i plays E with a probability of x_{bi} (killing a member of A with a probability of k_b) and R with a probability of $(1 - x_{bi})$. Each decision to play F lowers k_a by a constant δ , whereas each decision to play R increases k_a by δ .²⁵ Alternatively, if the message received is local and $y_i > \beta$, then b_i plays E with a probability of x_{bi} (killing a member of A with a probability of k_b) and R with a probability of $(1 - x_{bi})$.²⁶ If $y_i \leq \beta$, then b_i always plays R . These criteria are summarized in Table 3. After each event, members of B also update their beliefs as follows: if the proportion of a_i who played E at time t exceeded (lagged) the expectations of b_i —defined as y_i —then b_i increases (decreases) y_i by a constant λ .²⁷

If the scale of violence $v < s$ at time t , where s is a threshold that distinguishes limited violence from other levels of conflict, then all agents are sent a message to cease

24. For each event, we initially set k_a to be 0.5, implying that any member of group A who plays E has a 50% chance of killing a member of group B.

25. We set δ at 0.0001, meaning that for every 100 members of group B who flee (refrain from following the ethnic cue) during an event, k_a is reduced (increased) for that event by 0.01. Our intuition behind this updating is that flight reduces the number of "easy targets," whereas staying put leaves individuals more vulnerable to violence.

26. We set k_b to 0.5, that is, there is a 50% chance that any member of group B will successfully kill a member of group A. The fact that $k_a = k_b$ implies that neither group has an inherent advantage in its capacity to kill members of the other group. We do not update k_b because members of group A do not engage in flight.

27. Unlike k_a , beliefs are updated across events. Our intuition is that there is learning across the independent samples of members of group B who participate in an episode. We set λ at 0.01; thus, the maximum an individual's beliefs can change during an episode is 0.2, implying a 20% higher expectation that members of A will play E .

TABLE 3
Actions by Members of the Rival Ethnic Group

Beliefs	Type of Message	
	National	Local
$y_i > \beta$	$p(\text{flight}) = 1$	$p(\text{ethnic}) = x_{bi}$ $p(\text{refrain}) = 1 - x_{bi}$
$y_i \leq \beta$	$p(\text{ethnic}) = x_{bi}$ $p(\text{refrain}) = 1 - x_{bi}$	$p(\text{refrain}) = 1$

hostilities, ending the episode. If $v > n$ at time t , where n is a threshold that distinguishes widespread, extensive violence from other levels of conflict, then the subsequent message sent to agents at time $t + 1$ is “national violence.” Otherwise, the subsequent message is “local violence.”²⁸

RESULTS

We seek to evaluate the effect of four parameters—the type of message that is initially transmitted to the public, the rival ethnic group’s level of trust in members of the ethnic group in power, the strength of the genocidal norms that are enforced by the ethnic group in power, and the level of noise in the transmission of local messages to members of the ethnic group in power—on the scale of ethnic violence that is observed during episodes of conflict. To provide a basis of comparison throughout our analysis, we construct a “base model” that employs values for these parameters that should not skew the outcomes. Specifically, the rival ethnic group is, on balance, divided equally between trusting and distrusting individuals, genocidal norms are drawn from a uniform random distribution, and messages are transmitted without noise. In addition, the two ethnic groups are the same size and, as we already detailed, their members are evenly distributed across the spectrum of extremism.

Using these default parameter values, the average level of violence for the base model—generated from a collection of 200 episodes—is 2,677 deaths per episode. A notable result is the substantial variation in the scale of violence across episodes: a standard deviation of 1,151 deaths. In other words, even with the same general configuration, one observes marked differences in the outcomes of these ethnic encounters. Given the way we designed the model, there is no indication that the disparities are structurally deterministic.

One concern, however, is that the scale of violence may be dictated by the message that is transmitted at the outset of an episode. Figure 2, which displays the results from

28. To distribute the outcomes generated by our model among the three categories (limited, local, national), we set s equal to 0.17 (i.e., 17% of participants killed) and n equal to 0.27 (i.e., 27% of participants killed).

