

**Localized Ethnic Conflict and Genocide:
Accounting for Differences**

Ravi Bhavnani
David Backer
Department of Political Science
7601 Haven Hall
University of Michigan
Ann Arbor, MI 48109

August 13, 1999

Localized Ethnic Conflict and Genocide: Accounting for Differences

Abstract¹

This paper seeks to explain the variation in the scale of violence across episodes of ethnic conflict, using data from Rwanda and Burundi. To do so, we explore the “dark side” of in-group policing—when it is exploited for genocidal killing, instead of being employed to moderate ethnic hostilities. Our efforts build upon Fearon & Laitin (1996), who concede this mechanism could backfire if an ethnic group announces its intent to attack, rather than cooperate with, a rival ethnic group. We depart from them in developing a computational model that assumes individuals vary in their propensity to engage in violence, form independent beliefs about nominal rivals, and respond to catalysts, namely messages about ethnic violence that has transpired. In addition, members of the politically dominant ethnic group are subject to metanorms, i.e., sanctions for non-participation. Given these reasonable assumptions, our model yields substantial variance in the scale of violence across episodes. We further demonstrate that (1) the interaction between nominal ethnic rivals is rarely deterministic and, consequently, an emphasis on structural factors is somewhat misguided; (2) changes in inter-ethnic trust influence the patterns of ethnic conflicts—communities exhibiting high levels of trust are more likely to experience intense violence that subsides rapidly, in contrast to the moderate violence that is sustained over a longer period of time in less trusting communities; (3) stronger metanorms engender more extensive episodes of violence.

1 Introduction

Fearon & Laitin (1996) develop a social matching game model for interethnic dynamics that occur outside the purview of the state. Within this framework, they consider whether a cooperative outcome can be sustained given certain 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 scale 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 ethnic conflict, violence tends to escalate further under the latter.²

Apart from indicating that small groups are more likely relative to large groups to develop in-group policing, Fearon & Laitin never establish which ethnic interactions are regulated by each of the mechanisms. This omission is notable given that the relative prevalence of these two institutional solutions implicitly affects the nature of the ethnic violence one would expect to observe. Equally significant, Fearon & Laitin acknowledge—but do not theorize—that in-group policing could backfire if the leaders of one ethnic group wish to eliminate members of a rival ethnic group, instead of seeking to moderate hostilities and maintain cooperative relations.

¹ We are grateful to Robert Axelrod, Pradeep Chhibber, Douglas Dion, John Freeman, and Ken Kollman for their helpful comments on previous drafts. Earlier versions of this paper 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 Comparative Research Group seminar at the University of Michigan (April 2, 1999). We appreciate the input of the participants at these various gatherings. The analysis originated from Ravi Bhavnani's participation at the 1998 Summer Economics Workshop at the Santa Fe Institute. He would like to thank John Miller, Scott Page, and the other workshop participants for their guidance. All faults remain our responsibility alone.

² James Fearon and David Laitin, “Explaining Interethnic Cooperation,” *American Political Science Review* 90(4) [December]: 715-735.

TABLE 1
Analytical Frameworks

		Fearon & Laitin		Bhavnani & Backer	
		GROUP A		GROUP A	
		<i>Spiral</i>	<i>In-Group Policing</i>	<i>Individual Propensities</i>	<i>In-Group Sanctions</i>
GROUP B	<i>Spiral</i>	Higher violence			Extreme violence
	<i>In-Group Policing</i>		Lower violence	Limited violence	
		GROUP B		GROUP B	
		<i>Individual Propensities w/ High Trust</i>		<i>Individual Propensities w/ Low Trust</i>	

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 in which a state, controlled by a politically dominant ethnic group, has the capacity to convey a message to the public that this group is engaged in aggressions against its rivals. We then explore the possibilities of inter-ethnic violence when such messages are transmitted. Thus, our paper looks at the “dark side” of institutionalized in-group policing—when it is exploited for genocidal killing instead of being employed to moderate ethnic hostilities. As part of our efforts to explain the variation, we evaluate three parameters (inter-ethnic trust, metanorms, and noise in the transmission of messages) we introduce to assess the relationship between individual-level attributes and the outcomes of inter-ethnic encounters. By contrast, Fearon & Laitin limit their analysis to two ideal types of institutional arrangements; in both contexts, individualism is driven purely by opportunism. Table 1 compares the respective analytical frameworks.

In Section 2, we present an empirical puzzle—variation in the scale of ethnic violence—based upon episodes of ethnic conflict in Rwanda and Burundi. Our comparison of cases leads us to discount the conventional explanations offered by Prunier and Lemarchand, highly regarded political historians of Rwanda and Burundi, respectively.³ In Section 3, therefore, we examine the theoretical approach advanced by Fearon & Laitin. While their framework provides a foundation for our analysis, we propose several modifications, in the interest of developing 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. In Section 5, we describe our major findings. Most notably, we discover that the scale of violence varies substantially across episodes even using reasonable default parameter values. We subsequently skew our three key parameters—levels of inter-ethnic trust, metanorms, and noise in the transmission of messages—in order to assess their association with episodes that exhibit extreme violence. These parameters are not novel; however,

³ See Gerard Prunier, *The Rwanda Crisis: History of a Genocide* (New York, NY: Columbia University Press, 1995); and Rene Lemarchand, *Burundi: Ethnic Conflict and Genocide* (Cambridge: Cambridge University Press, 1996).

they have neither been used to explain differences in the scale of ethnic violence, nor used together to construct a model of ethnic conflict. In Section 6, we provide examples from Rwanda and Burundi that illustrate how the tendencies we detect in our results operate in practice.

