1-1-2019

Mass Atrocities and their Prevention

Charles H. Anderton  
*College of the Holy Cross, canderto@holycross.edu*

Jurgen Brauer  
*Chulalongkorn University, brauer.jurgen@gmail.com*

Follow this and additional works at: [https://crossworks.holycross.edu/econ_working_papers](https://crossworks.holycross.edu/econ_working_papers)

Part of the Behavioral Economics Commons, Economic Theory Commons, Human Rights Law Commons, International Economics Commons, International Relations Commons, Labor Economics Commons, Law and Economics Commons, Law and Gender Commons, Political Economy Commons, and the Public Economics Commons

Recommended Citation  

This Working Paper is brought to you for free and open access by the Economics Department at CrossWorks. It has been accepted for inclusion in Economics Department Working Papers by an authorized administrator of CrossWorks.
Mass Atrocities and their Prevention

By

Charles H. Anderton and Jurgen Brauer

January 2019

COLLEGE OF THE HOLY CROSS, DEPARTMENT OF ECONOMICS
FACULTY RESEARCH SERIES, PAPER NO. 19-01*

Department of Economics and Accounting
College of the Holy Cross
Box 45A
Worcester, Massachusetts 01610
(508) 793-3362 (phone)
(508) 793-3708 (fax)

https://www.holycross.edu/academics/programs/economics-and-accounting

*All papers in the Holy Cross Working Paper Series should be considered draft versions subject to future revision. Comments and suggestions are welcome.
Mass Atrocities and their Prevention†

Charles H. Anderton* and Jurgen Brauer**

* Professor of Economics and Distinguished Professor of Ethics and Society, College of the Holy Cross, Worcester, MA, USA

** Emeritus Professor of Economics, Augusta University, Augusta, GA, USA and Visiting Professor of Economics, Chulalongkorn University, Bangkok, Thailand

January 15, 2019

Abstract: Counting conservatively, and ignoring physical injuries and mental trauma, data show about 100 million mass atrocity-related deaths since 1900. Occurring in war- and in peacetime, and of enormous scale, severity, and brutality, they are geographically widespread, occur with surprising frequency, and can be long-lasting in their adverse effects on economic and human development, wellbeing, and wealth. As such, they are a major economic concern. This article synthesizes very diverse and widely dispersed theoretical and empirical literatures, addressing two gaps: a “mass atrocities gap” in the economics literature and an “economics gap” in mass atrocities scholarship. Our goals are, first, for noneconomists to learn how economic inquiry contributes to understanding the causes and conduct of mass atrocities and possibly to their mitigation and prevention and, second, to survey and synthesize for economists a broad sweep of literatures to serve as a common platform on which to base further work in this field.

JEL codes: B55, D71–D74, D91, F55, H56, J15, K38, P16

† For assistance with research and references, we thank Roxane Anderton and Brittany Coscio. For comments on drafts of this paper, we thank scholars drawn from economics, law, sociology, and political science, namely Daniel Arce, Chris Blattman, Hollie Brehm, Raul Caruso, Jacqueline DeMeritt, Mark Drumbl, Mario Ferrero, Lisa Hultman, Sebastian Ille, Matthew O. Jackson, Matthew Krain, Zoë Marriage, J. David Richardson, Ron P. Smith, Erik Snowberg, and Anastasia Voronkova. The usual disclaimer applies.
1. Mass Atrocities: Why Care?

Mass atrocities are episodes of large-scale violence committed against unarmed populations. In war- and peacetime, they occur with perplexing frequency and geographic reach; data suggest an average onset of nearly two mass atrocities per year since 1900, occurring on all continents except Antarctica. They include the infliction of injuries and deaths on civilians in the intuitively obvious meaning, and also forms of gendered and age-selective violence, forced removals of victims from their homelands, and the psychological trauma resulting from any such acts both for the surviving direct victims and for indirectly affected people such as family members. Mass atrocities—of which genocides are but the most prominent example—can be events of enormous scale. For example, the sum of all deaths from all domestic and transnational terror events recorded for the nearly 50 years since 1970 is about 380,000 (University of Maryland, 2018). This is far below the number killed in the 1994 Rwandan genocide alone (500,000 to 800,000 deaths). Mass atrocity data show, in all, a conservative estimate of about 100 million deaths since 1900, three times as many deaths as from all natural disasters (see Section 2). Kiernan (2007), Pinker (2011, pp. 320–43), and White (2012) provide overviews, and evocative descriptions, of varieties of mass atrocities across history and geography. ¹

Mass atrocities can generate inordinately long-lasting adverse effects on economic, political, and cultural development. For example, Soudis, Inklaar, and Maseland (2016), in a cross-country model of nation-states suffering an episode of genocide in the post-Holocaust era, find that the cumulative effect on per-worker GDP shows a rapid decline of about 10 percent, which then persists even ten years post-genocide onset. For a country-specific case, Hodler (2019) finds that Rwanda’s GDP declined 58 percent in 1994, its year of genocide, and then took 17 years to

¹ To be clear, we mean human populations. We do not review literature regarding nonhuman species.
return to its estimated counterfactual level. Although there are cases in which forced population homogenization—which is what mass atrocities often amount to—appear to improve the economic fortunes of an atrocity-imposing group (for example, Indonesia post-1966; see Murshed and Tadjoeddin, 2016), empirical case studies also show that economic and social development can remain attenuated for half a century and longer (Acemoglu, Hassan, and Robinson, 2011; Urbatch, 2017). On occasion, development ceases altogether for the atrocity-suffering group, as in the case of the Native American population in California in the 1800s (Madley, 2016). Even before inventing the word genocider, Lemkin (1933) noted that “acts of extermination [include] actions undertaken to ruin the economic existence of the members of a collectivity” (our emphasis). The causal mechanisms at work point, especially, to demographic disruptions (Kugler, 2016) and to changes in cultural norms, beliefs, and values in surviving populations that adversely affect social fabric, political participation, and economic activity (Nunn and Wantchekon, 2011; Grosfeld, Rodnyansky, and Zhuravskaya, 2013).