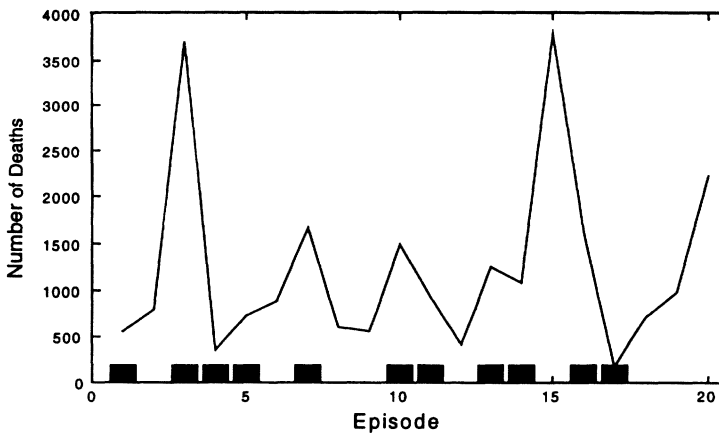


Figure 2: The Effect of the Initial Message on the Scale of Ethnic Violence

NOTE: Each episode, composed of a maximum of 20 events, used the following parameter values: $\beta = 0.5$, $m \sim U[0,1]$, Noise = 0, $A = 500$, and $B = 500$. The episodes with bars were initiated by a local message, whereas the episodes without bars were initiated by a national message.

a sample run of 20 episodes, indicates that the variance cannot be attributed simply to the initial messages. Although episodes that exhibit high levels of violence are generally more common when the initial message is national, such an extreme outcome is hardly guaranteed even under those circumstances. Some episodes initiated by national messages exhibit low levels of violence comparable to many of the episodes initiated by local messages. An episode initiated by a local message may also yield a high level of violence—the third episode in Figure 2 is a striking example. Thus, an initial national (local) message is neither sufficient nor necessary for an episode to generate a high (low) level of violence. Instead, the aggregate scale of violence reflects an endogenous process whereby individuals' reactions to the initial message determine the content of subsequent messages and the progression of an episode. Consequently, we have reason to investigate whether the other parameters of interest can be manipulated to induce levels of violence that are consistently higher than those we observed under the base model.

The first parameter we consider is the level of trust members of the rival ethnic group have in members of the ethnic group in power. As we discussed earlier, we assume individuals can develop independent beliefs about how nominal ethnic rivals are likely to behave during episodes of conflict. When surveyed across a population, these beliefs effectively translate to a measure of interethnic trust. In the model, we can vary this parameter by changing β : a higher (lower) value amounts to a decrease (increase) in the proportion of members of group B who believe that members of group A will engage in violence. We reason that a higher (lower) share of "trusting" members would make group B more (less) susceptible—due to a greater (smaller) incidence of individuals failing to flee or defend themselves—to unexpected attacks carried out or assisted by supposedly "trustworthy" neighbors. With more members of group B

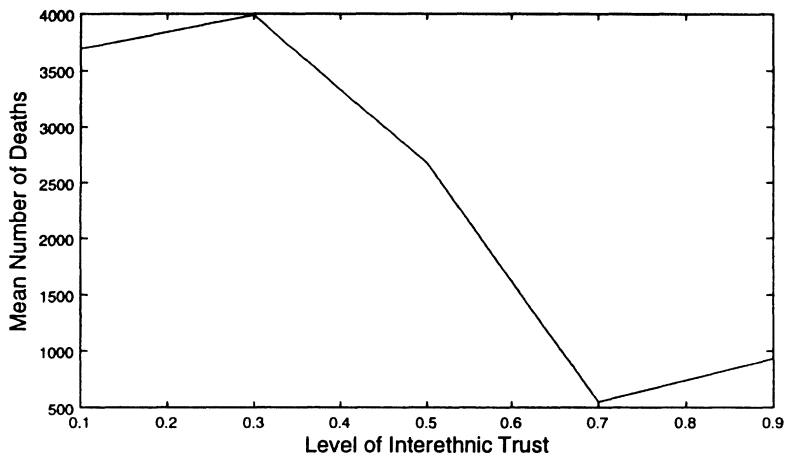


Figure 3: The Effect of Interethnic Trust on the Scale of Ethnic Violence

NOTE: These summary results are based on 200 episodes, each composed of a maximum of 20 events, using the following parameter values: $m \sim U[0,1]$, Noise = 0, $A = 500$, and $B = 500$.

being killed than is otherwise the case, we expect that greater trust will induce higher levels of violence.