2 An 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.⁴ The instinct might be to describe this event as simply the latest round in inter-ethnic warfare that has characterized the Great Lakes region since the post-colonial era began in the late 1950s. Upon further examination, however, one detail stands out in stark relief: the massacre, though genocidal in nature, remained geographically confined and limited in scale.⁵ 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 empirical 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 500,000-800,000 Tutsi were killed by civilian militias and the Hutu-dominated army following the assassination of President Juvénal Habyarimana. Less well known are the prior 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. This juxtaposition of events presents a similar enigma: why didn't the violence escalate further in 1963-4, 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 (see next page) offers further details on these episodes, which we culled from the numerous incidents of ethnic conflict in Rwanda and Burundi since 1959.⁶ We use this empirical background to assist our investigation into causes of the variation in the scale of ethnic violence.

⁴ 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.

⁵ By “genocidal”, we mean killings committed with an evident intent of eliminating members of an ethnic, racial or religious group, the standard that is specified in the *United Nations Convention on the Prevention and Punishment of the Crime of Genocide*. Localized ethnic massacres, however, do not conform to formal 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.

⁶ We synthesized the details from Robert Block, “The Tragedy of Rwanda,” *The New York Review of Books* 41 [October 20, 1994]: 3-8; Jean-Pierre Chrétien, “Burundi: The Obsession with Genocide,” *Current History* 95 [May 1996]: 206-210; Mike Edwards, “Central Africa’s Cycle of Violence,” *National Geographic* 191(6) [June 1997]: 124-133; Fergal Keane, *Season of Blood: A Rwanda Journey* (New York, NY: Penguin Books, 1995); Rene Lemarchand, “Burundi: The Politics of Ethnic Amnesia,” Chapter 6 in Helen Fein (ed.), *Genocide Watch* (New Haven, CT: Yale University Press, 1992); Rene Lemarchand, *Burundi: Ethnic Conflict and Genocide* (Cambridge: Cambridge University Press, 1996); Gérard Prunier, *The Rwanda Crisis: History of a Genocide* (New York, NY: Columbia University Press, 1995); Filip Reyntjens, *Burundi: Breaking the Cycle of Violence* (London: Minority Rights Group, 1995); and Frank Smyth, “Blood Money and Geopolitics,” *The Nation* 258 (May 2, 1994): 585-588.

TABLE 2
Select Episodes of Ethnic Violence in Rwanda and Burundi

Event (Classification)	Sequence of Events	Scale and Scope of Violence/Displacement
Episode 1 1963-4 Rwanda (<i>localized massacre</i>)	An attack by Tutsi exiles is stymied by the Rwandan army 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 (<i>genocide</i>)	After a failed coup attempt by Hutu politicians, Hutu unrest erupts in Nyanza Lac and Rumonge (south), Cankuzo (east), and Bujumbura (capitol), 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 (<i>localized massacre</i>)	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 (<i>genocide</i>)	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 Tusti refugees

These countries are natural candidates for exploratory analysis since the similarities between them—their ethnic divisions, demographics, economic contexts, etc.—facilitate comparison. The cases share another central characteristic: the ethnic group in power reacting to a provocative event attributed to its ethnic 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 localized massacres while the other result in a systematic genocide. Our primary objective, therefore, is to account for these differences.⁷

2.1 Conventional Explanations

Prominent accounts of ethnic conflict in Rwanda and Burundi suggest three reasons for variation in scale of ethnic violence. Lemarchand's 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.⁸ Yet the 1972 uprisings, which involved only marginally greater organization than the 1988 unrest, provoked reprisals that were far more extensive and considerably more methodical.⁹ This evidence undermines the premise that extreme reprisals reflect provocations that are particularly threatening to the political establishment.

⁷ This paper 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.

⁸ 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.

⁹ 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), p. 93.

Lemarchand also posits that the 1988 episode remained confined because the Tutsi elite anticipated that more extensive violence would have, unlike in the case of the 1972 genocide, serious international repercussions. He attributes the change in outlook to advances in human rights monitoring and media coverage.¹⁰ The 1994 genocide in Rwanda supplies a counterfactual: despite exposure and condemnation of events by foreign representatives as well as journalists and human rights activists, the international community evidently was not treated as an obstacle.¹¹

Finally, Prunier attributes the massive scale and systematic scope of the 1994 genocide in Rwanda to the elaborate “final solution” developed by high-ranking Hutu extremists.¹² He deduces that this agenda enabled the perpetrators to respond with such speed and efficiency that they overwhelmed any efforts at resistance or escape. Two pieces of evidence, however, cast a shadow on this admittedly intuitive assessment. One is the apparent absence of such a plan prior to the Burundian genocide of 1972.¹³ Thus, advance planning may not be necessary to accomplish systematic, nationwide killings, even if some measure of coordination and synchronization is implicit in the character of the event. In addition, the genocidal agenda that was conceived in Rwanda—far from creating an absolute information asymmetry that gave the perpetrators an insurmountable advantage—had likely been common knowledge for several months.¹⁴ For example, a Tutsi refugee described to BBC journalist Fergal Keane 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.¹⁵

Thus, advance planning is not sufficient: one must account for the Tutsi who discounted the information at their disposal, leaving themselves vulnerable, however unwittingly, to attack.¹⁶

¹⁰ See Lemarchand (1996), pp. 128-130.