Still, in addition to obvious humanitarian and economic development concerns, why should mass atrocities interest economists (and why should economics interest mass atrocity scholars)? We have two answers: Economics can help elucidate the mechanisms by which mass atrocities are carried out and it can contribute to an improved understanding of why mass atrocities occur (and recur) in the first place. Starting with the standard core of economic theory, this can be done by examining the rationality of the means and the rationality of the ends—the ‘how’ and the ‘why’ of mass atrocity (Ferrero, 2017a)—and then be enriched with insights drawn from psychology, sociology, and political analysis. For instance, paradigms drawn from behavioral and identity economics, based on theories of preference formation (Akerlof and Kranton, 2000), motivated beliefs (Gino, Norton, and Weber, 2016), and of beliefs as assets (Bénabou and Tirole,
2011), can help explain both atrocity leaders’ and followers’ behaviors. Social network economics can show how mass atrocities may emerge spontaneously and unpredictably and how they diffuse over a network of atrocity perpetrators. And concepts in politics such as power, ideology, and the instrumental use of hate, in addition to transaction costs, the difficulty of benefit-capture, and free-riding in collective decisionmaking contexts can help explain why mass atrocities, once started, are so hard to stop.

The economic theories surveyed in this article help to more fully address questions related to the rationality and nonrationality of mass atrocities, compliance by large (often oppositional) segments of populations in mass atrocities, and how an entire social structure of mass atrocity architects, bureaucrats, on-the-ground perpetrators, suppliers of the means of murder, silent bystanders, conformers, enthusiasts, opportunists, resisters, third parties—and victims—coheres, operates, and evolves. In the process, we also raise questions regarding overlooked events (many mass atrocity events remain uncaptured in databases), overlooked actors (victims’ choices and circumstances are strangely underresearched), and overlooked consequences (for example, interventions that backfire and make things worse).

The article proceeds as follows. Section 2 characterizes the study of mass atrocities in terms of definition, scholarly context, actors at play, and data. Focusing more on the causes and conduct of mass atrocities than on their consequences, Section 3, the heart of the article, then explores four entwined strands of economic theory as they pertain to mass atrocity. Empirical evidence on mass atrocity risk, prediction, and intervention is summarized in Section 4. Section 5 takes an illustrative look at atrocity prevention through the lens of the economics of international law. Section 6 concludes with a short reflection on what has been done, has not been done, and can be done. Our main contributions are, first, to synthesize very diverse and
dispersed work across streams of economic theory and, second, to comprehensively survey and assess the empirical literature.

2. Characterizing Mass Atrocities

In this section, we briefly discuss aspects of the definition of mass atrocity (Section 2.1), scholarly history and context (2.2), mass atrocity actors (2.3), and relevant data (2.4).

2.1 Definition

The phrase mass atrocity is an umbrella term, covering a variety of concepts and offenses. Much of the genocide studies literature has settled, for now, on the term atrocity crime, avoiding the qualifier “mass” but adding “crime” (Straus, 2016). Article I of the Statutes of Rome, establishing the International Criminal Court (ICC), uses the phrase “most serious crimes of international concern” and, Article 5 (and amendments) defines four such crimes: genocide, crimes against humanity, war crimes, and the crime of aggression. Genocide, for instance, “means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such: (a) Killing members of the group; (b) Causing serious bodily or mental harm to members of the group; (c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part; (d) Imposing measures intended to prevent births within the group; (e) Forcibly transferring children of the group to another group.” In genocide, people are singled out for who they are (or deemed to be) rather than for what they (may) have done.

Crimes against humanity are callous acts committed in a widespread or systematic manner against civilians as individuals rather than as members of an enumerated group; war crimes

---

2 The Rome Statute text defining genocide is identical to that of the UN Genocide Convention (United Nations, 1951).
constitute serious violations of the laws and customs applicable to armed conflict between or within states. In contrast, the crime of aggression refers to “the use of armed force by a State against the sovereignty, territorial integrity or political independence of another State,” and thus need not involve atrocities. Genocides and crimes against humanity can occur in wartime and peacetime, whereas war crimes can occur in war only. Politicides, the massive killing of political opponents, are not a type of genocide under international law but scholars and others often treat them equivalently. They may, or may not, constitute a crime against humanity or a war crime. Rummel (1997) popularized the term democide (“the murder of any person or people by a government, including genocide, politicide, and mass murder”). But democide is not defined in law. Another term used in the literature is ethnic cleansing, which usually means the forced expulsion of a people-group from a territory of habitual settlement. It, too, is not a crime recognized in law.

Despite the broad coverage these concepts and terms appear to provide, it is unclear whether, for instance, the killing of several tens of thousands of Viet Cong suspects during the second Indochina war, or the state-sanctioned massive killing of suspected narcotics dealers in the Philippines over the past few years, or even the killings of tens of thousands of people in Mexico’s narcotics-related violence since the mid-2000s are “most serious crimes of international concern.” Even if not, these and other cases, such as politically motivated or otherwise recklessly induced famines for purposes of political control (for example, Ukraine, China, Ethiopia), are atrocities on a massive scale.\(^3\) Mass atrocity databases thus include at least some such cases. Importantly, perpetrators of atrocities need not be states, but nor it is clear that perpetrators then must be nonstate actors. For instance, past or current practices of treating females as male property (Europe), footbinding (China), sex-selective abortions (India), and

---

\(^3\) Hryn (2008) characterizes the 1930s Ukrainian famine in a Soviet context as “hunger by design.”
female genital cutting (parts of Africa, Asia, and the Middle East) are examples of widespread, massive gendered violence even if at times locally condoned, rising to the status of social norms, and unrecognized as crimes in law. These practices, neither state nor nonstate but perpetrated by “society,” are instances of massive violence, with attendant economic damage.