Our results, depicted in Figure 3, contradict our original intuition: on balance, low levels of trust are actually associated with substantially greater violence than we observed under the base model, whereas high levels of trust result in substantially less violence. The primary explanation for this relationship is the different patterns in the duration of episodes, which the mean values (i.e., the average scale of violence per episode) effectively conceal. That is, the level of trust among members of the rival ethnic group affects their behavior to such an extent that it dictates whether episodes of violence are consistently sustained or instead dissipate.

To demonstrate this point, we display (see Figure 4) representative patterns of violence for each of the levels of interethnic trust we evaluate above: equal trust and distrust ($\beta = 0.5$), which corresponds to the base model; moderate distrust ($\beta = 0.3$); high distrust ($\beta = 0.1$); moderate trust ($\beta = 0.7$); and high trust ($\beta = 0.9$). On the x -axis, we plot 20 distinct episodes of ethnic violence. The y -axis plots the events that comprise each individual episode. The z -axis plots the scale of violence for each event within each episode.

With moderate or high levels of distrust, at least half of the episodes continue for the full term of 20 events. The intensity of violence is relatively normal, but aggregate levels of violence are high due to the long duration of episodes. In contrast, with moderate to high levels of trust, most episodes end within 10 events. Despite the fact that the episodes are often more intense while they last, the consistently shorter episodes translate into lower aggregate levels of violence. These results follow from the symmetry in the actions of the members of the rival ethnic group. Distrusting individuals are prone to flee whenever the message is national and engage in violence whenever the message is