¹¹ Prunier (1995) indicates that the genocide occurred partly because of the lack of intervention by the international community; he highlights the reluctance of the U.S. to become engaged in the conflict. Foreign military forces were on the scene, however, and there is no indication that the perpetrators anticipated this sort of muted response.

¹² See Prunier (1995), pp. 237-248.

¹³ A common Hutu myth imputes responsibility for these killings to Burundi's Minister of Foreign Affairs Arthemon Simbananiye, who is alleged to have masterminded efforts to provoke a Hutu uprising in order to justify reprisals designed to achieve ethnic parity between (minority) Tutsi and (majority) Hutu. See Lemarchand (1996), p. 27.

¹⁴ 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 Jamie Metz, “Information intervention: when switching channels isn't enough,” *Foreign Affairs* 76 [November/December 1997], p. 15.

¹⁵ Fergal Keane, *Season of Blood: A Rwanda Journey* (New York, NY: Penguin Books), p. 88. At the very least, radio broadcasts ensured that as the plan was being set into motion, many Rwandan Tutsi 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.

¹⁶ See Keane (1995), p. 10. *Inyenzi* (cockroaches) is a derogatory name among Hutu for the Tutsi rebels.

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

3 A Theoretical Framework

Having discounted these explanations, we turn to the theoretical framework proposed by Fearon & Laitin. Whereas the ethnic conflict literature concentrates on the origins of group identity and the sources of hostility, they offer useful insights concerning the dynamics of episodes of conflict. We use their framework as a foundation for our analysis, which emphasizes factors they examine only tangentially or in a manner that is largely incompatible with our cases.

In a departure from the literature, Fearon & Laitin 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 there are relatively few episodes of violence is the “decentralized, nonstate institutions [that] often arise to mitigate problems of opportunism in interactions between individuals of different ethnic groups.”¹⁷ These social institutions engender two types of non-violent outcomes, as confirmed by their social matching game model of ethnic conflict. The first type results from *spiral equilibria*, under which cooperation between ethnic groups is induced by the 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 fellow group members. Both types of equilibria implicitly moderate the frequency and the scale of episodes of ethnic conflict.

We have several reservations about Fearon & Laitin's 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.¹⁸ They offer two justifications for assuming this 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, ethnic groups. Both presuppositions are unusual given their concern that “[e]xisting 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.”¹⁹ Moreover, Fearon & Laitin explore the consequences of encounters that are not anonymous—i.e., members of one ethnic group can distinguish members of other ethnic groups. The fact that individuals can reliably identify those who cheat or exploit them as ethnic outsiders requires some familiarity and thus frequency of interaction. Yet they still assume inter-ethnic interactions are rare and the population segments homogenous—premises which support their equilibria. Even leaving aside issues of consistency, their assumptions are not compatible with the cases of Rwanda and Burundi, where ethnic heterogeneity is the norm.

Second, Fearon & Laitin acknowledge that self-policing could have brutal consequences: a group might use sanctions to induce members to participate in attacks against rivals, rather than to punish them for doing so. In-group policing equilibria would then be associated with high levels of violence, potentially well in excess of the levels observed under spiral equilibria. They fail, however, to describe the circumstances under which such reversals of function take place.

¹⁷ Fearon & Laitin (1996), p. 715.

¹⁸ Fearon & 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.

¹⁹ Fearon & Laitin (1996), p. 731.

Our third reservation concerns Fearon & Laitin's 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 they allude to—"mistakes, misinterpretations, drunkenness, sudden passions, or unobservable variations in payoffs from encounter to encounter"—are hardly the exception in multi-ethnic societies such as Rwanda and Burundi.²⁰ This observation suggests the conditions required to foster ethnic coexistence may not prevail as universally as they contend.

Finally, we have qualms about Fearon & Laitin's answer to a fundamental question: Why in some cases do inter-ethnic relations remain cooperative for a long time but periodically lapse into spiraling violence, while 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 developing policing). This omission is notable given that the relative prevalence of these two institutional solutions implicitly affects the frequency and scale of ethnic violence.

4 A 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 non-participants, conditional participants, and willing participants within an assembled crowd.²¹ The underlying logic clearly applies to Rwanda and Burundi, in addition to being 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 frequent interactions (and even some intermarriage) between members of the two major ethnic groups. Consequently, there is no reason to expect dramatic information asymmetries, though 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 their ethnic rivals. Here, we follow Kandori (1992), who observes that individuals exhibit multiple behavioral attributes: e.g., reputation can manifest independent of mere group affiliation. Because these attributes inevitably provide a range of cues to others, the interaction between any two individuals—even ethnic rivals—is rarely deterministic.²²

²⁰ Fearon & Laitin (1996), p. 723.