At its broadest, the subject matter, then, is the class of atrocities that are “massive,” carried out by whomever and for whatever underlying political, economic, or cultural reason. It includes massive crimes not deemed atrocities and mass atrocities not deemed crimes. For science, such a broad definition makes the operational setting of conceptual and quantitative boundaries difficult. But for economists at least, any violation of people’s physical, mental, and cultural wellbeing on a massive scale is relevant because it sabotages the development of human capital needed for the betterment of society.

The legal texts, domestic or international, do not help much as they restrict attention to a limited class of atrocities and eschew quantification. In contrast, scientists create codeable definitions and metrics but often restrict them to killings, which is incomplete as atrocities cannot be restricted to killing alone: The UN Genocide Convention, for instance, lists killing as merely one of five enumerated acts, each of which can constitute genocide. For example, “(d) Imposing measures intended to prevent births within the group; [and] (e) Forcibly transferring children of the group to another group” arguably constituted acts of genocide in Australia’s treatment of its indigenous populations well into the twentieth century (Tatz, 2013).

Just as the conceptual boundaries of “atrocity” are blurred in scholarship and law, so is the numeric definition of “massive,” if only because it is unclear which time scale to use. The literature frequently employs a cutoff point of 1,000 or more noncombatant civilians killed over

---

4 Relatedly, the International Criminal Tribunals for the former Yugoslavia (ICTY) and Rwanda (ICTR) “have stated clearly that no numerical threshold had to be crossed for atrocities to count as genocide” as long as certain aspects of the UN Convention’s definition of genocide are met (Meierhenrich, 2014, p. 29).
some period of time, implying that if 800 are killed within that time, and another 800 in the ensuing time period, an atrocity is not counted as such. Indeed, one of the dangers in scholarship and policymaking is to equate mass atrocities with genocidal mass killings, leading to a “canon of general acceptance” which includes Armenia, the Holocaust, Cambodia, Rwanda, the former Yugoslavia, and Sudan’s Darfur region “but virtually no other cases” (Bloxham and Moses, 2010, p. 4). For this reason we refer in this article to many additional atrocity cases worldwide, pre- and post-Holocaust.

In a classic text, *Statistics of Deadly Quarrels*, Lewis Richardson (1960) starts his death counts with $n=1$. This removes aggregation as a definitional problem. Indeed, time- and geocoded event-based datasets such as the Armed Conflict Location and Event Dataset (ACLED) go into a scientifically desirable direction, limiting aggregations of counts, geographic space, and time, leaving aggregation to researchers’ specific interests. Consequently, analysts need not agree on a single dependent variable, nor search for a single “best” model of mass atrocities. The set of appropriate theoretical and empirical approaches and models therefore is wide-ranging and can be eclectic (Sections 3 and 4). That said, we must reckon with the literature’s current conventions of counts, space, and time, more so since most of the relevant databases are built to reflect these conventions. Thus, much of the literature covered deals with cases of 1,000 or more deaths of unarmed civilians. But this should not blind readers to the much broader scale, scope, and severity of mass atrocities.

2.2 A Brief History of Mass Atrocity and Genocide Studies

In the two decades following the Holocaust, historians and social scientists paid little scholarly attention to Nazi-Germany’s efforts to exterminate Jews and other people-groups. Nor were other mass atrocities—such as mass killings of Herero and Nama people in then South-West
Africa in the early 1900s, slaughter of Armenians in the late Ottoman empire, destruction of indigenous peoples in Australia and the Americas, weaponization of hunger in the man-made mass starvation of Ukrainians in the 1930s, and cases of mass elimination of civilians during wars and empire building throughout ancient and medieval history (Pinker, 2011)—subject of much social scientific inquiry in the immediate years following the Holocaust. A breakthrough occurred with the publication of Raul Hilberg’s (1961) study, *The Destruction of the European Jews*. Considered seminal, Hilberg is credited by genocide scholars (Meierhenrich, 2014, pp. 5–6) in helping to create an academic field of study focused upon the origins and consequences of the Holocaust, later extended to the academic study of other mass atrocities. (The field is often named “genocide studies” in reflection of its early focus on genocides.) By the 1970s and 1980s, “first-generation genocide scholarship” (Straus, 2007a, p. 480) used concepts and tools from the social sciences and rigorous historical analyses, led by sociologists (for example, Dadrian, Fein, Horowitz, Kuper, Porter), social psychologists (Charny, Staub), political scientists (Harff, Krain, Melson, Rummel), and historians (Adalian, Browning, Chalk, Hovannisian). These pioneers’ work was significant for at least three reasons. First, they emphasized the intentional aspects of perpetration by mass atrocity architects and “troops” as well as the various incentives that nonperpetrators had to be acquiescent bystanders. Many of these scholars emphasized that mass atrocities were argued (or perceived) by perpetrators as instrumentally rational in achieving certain goals for political leaders, including destroying an emerging threat, positioning oneself more strongly during war, eliminating a persistent rival, acquiring land or other resources, and/or establishing reputation. Importantly, as early as 1979, sociologist Helen Fein emphasized that to account for genocide, “we must consider how it may be motivated or appear as a rational choice to the perpetrator” (Fein, 1979, p. 7). Second, the first generation of empirical and historical case
studies established that architects and on-the-ground perpetrators of mass atrocities are, by and large, “ordinary people,” scoring, for example, within normal ranges on standard personality tests, and that their behaviors, outside of atrocity perpetration, in such contexts as family life and community relations were normal and even commendable (Waller, 2007, chs. 5–8). Moreover, depending on the case at hand, atrocity perpetrators could number in the many thousands. This, third, led these early scholars to wrestle with the relative importance of individual agency and of psychological, social, and political contexts in understanding the recruitment into and behavior of perpetrators, as well as actions and inactions by other actors in mass atrocity circumstances.