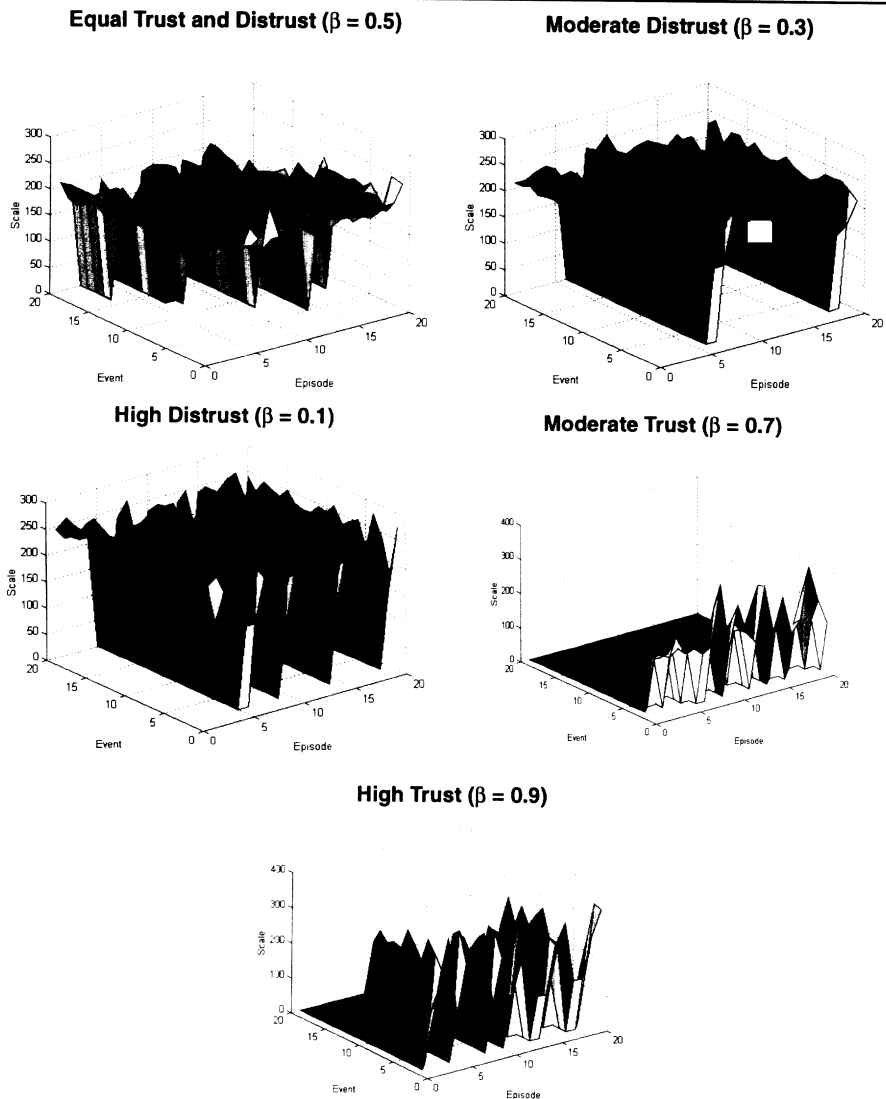


Figure 4: The Effect of Interethnic Trust on Patterns of Ethnic Violence

NOTE: Each episode, composed of a maximum of 20 events, used the following parameter values: $m \sim U[0,1]$, Noise = 0, $A = 500$, and $B = 500$.

local. Consequently, any episodes that exhibit national messages are at risk of not generating enough violence to sustain the episode, because relatively few members of group B are likely to contribute to the violence. Episodes that begin with or subsequently generate local messages have a greater chance of being sustained, because

more members of group B are likely to engage in violence, and the threshold for continuing at that modest level becomes easier to attain.

The same logic also accounts for the nonlinear relationship at the two extremes: increasing mistrust (trust) actually yields a decrease (increase) in the scale of violence. Moving from moderate to high levels of distrust lowers the probability that an episode will be sustained after an initial national message. On the other hand, moving from moderate to high levels of trust raises the probability that an episode will be sustained after an initial national message. This result suggests that the relationship between trust and violence is not necessarily straightforward. Within communities in which ethnic groups are particularly antagonistic—and their members are universally suspicious as a result—refugee flight can function as a built-in release valve that effectively limits the potential for violence. Meanwhile, multiethnic communities that are generally harmonious may experience unexpectedly severe episodes of conflict because most residents underestimate the threat of violence and yet get caught up in any clashes that erupt.

The second parameter we consider is the strength of the genocidal norm. In our base model, we allow these norms to vary on the unit interval, based on our assumption that members of the ethnic group in power always face some pressure from fellow members to follow ethnic cues during “national” episodes of violence. To strengthen the norms, we simply narrow the range over which the parameter m can vary.²⁹ As explained earlier, this corresponds to an increase in the certainty and/or severity of sanctions for nonparticipation.

As Figure 5 indicates, at low levels, this constraint has a minimal effect on the aggregate scale of violence. Our interpretation of this result is that the incremental influence of weak norms on the moderate members of group A (i.e., those with low probabilities of following the ethnic cue to begin with) is not substantial enough to measurably increase the overall level of violence. Instead, it appears that only a small number of conditional participants are prompted to join the fighting. The impact is more considerable when we limit the distribution of norms to the range $[0.5, 1]$: the mean number of deaths per episode is 14% higher than under our base model. Further limiting the range to $[0.75, 1]$ results in a mean value that is 24% higher than under our base model. These latter two results reveal that the definite risk of strong sanctions can induce levels of participation sufficient to engender more intense, lasting episodes of conflict.

The third parameter we consider is the level of noise in the transmission of “local” messages to members of A, such as might be caused by poor lines of communication between the central government and regional outposts or suggestive appearances by national officials.³⁰ Given the inherent uncertainty among members of B about the behavior of those in power—reflected in their varying levels of trust—we limit noise

29. This adjustment ensures that the effect of the genocidal norms—that is, the $(1 - x_i)m$ component of the formula that ultimately determines when a_i plays E —always equals or exceeds some nonzero lower bound.

30. We do not explore the scenario of national messages being transmitted with noise. We assume that the leaders of the ethnic group in power make a special effort under those circumstances to fully inform their members.