²¹ Susanne Lohmann, "A Signaling Model of Informative and Manipulative Political Action," *American Political Science Review* 87(2) [June 1993]: 319-333. 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 Douglas Dion, "Competition and Ethnic Conflict," *Journal of Conflict Resolution* 41 [October 1997]: 638-648, especially at page 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 inter-group conflict. See Susan Olzak, *The Dynamics of Ethnic Competition and Conflict* (Stanford, CA: Stanford University Press, 1992). For a similar assessment of Burundi, see Prunier (1995), p. 140.

²² Michihiro Kandori, "Social Norms and Community Enforcement," *Review of Economic Studies* 59 [January 1992]: 63-80.

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 & 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. Following Axelrod (1986), we term this a “metanorm”, i.e., an arrangement whereby a first-order norm is sustained under threat of sanctions for non-compliance. He finds that the stronger the metanorm—because enforcement is more likely and/or involves more severe punishment—the greater the likelihood of sustaining the norm.²³

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.²⁴ Section 4.1 offers a narrative of the sequence of events, while Section 4.2 provides a formal description of the behavior of the agents. We utilize this particular methodology, as opposed to the game-theoretic approach employed by Fearon & Laitin, in order to introduce variation in individual characteristics such as the propensity to engage in violence, the expectation that ethnic outsiders 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 in these characteristics over time, whereas in Fearon & Laitin’s model the individual agents retain the same attributes throughout the course of play. Finally, in our model decisions reflect the uncertainty of current engagement, rather than being based upon full knowledge of the outcomes of prior cross-group pairings as in Fearon & Laitin’s social matching game.

4.1 Sequence of Events

The model consists of individuals from two ethnic groups, *A* and *B*. 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 transmitted 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.²⁵ Individuals then react to the message by engaging in or abstaining from violence.²⁶

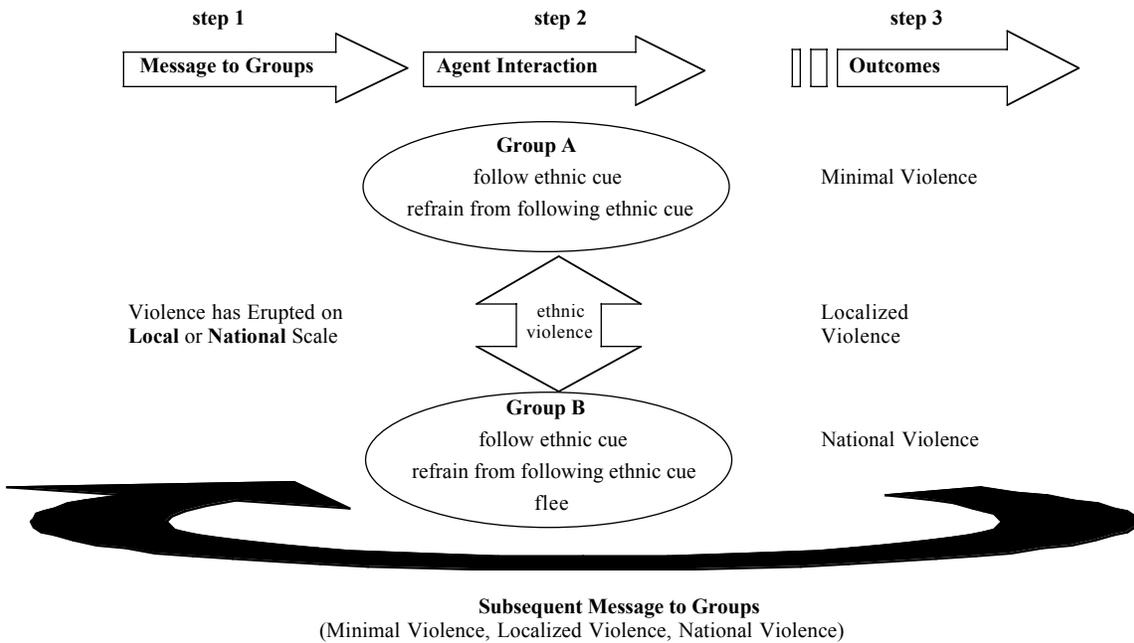
²³ Axelrod’s analysis focuses on the behavior of witnesses to crimes, but we envision that his findings are similarly applicable to a setting where the first-order norm entails eliminating ethnic rivals. See Robert Axelrod, “An Evolutionary Approach to Norms,” *American Political Science Review* 80:4 [December 1986]: 1095-1111. For another theorization of self-policing potentially engendering high-violence equilibria, see also James Coleman, *Foundations of Social Theory* (Cambridge, MA: Harvard University Press, 1989).

²⁴ Information about the computational model, which was programmed using Matlab v5.0 for Windows software, is available directly from the authors.

²⁵ Our computational model does not explicitly incorporate a spatial dimension. Instead, we utilize the proportion of agents that are casualties of ethnic violence to assess whether the violence is local or national in scope. In essence, the absolute scale of violence serves also as a proxy for the territorial scope of the violence.