From the early 1990s, a second generation of scholars has put the field on increasingly stronger theoretical and empirical foundations, also aiming to integrate genocide studies back into their home disciplines (see Straus, 2007a; Meierhenrich, 2014, pp. 9–18 for synopses). We highlight five themes that permeate this generation of scholarship. First, the rational and strategic aspects of mass atrocity perpetration became well-established in the field and across the social science disciplines (Valentino, 2004). Second, to arrive at richer models of the choices actors make in atrocity contexts, scholars argue that perspectives on the rationality of mass atrocity should be integrated with insights that emphasize nonrational aspects, drawn from psychology, sociology, and political science (Staub, 1989; Melson, 1992; Fein, 1993). Third, it is increasingly recognized that emphasizing context or individual agency in the social scientific analysis of mass atrocity must give way to research that includes both: “Structure and agency are inextricably intertwined,” write Bloxham and Moses (2010, p. 6). Later on, we will argue that modern economics is particularly well-suited and poised to help effect such multidisciplinary integration. Fourth, the latest scholarship is incorporating dynamic and feedback effects missing from most earlier studies, including social evolutionary processes that can lead to a tipping point into mass
atrocity, contagion effects in which mass atrocity acceptance among a population spreads like a “disease,” and mass atrocity as an emergent phenomenon in which seemingly trivial changes in a personal, social, or political condition can have dramatic effects on the risk or spread of mass atrocity. Finally, fifth, we note the striking gap in work by economists, relative to other scholars, on understanding mass atrocity as an economic proposition and on the challenges of mass atrocity risk and prevention. This gap is beginning to be addressed, as our article will show.

2.3 The Actors: Perpetrators, Victims, Bystanders

Three broad classes of actors in mass atrocities are perpetrators, victims, and bystanders (Hilberg, 1992; Ehrenreich and Cole, 2005). Among the set of perpetrators are architects (who envision, plan, and have the power, authority, and control necessary to drive atrocities), bureaucrats (who implement, supervise, and manage the machinery of destruction), and troops (who inflict the actual harm). All three share responsibility for the atrocities they jointly commit. Victims are classified by perpetrators as “others” (an out-group) based upon some identifier the victims effectively are unable to offset (Ehrenreich and Cole, 2005, pp. 218–9), even if victims attempt to do so by such means as cultural assimilation or religious conversion. Victims have been further distinguished by their degree of vulnerability leading into and during mass atrocity, their age and gender, and the nature of their victimhood (for example, death, injury, loss of family members, or loss of homeland or wealth) (Hilberg, 1992; Ehrenreich and Cole, 2005; Donà, 2018). Bystanders are socially and spatially situated between perpetrators and victims (Staub, 1989, p. 20; Donà, 2018, p. 2), sometimes outside the immediately affected society, for example in diasporas. Bystander actions can range from actively helping victims, to resisting perpetrators, to neither helping nor resisting, to refusing to actively participate in the destruction but agreeing with the ideology to do so, to disagreeing with the destruction ideology but being
willing to benefit from it as in opportunistic looting (Ehrenreich and Cole, 2005; Donà, 2018). Hilberg (1992) collapses these behaviors into three subcategories: helpers, onlookers, and gainers.

Actor identities are not fixed (Fujii, 2009). Behaviors can be fluid and multiplex as, for example, when a perpetrator who was killing Tutsi simultaneously acted to rescue a specific Tutsi who was a former soccer teammate (Donà, 2018, p. 12; see also Campbell, 2010) or of victims who are coerced to become victimizers (Drumbl, 2016). Individuals can also dramatically change their behavior over time. A well-known example is Oskar Schindler who joined the Nazi party and enriched himself from slave labor (Staub, 1993, p. 335) but later risked his wealth and life to help Jews.

Mass atrocity perpetrators, victims, and bystanders are usually conceived of in uniquely personalized ways. Yet they also encompass organizations such as parties, agencies, bureaus, firms, and international governmental organizations. For example, when the United Nations observes an on-going atrocity but does little to help, it can be classified as an onlooker—raising the issue of collective responsibility for harm done (Drumbl, 2007; Bellamy and Dunne, 2016).

2.4 Data and Scope of the Problem
Stitching together various databases, scholars know, for the years 1900–2017, of close to 200 mass atrocities in which states deliberately killed at least 1,000 of their own or other states’ noncombatant civilians per case; of these, data sources provide fatality estimates for 149 cases. Furthermore, data sources permit identification of mass killings perpetrated by nonstate actors from 1989–2017, again with 1,000 or more victims, of which at least 39 have occurred. Culled

5 Choosing the year 1900 as our cutoff for state-perpetrated mass atrocities (and 1989, for nonstate atrocities) is dictated, in part, by the limits of available databases. Our point is to numerically illustrate the sheer scale of atrocities and, at present, this only seems possible for the late modern and contemporary eras. It is desirable to extend databases even into the pre-historic era (Keeley, 1996). Certainly, scholarship in ancient and modern history
from various data sources (see Appendix), Figure 1 plots each case at the onset year. Time is displayed on the horizontal axis. The vertical axis shows victim counts, but on a log scale. We refer to this as the Mass Atrocity Log Scale (MALOS). The circles represent state-perpetrated mass atrocities; the triangles constitute nonstate cases. Also included are 14 “X” cases representing mass killings perpetrated by unidentified actors, 1995–2017. Although they cannot be tied to a unified decisionmaking process or party responsible, the emergence and increasing frequency of atrocities committed by “who knows who” are a palpable empirical phenomenon (Anderton, 2018).