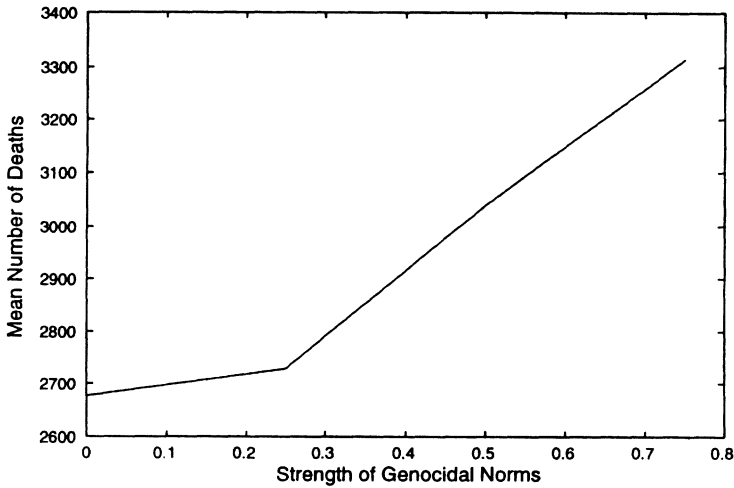


Figure 5: The Effect of Genocidal Norms on the Scale of Ethnic Violence

NOTE: These summary results are based on 200 episodes, each composed of a maximum of 20 events, using the following parameters: $\beta = 0.5$, Noise = 0, $A = 500$, and $B = 500$.

to members of group A.³¹ Our intuition is that noise should increase the scale of violence because any member of A who misinterprets the message in this fashion is, considering that sanctions for nonparticipation may be enforced, more apt to attack a member of B.

We initially introduce a 25% probability for each event that individual members of group A misinterpret a local message as a national message. This amount of noise results in a mean level of violence that is about 80% higher than under our base model (see Figure 6). Increasing noise beyond this point does not have an additional effect: the mean level of violence when there is a 50% probability remains essentially the same as when it is only 25%. This result could reflect the steep decline in local messages—and thus the higher incidence of flight by members of group B, which offsets group A's activity—generated by the model.

Our basic results are also robust with respect to the size of the ethnic group in power. Even if group A is the minority, under certain conditions the scale of violence can still exceed that of the base model. Suppose, for example, group A is half the size of group B (i.e., 250 members versus 500). With a moderate amount of noise and moderately strong genocidal norms, the mean scale of violence—based on 200 episodes—is 2,972 deaths per episode, or 11% higher.³² The joint effect of these factors is to ensure that a higher share of the ethnic group in power follows ethnic cues, counteracting its smaller size and the reduction in its capacity for violence.

31. Another potential avenue for future work is to evaluate the effects of noise on the interpretation of messages by members of both groups.

32. Each episode, composed of a maximum of 20 events, used the following parameter values: $\beta = 0.5$, $m \sim U[0.25, 1]$, Noise = 0.25, $A = 250$, and $B = 500$.

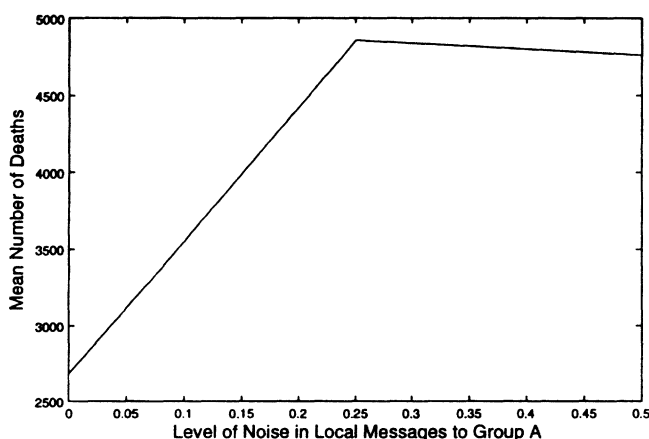


Figure 6: The Effect of Noise on the Scale of Ethnic Violence

NOTE: These summary results are based on 200 episodes, each composed of a maximum of 20 events, using the following parameter values: $\beta = 0.5$, $m \sim U[0,1]$, $A = 500$, and $B = 500$.

DISCUSSION

Our analysis addresses a basic empirical puzzle: Why in some cases does ethnic violence spiral out of control, whereas in other cases peace is quickly restored after violence breaks out? As we discussed earlier, Fearon and Laitin (1996) offer an inadequate explanation for this variation. They find that violence escalates further under spiral equilibria than under in-group policing equilibria but provide little indication as to when episodes are regulated by each of these mechanisms. Absent this link, their institutional solutions fail to account for disparities across episodes. Meanwhile, Gould (1999) attributes the escalation of interethnic disputes to challenges to group solidarity but does not clarify the circumstances under which the collective aspect of certain conflicts is sustained.