²⁶ To keep things simple, we base the mobilization to engage in violence entirely on individual agents’ independent responses to the public messages that are transmitted. No collective action is involved at this stage.

FIGURE 1
Sequence of Events in an Episode



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 applicable metanorms. According to the metanorm, 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.²⁷ Their actions are similarly based on the content of the message and their level of extremism, as well as on their beliefs of how the members of group *A* are likely to behave in response to the message.

Episodes—each comprised 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.²⁸ The

²⁷ Flight does not guarantee survival, though (as we detail later) it reduces the probability of an individual being killed by members of the ethnic group in power. Flight also implicitly reduces the extent of killing by the rival ethnic group. In subsequent work, we address the issue of flight from ethnic and political violence in greater detail.

²⁸ 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.

progression of an episode is dictated by the subsequent messages that are transmitted to new segments of the population, which in turn depend upon 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.

4.2 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 upon 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 $x_{ai} \sim U[0,1]$, and the strength of any metanorm. The metanorm 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 metanorms 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 metanorm 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 metanorm 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 metanorm—an increase in the probability of an individual following the ethnic cue—is inversely proportional to the individual's level of extremism. A strong metanorm 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 metanorm: a_i follows the ethnic cue with a probability of $[x_{ai} + (1 - x_{ai})m]$ and refrains with a probability of $1 - [x_{ai} + (1 - x_{ai})m]$. If a_i plays E , he then kills a member of group B with a probability k_a .²⁹ 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 , while each decision to play R increases k_a by δ .³⁰ 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 $(1 -$

²⁹ 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 percent chance of killing a member of group B .

³⁰ 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.

x_{bi}).³¹ 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_t —then b_i increases (decreases) y_i by a constant λ .³²

TABLE 3
Actions by Members of the Rival Ethnic Group

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

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 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 that is sent to agents at time $t + 1$ is “national violence.” Otherwise, the subsequent message is “local violence.”³³

5 Results

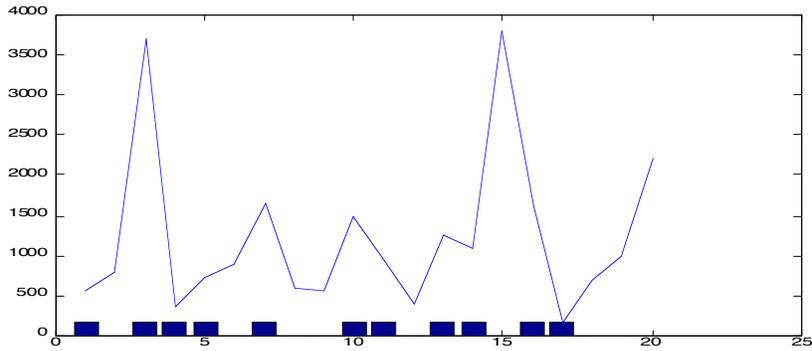
We seek to evaluate the effect of four parameters—the type of initial message that is transmitted, the rival ethnic group’s level of trust in members of the ethnic group in power, the strength of the metanorms that are enforced by the ethnic group in power, and the level of noise in the transmission of local messages to members of the ethnic group in power—on the scale of ethnic violence. To provide a basis of analysis and comparison, we construct a “base model” that employs default values for each of the key parameters, i.e., values that should not skew the outcomes. In particular, the rival ethnic group is, on balance, divided equally between trusting and distrusting individuals, metanorms are drawn from a uniform random distribution, there is no noise in the transmission of messages, and the two groups are the same size. Using these default values, the average level of violence for the base model—generated from 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. Given the way we designed the model, there is no indication the outcomes of these encounters are structurally deterministic.

³¹ We set k_b to be 0.5, i.e., there is a 50 percent 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.

³² Unlike k_a , beliefs are updated across events, our intuition being that there is learning across the independent samples of group B that 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 percent higher expectation that members of A will play E .

³³ In order 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 percent of participants killed) and n equal to 0.27 (i.e., 27 percent of participants killed).

FIGURE 2
The Effect of the Initial Message on the Scale of Ethnic Violence

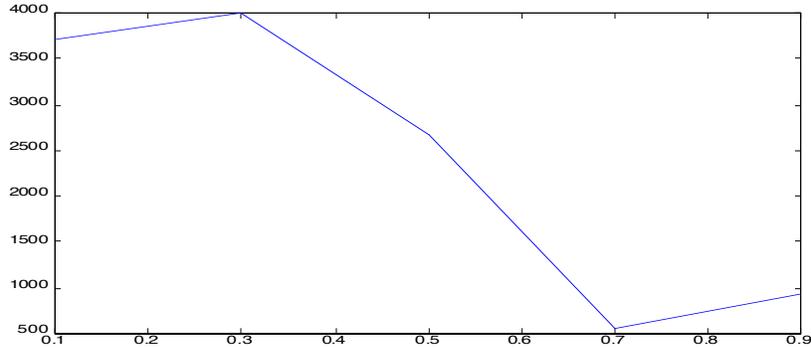


Note: Each episode, comprised of a maximum of 20 events, used the following parameter values: $\beta=0.5$, $m \sim U[0,1]$, Noise=0, $A=500$, and $B=500$. The episodes with bars were initiated by a local message, whereas the episodes without bars were initiated by a national message.