The estimated cumulative death toll depicted in Figure 1 is about 100 million people. This understatesthe true toll because the number refers to cases documented in readily available databases only and, even among documented cases, only to those for which fatality estimates are available. Even with these restrictions (including the 1,000+ restriction), the estimated fatalities fall within the range of the number of people who died due to bubonic plague—the Black Death—in the mid-1300s in Eurasia and Europe (about 75 to 200 million people) and far exceed total deaths worldwide from natural disasters (~33 million), 1900–2016 (EM-DAT, 2018). Not included in our counts are nonfatal physical injuries and mental trauma. Thus, if one were willing to assume—and scattered research suggests as much—that for every civilian killed,
another two are physically injured and a further two traumatized in other ways (for instance, witnessing and/or fleeing atrocity), then the combined casualty toll from the cases reported in Figure 1 alone reaches about half a billion people.

If the sheer scale of mass atrocities is enormous, so is their severity and brutality. For example, estimated fatalities totaled over just three genocides (Cambodia 1975–9, Pakistan 1971, and Sudan 1983–2002; total: ~6,837,000 people) far surpass the total estimated battle-related fatalities for all of the 347 intrastate wars fought from 1900–2015 (total: ~4,893,230 people; see Appendix). Indeed, six of the mass atrocities plotted in Figure 1 each have a MALOS score greater than the combined score for the aggregate of all battle-related fatalities for the 347 intrastate wars. The brutality of mass atrocities lies not only in who is being destroyed, and in what numbers, but in how people are destroyed. Atrocity survivors often report physical and mental torture prior to “mere” killing (for example, child dismemberment in front of parents or insertion of vermin through tubes into people’s intestines).

The data shown in Figure 1 raise questions about a dominant theme in the literature, namely, that mass atrocity is primarily “death by government” (Rummel, 1994; Pinker, 2011). In terms of events, Figure 1 documents 36 mass atrocities perpetrated by states, 1989–2017. Yet during the same time span, an additional 39 mass atrocities were perpetrated by nonstate actors (many acting independently), and in particular by rebel and militia groups (see Appendix). Regarding fatalities from 1989–2017, however, government mass atrocities have been more deadly (~ 1.9 million fatalities) than nonstate cases (~ 250,000). Regarding data on “low-level” violence

---

displaced people (IDPs) and refugees can number in the many millions of people per conflict, for example, in Colombia and Syria. For a review of studies on sexual violence and associated mental trauma and social consequences, 1981–2014, see Ba and Bhopal (2017).

9 The six are: China (onset 1925, MALOS=7.0), USSR (onset 1930, MALOS=6.81), USSR (onset 1932, MALOS=6.78), Germany (onset 1933, MALOS=7.0), Japan (onset 1935, MALOS=6.71), and China (onset 1949, MALOS=7.55). Note that the log scale obscures the difference between the smallest and the largest values (6.71; 7.55); the small log-difference of 0.84 translates into an extra 30 million deaths!
against civilians (VAC) in Africa (not shown in Figure 1), ACLED reports 47,835 VAC events, excluding riots, protests, and peacekeeping missions, and 178,397 associated fatalities from 1997 to December 15, 2018 (Raleigh, Linke, Hegre, and Karlsen, 2010). Among perpetrator categories, state forces (military and/or police) carried out 20.2 percent of VAC attacks and were responsible for 16.5 percent of the fatalities. The very large remainder is accounted for by other actors. (Rebel forces accounted for 19.6 percent of attacks, but 37.0 percent of fatalities, political and identity militias for 29.7 percent of attacks and 30.1 percent of fatalities, and unidentified perpetrators for 30.5 percent and 15.8 percent, respectively). For “low-level” violence, then, “death by government” may not be as prominent a feature.

A final point: At times the literature conflates mass atrocities with periods of war. This is not necessarily correct. For example, of the 138 mass atrocity cases for which onset years for both, wars and mass atrocities are available, 20 percent had no overlap years with a war period at all. Of the remaining 80 percent, one-third of the cases began in a year before war onset and then continued into the war period, another third had the same onset year as a war so that it is unclear whether mass atrocity or war came first in that year, and only 30 percent occurred after the onset year of a war (see Appendix).

In sum, mass atrocities encompass far are more than genocides and involve more than “mere” killing. States need not be the primary atrocity perpetrators. Even when limited to the 1,000+ people killed cases depicted in Figure 1, cases are frequent, fairly evenly distributed across time, occur in wartime and in peacetime, and lead to an enormous number of victims.

---

10 ACLED provides data on VAC events in which hundreds or even thousands of civilians are killed, but also events with zero fatalities. Zero fatality events should be important in atrocity research because they would include forced abductions, non-fatal rapes and beatings, destruction of cultural goods, and stealing of land and other goods.

11 According to the ACLED codebook (www.acleddata.com), political militias can operate in conjunction with a “recognized government, governor, military leader, rebel organization, political party, business elite, or opposition group”; identity militias “include … tribal, communal, ethnic, local, community, religious, and livelihood militias.”
3. Actors and Their Choices: Theory

This section discusses four sets of economic theories that contribute to the study of mass atrocities. These are standard theory’s core of rational choice and game-theoretic approaches as well as some theory of the firm considerations (Section 3.1), behavioral and identity economics in regard to nonstandard decisionmaking, beliefs, and preferences (3.2), social network economics and analysis of complex adaptive systems (3.3), and aspects of political economy (3.4). These tributaries intertwine; works cited in one section might well have been placed in another. We survey relevant literature, highlight contributions of particular interest, and conclude with an evaluation of the extant material (Section 3.5) before turning to empirical matters in Section 4.