We construct a computational model that yields variance in the scale of violence across episodes. By focusing on individual extremism and beliefs, as well as the impact of genocidal norms on individual behavior, we demonstrate that structural factors (institutions, planning, etc.) need not be instrumental in dictating the outcomes of episodes of ethnic conflict. Furthermore, we find that prior levels of violence are inadequate predictors of future levels of violence. Instead, even with our base model—deliberately configured so as not to skew the results—variation in individual characteristics is seemingly sufficient to generate variation in outcomes. Consequently, the conventional preference for tracking structural factors may be overstated; our results imply that comparable attention should be afforded to individual-level dynamics.

To evaluate specific hypotheses concerning individual characteristics, we adjust several parameters of interest—trust, genocidal norms, noise—that derive from our reservations concerning the theoretical frameworks presented by Fearon and Laitin

(1996) and Gould (1999). As a result of these manipulations, we identify three factors that increase the scale of violence. First, we find that low levels of trust result in greater violence than do high levels of trust. These patterns initially struck us as counterintuitive. On closer examination, we ascertain that differences in aggregate levels of trust influence the patterns of violence that emerge. In communities characterized by high levels of interethnic trust, violence is likely to peak and then subside rapidly, whereas communities characterized by low levels of interethnic trust are more likely to experience moderate levels of violence that are sustained over longer periods of time. The results vindicate our hypothesis that variation in the scale of violence is affected by individuals acting on independent beliefs about the trustworthiness of nominal ethnic rivals rather than being equally distrusting of outsiders, as Fearon and Laitin assume.

A related result is the nonlinear relationship at high levels of trust—past a certain point, greater trust is actually associated with increasing (rather than decreasing) intensity and scale of violence. This finding can be viewed as consistent with our original hypothesis that higher levels of trust will be associated with increasing levels of violence because they render the rival ethnic group more susceptible to unexpected attacks. This phenomenon is afforded some support by events in Rwanda. Although distrust between Rwanda's rival ethnic groups is prevalent, examples abound of Tutsi "trusting" Hutu in ways that facilitated, however inadvertently, the genocidal killings in 1994. The most powerful evidence consists of cases in which Tutsi approached local Hutu officials to seek assistance and relied on whatever support or advice they provided, only to be sincerely surprised when the situation backfired. In one particularly tragic story recounted by Keane (1995), the "trusted" local official is directly implicated in the killings:

The Tutsis decided to form a deputation and they went to Rusomo to see the Bourgmestre, Sylvestre Gacumbitsi. . . . Without his help the Tutsis knew that it was only a matter of time before they were all killed. . . . He told them to go to the church and try to find safety there. . . . The word spread quickly and thousands of Tutsis from the surrounding area fled to the church at Nyarubuye. . . . It was not long before the militia arrived at the church and began to attack the refugees . . . of all the things about the massacre what she cannot believe is that Gacumbitsi actually came and directed the killing of the people. He was the Bourgmestre and he had organized the killing. (p. 90)

Clearly, these victims never suspected that Bourgmestre Gacumbitsi would take an active role in the violence. Many other Tutsi had favorable opinions about their Hutu neighbors and local officials or at least expected that they would behave in a neutral manner and observe certain boundaries. A standard refrain in the accounts is that the belief that churches would be safe havens proved to be the undoing for countless thousands of helpless Tutsi. Some church officials aided and abetted the militias; many were ultimately powerless to resist their violent intrusions.³³

33. Human rights groups reported that more Tutsi were killed in Rwanda's churches than in any other type of site. See, for example, African Rights (1994). The Catholic Church was known to have close ties to the government in Rwanda, whereas in Burundi the two were often at odds. See Chua-Eoan (1996, 49). The former relationship is confirmed by Prunier (1995, 251).