Figure 2, which displays the results from a sample run of 20 episodes, indicates that the variance cannot be attributed to differences in 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 message is neither sufficient nor necessary for an episode to achieve a high level of violence. Similarly, an initial local message is neither sufficient nor necessary for an episode to result in only low-level 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 key parameters 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 that the members of the rival ethnic group have in the members of the ethnic group in power. As we discussed earlier, we assume that individuals have the capacity to develop independent beliefs about how their nominal ethnic rivals are likely to behave during episodes of conflict. When surveyed across an entire population, these beliefs effectively translate to a measure of inter-ethnic 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 the members of group A will follow the ethnic cue and 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 being killed than is otherwise the case, our expectation is that greater trust will be associated with higher levels of violence.

FIGURE 3
The Effect of Trust on the Scale of Ethnic Violence



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

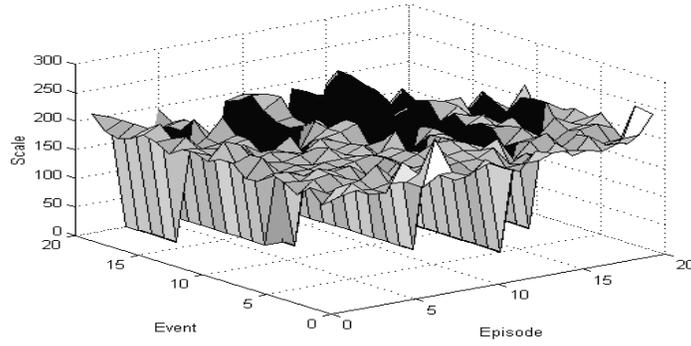
Our results, depicted in Figure 3, contradict our original intuition: on balance, low levels of trust actually induce substantially greater violence than we observed under the base model, whereas high levels of trust result in substantially less violence. The primary explanation for this relationship is 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 beginning on the next page) representative patterns of violence for each of the levels of inter-ethnic 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.

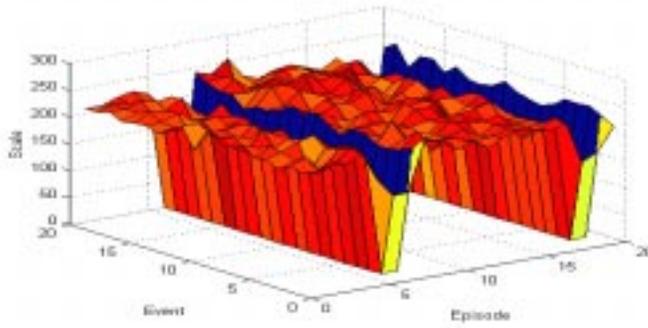
With moderate or high levels of distrust, at least half of the episodes continue for the full term of 20 events. The intensity of violence is relatively normal, but aggregate levels of violence are high due to the long duration of episodes. In contrast, with moderate to high levels of trust, most episodes 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 to engage in violence whenever the message is 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.

FIGURE 4
The Effect of Trust on the Patterns of Ethnic Violence

Equal Trust and Distrust ($\beta=0.5$)



Moderate Distrust ($\beta=0.3$)



High Distrust ($\beta=0.1$)

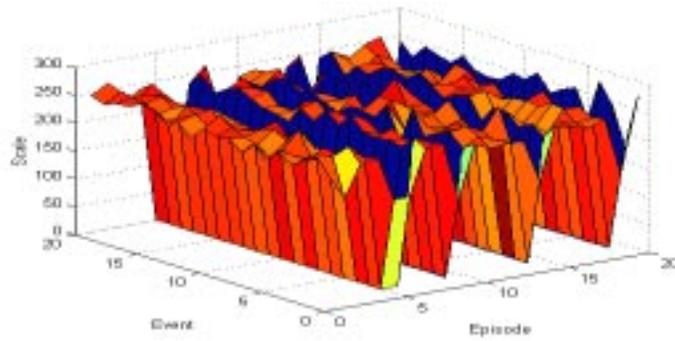
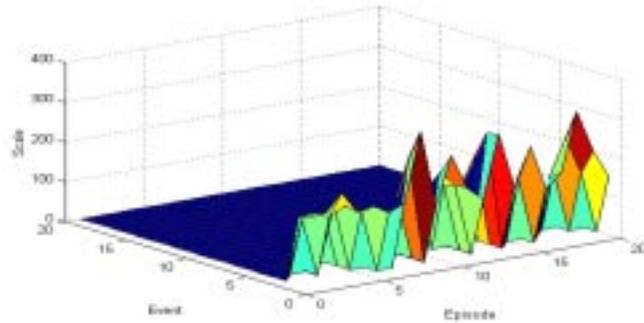
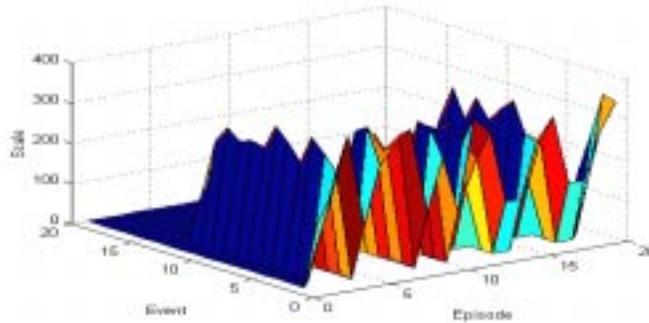


FIGURE 4 (cont.)