3.1 Tributary 1: Standard theory

3.1.1 Rational choice theory

Lionel Robbins argued that “any kind of human behavior falls within the scope of economic generalisations” (1937, p. 16). This includes war—of which Robbins made a special point—peace, and, in our case, mass atrocities or their absence. This is so because in Robbins’ choice-theoretic view even a dictator bent on group removal has to choose how best to deploy scarce means to pursue desired ends (Robbins, 1937; Fein, 1979; Valentino, 2004). Standard theory does not interrogate preferences (Stigler and Becker, 1977) but studies how the fulfillment of given, subjective preferences is objectively constrained by the resources needed to defray the expense of obtaining the ends. If, given the resources, killing costs “too much” and thereby limits acquisition of other preferred (nonkilling) goods, then killing may not take place. Similarly if, given the prices of killing and nonkilling, resources are too few then, again, killing may not be chosen. Thus, behavior is subject to the constraint of scarce means. Even if the dictator already
has decided on some amount of killing, Robbins’ view comes into play. If, given available resources, mass killing by shooting is “too costly” relative to killing in gas chambers, then killing in gas chambers it is. Either way, in standard theory the “problem” of mass atrocity becomes one of a general class of mathematical problems called constrained optimization. So long as preferences are given and consistent (complete, transitive, context independent), the dictator is said to act rationally—not in regard to the content of his preferences, of course, but in the way he goes about seeing them fulfilled (Wintrobe, 2007).

Ignoring for now the why of mass atrocity, this way of studying them can yield insights into their decisionmaking mechanics, offer predictions and testable hypotheses, and suggest what it may take to mitigate, or altogether prevent, them. For example, Anderton and Brauer (2016) use Lancaster’s (1966) input attributes model to analyze the many input possibilities available to a group pursuing mass atrocity. Their work reveals that protection efforts to constrain an elite’s use of one or several of the destruction inputs may have relatively little effect, and sometimes no effect, in protecting the targeted group. Under such circumstances, protection policy cannot focus on just one or several inputs available to mass atrocity architects. To be effective, prevention requires holistic policies, raising the unit costs of virtually all inputs (or lowering their productivity), reducing resources available for mass destruction, or altering the preference set of the ruling elite altogether so that mass atrocity is not utility-enhancing—a result readily available in the counterterrorism literature as well (Landes, 1978; Enders and Sandler, 2011). Lancaster’s model can also be used to analyze the role of any one key input for achieving a particular form of mass repression such as ideological control of substantial numbers of people. For example, Stalin’s mass resettlement of millions of people was designed, in part, to keep

---

12 See Semelin (2007, pp. 233-6) for documentation of wide-ranging “mass killing technologies,” including those deployed during colonial conquests in Africa in the nineteenth century and in the Ottoman Empire, Germany, Poland, Bosnia, and Rwanda in the twentieth century.
regime opposition weak, but this was keenly dependent on transportation networks. Geocoding some 2.65 million arrest records overlaid on railroad maps, Zhukov (2016) provides empirical evidence that Stalin obeyed the law of demand during the regime’s population control campaigns over the period 1945–59. Specifically, all else equal, because of the higher cost, Stalin was less willing to coerce more remotely located civilians to move. The further away they lived from a railway station and the greater the distance to a forced labor camp (Gulag), the higher the likelihood that they received other sentences such as incarceration or property confiscation.

As this example shows, the constrained optimization model can be applied to mass atrocity as a means to an end. Assume a production function in which mass killing of civilians is one of the inputs for achieving political or territorial control and the ruling elite is resource-constrained. Such models generally show that mass killing acts will occur when they are “cheap” and productive at the margin relative to other means of achieving control such as directly contesting rebels (Anderton, 2014a; Anderton and Brauer, 2016). Such models imply a downward-sloping demand function for mass atrocity in which the higher the per-unit cost of people-group destruction, the lower the amount of mass atrocity actions demanded. Again, Zhukov (2016) provides empirical evidence for such behavior. Specifically, using 80,000 geocoded events, he finds that the number of acts of violence against civilians by governments and rebels in Africa over the period 1997–2010 are greater, the lower are the logistical costs of such acts (due to high road density and distance from a country’s capital city), all else equal. Furthermore, Anderton and Ryan (2016) use a Cobb-Douglas production function, Stone-Geary utility function, and concepts from the rational addiction literature to derive a demand function for atrocity in which lower-level attacks in period $t$ lead to high-level attacks in period $t + 1$ owing to a habituation

(addiction) mechanism.

While constrained optimization models have been applied to mass atrocity architects, virtually no such models have been applied to other actors such as “troop” perpetrators, bystanders, or victims (Brauer and Anderton, 2014). Brauer and Caruso (2016) provide guidance for how this work can unfold with applications to five pre-Holocaust genocides. Such work may help to explain behavior that currently seems incomprehensible. For example, extremely brutal means of destruction (rape, torture, forced killing of family members) can be instrumentally rational (von Joeden-Forgey, 2016), demonization and dehumanization making killing psychologically less costly or “burdensome” to the perpetrators (Pinker, 2011). In some cases, the threat of brutal killing can be financially rewarding when, for example, some victims of the 1994 Rwandan genocide paid their killers to shoot them instead of hacking them to death with machetes (des Forges, 1999, p. 216). The approach may also help explain seemingly puzzling actions of agents in mass atrocity contexts, for example, the aforementioned outcome in which an agent acts as both perpetrator and rescuer (the “utility” of killing a perceived foe as opposed to the “disutility” of killing a friend). Recognizing the role of price and resource constraints, economists are particularly well-equipped to interrogate the ‘how’ of mass atrocities and to elucidate their mechanics, possibly leading to fresh insights into intervention, mitigation, or prevention.