Second, we find that genocidal norms—that is, arrangements involving sanctions for those failing to participate in ethnic killings—lead to a marked increase in the scale of violence. This result supports our original hypothesis that Fearon and Laitin's (1996) in-group policing equilibria can indeed intensify, rather than simply mitigate, the scale of ethnic violence. The conclusion is also corroborated by events in Rwanda and Burundi. Prunier (1995) refers on several occasions to the threat of sanctions faced by those who failed to carry out their assigned roles. Similarly, Keane (1995) describes how Hutu neighbors of one Tutsi family “were kind and offered to help them. But they were threatened by others who promised to tell the Interahamwe that they were helping the Inyenzi” (p. 88).³⁴ In fact, one of the notable aspects of the 1994 genocide was the killing of some 10,000 to 30,000 Hutu moderates who refused to take part in the violence (Prunier 1995, 265). In the process, ethnic difference is effectively superseded by political divergence: moderates are deemed collaborators, and boundaries are redrawn with the objective of restoring group solidarity and preserving group supremacy. The Rwandan genocide shares this feature in common with the 1972 genocide in Burundi. Lemarchand (1996) reports that “the few Tutsis who tried to interpose did so at their own peril . . . the killing of Hutu seemed to have become part of the civic duty expected of every Tutsi citizen” (pp. 97-98).

Genocidal norms, therefore, provide a potent mechanism by which victims' otherwise reliable impressions about their neighbors are confounded, with such unexpectedly devastating results. Some of the perpetrators appear to have been reluctant participants, carrying out acts of violence only under the threat of death if they refused. The larger class of perpetrators—including those beyond just the core extremists—provides more personnel to execute the genocidal agenda and makes escaping more difficult for members of the targeted group. The threat of sanctions also extends the violence over a wider range of targets. Whereas some Hutu may have willingly redressed their grievances against particular Tutsi, under these conditions they were pressed to be indiscriminate. Moreover, the fact that such a pernicious variety of norms operates within ethnic groups is only exacerbated by the difficulty of identifying such forms of compulsion. The logical outcome of these factors is a more extensive and efficacious episode of ethnic violence.

Third, we find that even low levels of noise—leading some members of the ethnic group in power to misinterpret violence as being more far-reaching—substantially increase the scale of violence. This result bolsters our concerns about the fragility of Fearon and Laitin's (1996) two cooperative equilibria. Events in Rwanda and Burundi support our earlier contention that the conditions necessary to mitigate ethnic violence may not prevail as universally as they contend. Prunier (1995) describes one set of circumstances that suggests how misinterpretations may have contributed to the genocidal violence in Rwanda:

A common feature of all the massacres is that they were preceded by political meetings during which a “sensibilization” process was carried out. These seemed to have been designed to put the local peasants “in the mood,” to drum into them that the people they

34. The *Interahamwe* were the local Hutu militias that were organized beginning in 1991. See Block (1994) for a detailed discussion of their formation and activities.

were soon to kill were *ibiyitso*, i.e., actual or potential collaborators of the RPF arch-enemy. These meetings were always presided over and attended by the local authorities with whom the peasants were familiar; but they also usually featured the presence of an “important person” who would come from Kigali to lend the event an aura of added respectability and official sanction. (pp. 137-38)

Some of the Hutu who attended these gatherings took part in a series of massacres that preceded the 1994 genocide, perhaps under the misconception that a broader agenda was active. The presence of the national observer insinuates that violence is being coordinated at the center and is set to occur on a nationwide basis. If such suggestive appearances occur on a large scale and/or with some regularity, the likely effect is to precipitate an extreme episode of ethnic violence.

Finally, we demonstrate that even weakly enforced genocidal norms, when combined with moderate noise, can offset the minority status of the dominant ethnic group. This result clarifies how genocide was feasible in Burundi, where the ruling Tutsi constitute only 15% of the population. As we described previously, many Tutsi moderates were killed for failing to join or otherwise resisting the 1972 violence. Lemarchand (1992) also describes how Tutsi leaders had stoked fears of Hutu domination by associating their demands for fuller participation in political affairs with an effort to gain political ascendancy. These activities polarized ethnic relations to the point where the reprisals for an opportunistic Hutu uprising were exceptionally brutal and systematic.³⁵ Cultivating such antagonism, even if no formal program of genocidal violence is afoot, fosters more extensive and severe reactions that would ordinarily be confined to localized clashes.