Moderate Trust ($\beta=0.7$)



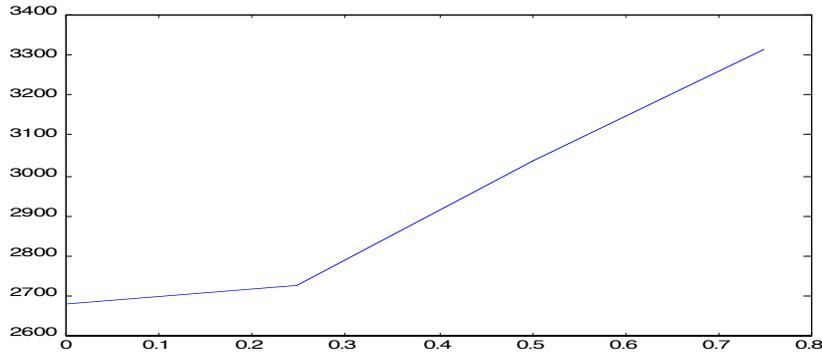
High Trust ($\beta=0.9$)



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

The same basic logic also accounts for the non-linear 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 where 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, multi-ethnic 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.

FIGURE 5
The Effect of Metanorms on the Scale of Ethnic Violence



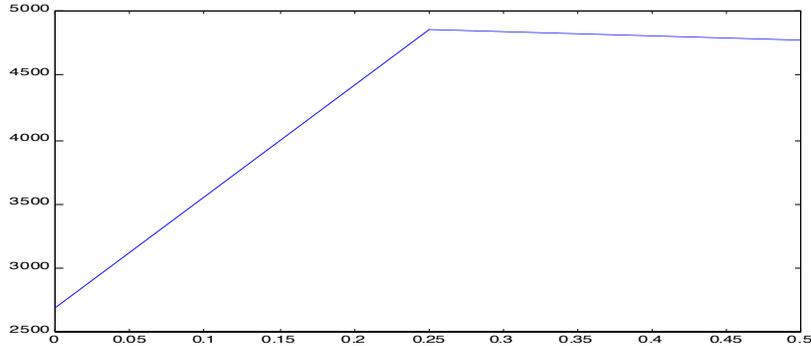
Note: These summary results are based on 200 episodes, each comprised of a maximum of 20 events, using the following parameters: $\beta=0.5$, Noise=0, $A=500$ and $B=500$.

The second parameter we consider is the strength of metanorms. In our base model, we allow metanorms 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 metanorms, we simply narrow the range over which the parameter m can vary.³⁴ As we explained in earlier, this corresponds to an increase in the certainty and/or severity of sanctions for non-participation.

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 metanorms 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 metanorms to the range $[0.5, 1]$: the mean number of deaths per episode is 14 percent higher than under our base model. Further limiting the range to $[0.75, 1]$ results in a mean value that is 24 percent 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.

³⁴ As noted above, this adjustment ensures that the metanorm effect—i.e., the $(1 - x_i) m$ component of the formula that ultimately determines when a_i plays E —always equals or exceeds some non-zero lower bound.

FIGURE 6
The Effect of Noise on the Scale of Ethnic Violence



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

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.³⁵ 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 to members of group A .³⁶ 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 the metanorms that ordinarily apply, more apt to attack a member of B .

We initially introduce a 25 percent 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 percent 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 percent probability remains essentially the same as when it is only 25 percent. 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 as versus 500). With a moderate amount of noise and moderately strong metanorms, the mean scale of violence—based on 200 episodes—is 2972 deaths per episode, or 11 percent higher.³⁷ 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.

³⁵ We do not explore the scenario of national messages being transmitted with noise. We assume the leaders of the ethnic group in power make a special effort under those circumstances to fully inform their members.

³⁶ In other work, we evaluate the effects of noise on the interpretation of messages by members of both groups.

³⁷ Each episode, comprised 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 $B=500$.

6 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 & Laitin 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 these mechanisms. Absent this link, their institutional solutions fail to account for the disparity in violence across episodes.

Using a computational model, we are able to generate variance in the scale of violence across episodes. At the initial stage, we deliberately set our parameters so as not to skew the outcomes: the rival ethnic group is evenly distributed between trusting and distrusting members; levels of extremism and metanorms are drawn from uniform random distributions; the initial messages are randomly determined and (like all subsequent messages) transmitted without noise; and the two ethnic groups are the same size. The fact that the model generates variance even with these defaults provides evidence that the parameters in our model affect the scale of violence.