3.1.2 Classic game theory

By taking account of strategic interdependence among actors, classic (as distinct from evolutionary) game theory models of mass atrocity are an extension of rational choice and decision theory. Literature often points to strategic aspects of mass killings.\textsuperscript{14} Perpetrators could

intentionally mass kill civilians for a variety of reasons including to eliminate a threat to the elite’s ideological or physical control of the state, to undermine civilian support for a rebel group contesting the elite, to create terror and thus subservience among the population, and to eliminate a rival group once and for all. Of course, nonstate groups may also mass kill civilians for similar reasons (Wood, 2010; Eck, 2015; Ottmann, 2015; Vargas, 2016). Most game theory models of mass atrocity posit a ruling authority contesting a potential or actual rebel group in which civilians can be manipulated (for example, repressed, forcibly relocated, or killed) by one or both sides and in which each side has an objective function that includes political, territorial, or resource control.15

An important model along these lines is Esteban, Morelli, and Rohner (2015, hereafter EMR) who consider two potentially warring groups with distinct initial populations, one of which is in power. In each period of an infinite horizon game, a divisible surplus is available, generated from a per-period exploitable natural resource and output produced by labor drawn from the two populations according to a labor productivity parameter. At the beginning of period $t$, production occurs, and the group in power announces the division of surplus based on a fairness parameter, which may be bound from below based on domestic and international norms. The two groups then decide whether to fight, with both needing to choose peace for peace to prevail. If war occurs, a portion of the surplus is lost. The winner retains (or obtains) power, keeps the remaining surplus, and decides whether and how many to mass kill from the other group, a factor which may be bound from above based on international norms.

In the model, the fundamental tradeoff that determines the decision to mass kill is between the gains and costs of group elimination for the winner following war. Gains include the
substantial control of discounted future streams of the surplus and a weakening of the other group, which could prove beneficial in the event of future violence. Costs involve the loss in discounted future streams of production in labor-intensive sectors owing to elimination of part of the population. The game permits multiple subgame perfect equilibria (SPE) and, even though equilibria without violence exist, EMR show that the surplus-maximizing SPE can display war and mass killing.\textsuperscript{16} Hence, one cannot necessarily escape the mass killing result by coordinating onto a surplus-maximizing equilibrium. This is a disturbing result.

Comparative static predictions of the EMR model are important. Tweak any parameter in the model, and the bargaining space can shrink, making the commitment problem more urgent. Thus, war and mass killing are most likely to occur when natural resources are relatively more important than people in producing surplus. And since people consume surplus, cutting down on their numbers can result in the tragic economic logic that lay behind Nazi-era Germany’s schemes to work to death and destroy “useless eaters” (Browning, 2004, p. 295).\textsuperscript{17} Also noteworthy are the “backfire” results that emerge from the EMR model. In contrast to an unrestricted upper bound, new norms that tighten the constraints on an actor’s power to mass kill can increase killing incentives as a smaller future surviving population of “others” may be desirable from the perpetrators’ perspective. Similarly, if domestic or international actors insist on more stringent fairness norms in surplus division by the group in power, incentives to mass

\textsuperscript{16} There is a commitment problem in the model. Each player, when in power, could lose power in the future and face elimination. Hence, any player in power today (say A) will, under certain parametric conditions, want to eliminate player B to prevent the future possibility that player A is eliminated. Even in the version of the model where there are limits to the exercise of power, a commitment problem still exists. Specifically, any player in power today (say A) will, under certain parametric conditions, want to weaken player B to diminish the potential that player A is weakened in the future. For applications of the bargaining theory of war to a variety of commitment problems in mass atrocity contexts, see Anderton (2010, pp. 465-73). In the civil war literature, exploration of the commitment problems is much more developed (Blattman and Miguel, 2010).

\textsuperscript{17} German expansion to the east through Poland and into Russia from September 1939 to March 1942 eventually led to conquests of areas with little or no industry. “Since fewer workers were needed for economic purposes, the remaining population appeared as ‘useless eaters’” (Browning 2004, p. 295).
kill can increase as the group in power will not gain a large enough share of the surplus to make peace and the avoidance of mass killing worthwhile.

In contrast to literature that focuses on strategic wrangling between a ruling and a potential or actual rebel group, Vargas (2016) considers interdependencies among nonpassive victims and two groups of combatants who contest control of territory. The probability that a group secures territorial control is a function of the size of its civilian support network. The probability that a civilian is killed by one group given her support of the other is determined by a Tullock-like contest success function encompassing the fighting efforts and effectiveness parameters of the armed groups (Tullock, 1975, 1980). Armed groups simultaneously choose their fighting effort to maximize their civilian support network. This, in turn, maximizes their probability of winning the contested territory. Civilians, meanwhile, strategically choose which group (or none) to support. In supporting one group, heterogeneous individuals incorporate in their utility functions the probability of being killed by the other group and also an idiosyncratic parameter reflecting each civilian’s bias toward supporting a group. A perfect-information sequential game is solved by backward induction, beginning in stage 2 with each armed group choosing its fighting effort to maximize its payoff. The resulting best-response functions lead to equilibrium fighting efforts. Backing up to the first stage, each civilian takes the equilibrium fighting efforts of the armed groups into account to determine her or his probability of being killed by one group when supporting the other. Given the distribution of bias parameters across civilians, the equilibrium number of civilians compliant with each armed group is determined.

We emphasize three of the model’s results. First, noncombatants are strategically killed by both sides in the contest, a result frequently observed in reality. Second, owing to a preference for survival, it can be extremely difficult for civilians to resist supporting armed groups or to
remain neutral when combatants are willing to conduct atrocities. Moreover, as their utility calculus changes, civilians can change the side they support, as again is frequently observed. Survival trumps loyalty. Third, the model applies to civilians unable to flee the violence (called “stayers” in Ibáñez and Moya, 2016), a generally underresearched topic in the field.