CONCLUSION

This article offers a novel perspective on the subject of ethnic conflict, which afflicts almost every region of the world and preoccupies numerous scholars and policy makers. Our research is motivated by a compelling empirical puzzle—variation in the scale of violence across episodes—but also draws inspiration from our desire to promote efforts to minimize bloodshed by providing further insight into the dynamics of episodes of conflict. We believe the methodology of computational modeling provides useful leverage on these phenomena because of its ability to represent variation in individual attributes within ethnic groups and over time, as well as the uncertainties characteristic of interactions between nominal rivals. Equally significant, we generate predictions about the duration and magnitude of collective violence that can potentially be tested using various forms of empirical analysis, including event histories of recent ethnopolitical conflicts such as those in Chechnya, East Timor, and Kosovo.

At this stage, our model is a relatively simple abstraction of this general class of events, informed by circumstances in two of the notable flash points during recent

35. In this context, a tacit policy of denying Hutu scholarships for foreign study after 1968 effectively overlapped with periodic military and police operations that targeted Hutu politicians who were allegedly plotting to overthrow the government of President Michel Micombero. See Lemarchand (1992, 76-78).

years: Rwanda and Burundi. In the future, we intend to explore a series of modifications and extensions (e.g., examining protest cycles, introducing a spatial dimension, tracking flight) that we expect will enable us to depict and study these conflicts in a more comprehensive fashion.

REFERENCES

- African Rights. 1994. *Rwanda: Death, despair and defiance*. London: African Rights.
- Axelrod, Robert. 1986. An evolutionary approach to norms. *American Political Science Review* 80 (4): 1095-1111.
- Block, Robert. 1994. *The tragedy of Rwanda*. *The New York Review of Books* 41 (October 20): 3-8.
- Chrétien, Jean-Pierre. 1996. Burundi: The obsession with genocide. *Current History* 95:206-10.
- Chua-Eoan, Howard. 1996. Of death and defiance: Hate kills Burundi's religious fighter for peace. *Time*, 23 September, 49.
- Coleman, James. 1989. *Foundations of social theory*. Cambridge, MA: Harvard University Press.
- Dion, Douglas. 1997. Competition and ethnic conflict. *Journal of Conflict Resolution* 41:638-48.
- Edwards, Mike. 1997. Central Africa's cycle of violence. *National Geographic*, June, 124-33.
- Fearon, James, and David Laitin. 1996. Explaining interethnic cooperation. *American Political Science Review* 90 (4):715-35.
- Gould, Roger. 1999. Collective violence and group solidarity: Evidence from a feuding society. *American Sociological Review* 64:356-80.
- Hardin, Russell. 1995. *One for all: The logic of group conflict*. Princeton, NJ: Princeton University Press.
- Hechter, Michael. 1987. *Principles of group solidarity*. Berkeley: University of California Press.
- Heckathorn, Douglas. 1990. Collective sanctions and compliance norms: A formal theory of group-mediated social control. *American Sociological Review* 55 (3):366-84.
- Kandori, Michihiro. 1992. Social norms and community enforcement. *Review of Economic Studies* 59: 63-80.
- Keane, Fergal. 1995. *Season of blood: A Rwanda journey*. New York: Penguin.
- Lemarchand, René. 1992. *Burundi: The politics of ethnic amnesia*. In *Genocide watch*, edited by Helen Fein, chap. 6. New Haven, CT: Yale University Press.
- . 1996. *Burundi: Ethnic conflicts and genocide*. Cambridge, UK: Cambridge University Press.
- Lohmann, Susanne. 1993. A signaling model of informative and manipulative political action. *American Political Science Review* 87 (2):319-33.
- Metzl, Jamie. 1997. Information intervention: When switching channels isn't enough. *Foreign Affairs* 76 (November/December):15-20.
- Olzak, Susan. 1992. *The dynamics of ethnic competition and conflict*. Stanford, CA: Stanford University Press.
- Prunier, Gérard. 1995. *The Rwanda crisis: History of a genocide*. New York: Columbia University Press.
- Reyntjens, Filip. 1995. *Burundi: Breaking the cycle of violence*. London: Minority Rights Group.
- Smyth, Frank. 1994. Blood money and geopolitics. *The Nation* 258:585-88.