By focusing on the variation in individual extremism and beliefs, as well as in the impact of metanorms on individual behavior, we demonstrate that structural factors—institutions, planning, etc.—need not be instrumental in dictating the outcome of episodes of ethnic conflict. Moreover, our analysis indicates that prior levels of violence are inadequate predictors of future levels of violence. Instead, individual characteristics appear to be sufficient to generate variation in the scale of violence. Consequently, the conventional preference for tracking structural factors appears somewhat misguided; our results indicate that comparable attention should be afforded to individual-level dynamics. To evaluate specific hypotheses concerning the effects of individual characteristics, we adjust several parameters of interest—trust, metanorms, and noise—that derive from our reservations concerning Fearon & Laitin’s theoretical framework. 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 higher average levels of violence than do high levels of trust. These patterns initially struck us as counter-intuitive. On closer examination, we determine that differences in aggregate levels of trust influence the patterns of violence that emerge. In communities characterized by high levels of inter-ethnic trust, violence is more likely to peak and subside rapidly, whereas communities characterized by low levels of inter-ethnic trust are more likely to experience moderate levels of violence that are sustained over long periods of time. The results vindicate our hypothesis that variation in the scale of violence is affected by individuals acting upon independent beliefs about the trustworthiness of nominal ethnic rivals, rather than being equally distrusting of all outsiders, as Fearon & Laitin effectively assume.

A related result is the non-linear 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. The finding can be viewed as consistent with our original hypothesis that higher levels of trust are associated with increasing levels of violence because they render the rival ethnic group more susceptible to unexpected attacks by members of the ethnic group in power. This phenomenon is afforded some support by empirical evidence from 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 where Tutsi approached local Hutu officials to seek assistance and relied on whatever support or advice they provide, 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.³⁸

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 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.³⁹

Second, we find that genocidal metanorms—i.e., 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 & Laitin’s in-group policing equilibria can indeed intensify, rather than simply mitigate, the scale of ethnic violence. Our finding is 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.”⁴⁰ In fact, one of the notable aspects of the 1994 genocide was the killings of some 10,000 to 30,000 Hutu moderates who refused to take part in the violence.⁴¹ The Rwandan genocide shares this feature in common with the 1972 genocide in Burundi. Lemarchand (1996) reports that “[t]he 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.”⁴²

Metanorms, 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

³⁸ Keane (1995), p. 90.

³⁹ Human rights organizations reported that more Tutsi were killed in Rwanda’s churches than in any other type of site. See, for example, African Rights, *Rwanda: Death, Despair and Defiance* (London: 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 with one another. See Howard Chua-Eoan, “Of Death and Defiance: Hate kills Burundi’s religious fighter for peace,” *Time* 148 (September 23, 1996): 49. The former relationship is confirmed by Prunier (1995), p. 251.

⁴⁰ Keane (1995), p. 88. 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.

⁴¹ Prunier (1995), p. 265.

⁴² Lemarchand (1996), pp. 97-98.

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 the most pernicious variety of metanorms 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 & Laitin’s 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 were soon to kill were *ibyitso*, 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.⁴³

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 moderate metanorms, when combined with moderate noise, can offset the effects of the minority status of the ethnic group in power. This result clarifies how genocidal violence was feasible in Burundi, where the ruling ethnic group (Tutsi) constitutes only 15 percent of the population. As we described earlier, members of this group were subject to severe sanctions during the 1972 violence—moderates were killed for failing to join or otherwise resisting the violence. Lemarchand (1992) also describes how Tutsi leaders stoked fears of Hutu domination by associating demands for fuller participation in political affairs with an effort by the educated Hutu elite to gain political ascendancy. The indication is that these anti-Hutu activities polarized ethnic relations, to the point where the reprisals for the opportunistic Hutu uprising in 1972 were exceptionally brutal and systematic.⁴⁴ Cultivating such antagonism at the national level, even if no formal program of genocidal violence is afoot, fosters more extensive and severe reactions to ethnic conflict that would ordinarily be confined to localized clashes.

⁴³ Prunier (1995), pp. 137-138.

⁴⁴ 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), pp. 76-78.

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 [December]: 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 [May]: 206-210.
- Chua-Eoan, Howard. 1996. "Of Death and Defiance: Hate kills Burundi's religious fighter for peace," *Time* 148 [September 23]: 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 [October]: 638-648.
- Edwards, Mike. 1997. "Central Africa's Cycle of Violence," *National Geographic* 191(6) [June]: 124-133.
- Fearon, James and David Laitin. 1996. "Explaining Interethnic Cooperation," *American Political Science Review* 90(4) [December]: 715-735.
- Kandori, Michihiro. 1992. "Social Norms and Community Enforcement," *Review of Economic Studies* 59 [January]: 63-80.
- Keane, Fergal. 1995. *Season of Blood: A Rwanda Journey*. New York, NY: Penguin Books.
- Lemarchand, René. 1992. "Burundi: The Politics of Ethnic Amnesia," Chapter. 6 in Helen Fein (ed.) *Genocide Watch*. New Haven, CT: Yale University Press.
- _____. 1996. *Burundi: Ethnic Conflict and Genocide*. Cambridge: Cambridge University Press.
- Lohmann, Susanne. 1993. "A Signaling Model of Informative and Manipulative Political Action," *American Political Science Review* 87(2) [June]: 319-333.
- 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, NY: 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 [May 2]: 585-588.