Other than DeMeritt (2015a, 2015b), we are unaware of formal models of principal-agent games between governments and subcontracted militia groups or between external actors and rebel groups in which civilians are at risk of mass atrocity (but see Eck, 2015; Salehyan, Siroky, and Wood, 2014, for empirical treatments). Furthermore, there are few formal models of strategic interactions among perpetrators, bystanders, and victims (but see evolutionary models in Section 3.4) or of strategic aspects of refugee outcomes (for exceptions, see Zeager and Bascom, 1996; Zeager, 1998, 2002, 2005). In addition, scholars have raised the possibility of moral hazard associated with bystander interventions. For example, a rebel group could strategically goad a government into mass killing so as to induce bystander intervention (Kuperman and Crawford, 2006; Kuperman, 2008; Kydd and Straus, 2013). We take up intervention in the empirical section below. Developments in the theory of dictatorship also point to important implications for mass atrocities (Wintrobe, 1990, 1998, 2007; Verwimp, 2003a), again addressed later on.

3.1.3 Theory of the firm

Mass destruction requires the acquisition of inputs from and delivery of destruction to potentially many locales. But the nature of atrocity production varies across cases, for example, in the Nazi-era Holocaust from high capital intensity, strong complicity of numerous firms, and geographically widespread tentacles of destruction, on the one hand, to the high labor intensity, relatively little direct business complicity, and more localized deaths in the Rwandan genocide,
on the other hand. Regardless of the variance across cases, principles of managerial economics, industrial organization, and economics of information (propaganda) apply. A key contribution is Breton and Wintrobe (1986), who explain how Nazi bureaucracy (encompassing leaders, agencies, bureaus, and firms) became such a monumental industry of mass killing. The static model of bureaucracy assumes that bureaus have given assignments in which they are monopolists over their assigned areas but Breton and Wintrobe (1986) insist that the Holocaust “cannot be understood by using such a model of bureaucracy” (p. 909). Instead, superiors and subordinates traded with each other through “informal services” and “informal payments” (p. 909), and bureaus competed with one another for jobs, resources, and the generation of new ideas, policies, and projects to please Hitler and other Nazi leaders. The key insight is that Hitler fostered a tournament-like “Schumpeterian competition or entrepreneurship” among bureaus toward people-group destruction (Breton and Wintrobe, 1986, p. 909), a rent-seeking contest. Within this dynamic world of new ideas and programs, “schemes [to destroy Jews] were constantly being put forward by rival power centers or rival entrepreneurs, and Hitler would choose among them” (Breton and Wintrobe, 1986, p. 912).

Breton and Wintrobe (1986) provide important insights. We emphasize two. The first concerns the notion of efficiency. Some scholars cite the frightening efficiency of the Holocaust; others emphasize its inefficiencies. Thus, Stokes and Gabriel (2010, p. 465–6) quote Grey (2005) on the idea that the Nazi’s industrial organization and bureaucratic systems of rules made genocide “as technically efficient as genocide could be” but also quote Bauer (2001) for whom

---


the Nazi bureaucracy was “often a fumbling, ineffective, contradiction-ridden machine.” Breton and Wintrobe (1986) make a different argument: If Hitler and other key leaders created a messy cauldron of bureaus and businesses with overlapping interests and claims, the unleashed dynamism would do far more to mass-destroy people than rearranging a static flowchart of organizational responsibilities could ever do.

Second, Breton and Wintrobe’s work speaks to shortcomings in the literature on the dark side of organizations, including economic theories of organizations. The literature recognizes that organizations can make mistakes, harbor misconduct, encourage opportunistic behavior among its divisions, and appeals to cognitive gaps in leadership and to groups, ideology, and peer pressure to explain how organizations can end up on a slippery slope toward disaster, but such going awry is generally treated as unintended. What is missing is the intentional creation of misconduct and disaster as one finds in Breton and Wintrobe (1986). This is no small distinction. In the “dark side” literature, information erroneously becomes incomplete; in the Nazi bureaucracy, incomplete information is purposely fostered because it aids in the entrepreneurship of destruction. As Breton and Wintrobe (1986, p. 914) note, “another indicator of competition among bureaus in the Nazi state is the imprecision of the orders emanating from the top. When orders do not have specific content and when they are not directed at anyone in particular, they will elicit a large response from diverse quarters” (our emphasis).

Moreover, the dark side literature is not nearly dark enough. In the economics of people-group destruction, the harsh economic rationalizations are almost unfathomable. For example, in accounting documents that survived the Holocaust, prisoners who were worked to death generated a profit stream and then a “scrap value” after they died encompassing their clothing,

---

valuables (including extracted fillings), hair (for soap), and ashes (for fertilizer) (Lippman and Wilson, 2007, p. 288). As the war progressed, resource shortages affected Nazi techniques of destruction and cost-benefit analysis was used to optimize the killing routines. For example, analysis showed that people could be killed with less gas if they were made to die longer (15 to 20 minutes) rather than shorter (7 minutes). Hence, victims were made to suffer longer, because it lowered the cost of going about the dark business (Lippman and Wilson, 2007, p. 290). Similarly, children could be gassed and then burned after they were dead, but it was cheaper to burn them alive, and so they were (Lippman and Wilson, 2007, p. 289).

3.2 Tributary 2: Behavioral and identity economics

Beyond the ‘how’ of mass atrocity lies the question of ‘why’. Behavioral and identity economics—confluences of psychology, sociology, and economics—can address aspects of both. Differentiated from standard theory in three ways—nonstandard decisionmaking, nonstandard beliefs, and nonstandard preferences (Akerlof and Kranton, 2000; Rabin, 2002; DellaVigna, 2009)—we illustrate each in turn.

3.2.1 Nonstandard decisionmaking

Decisionmaking in standard theory does not always match decisionmaking observed in practice (Allais, 1953; Simon, 1955). Consider, for example, decision framing and emotion in the context of mass atrocity intervention. Slovic, et al. (2016, p. 613–4) write that the “repetitiveness of [mass] atrocities, which too often have been ignored by powerful people and nations and by the general public, calls for an explanation that may reflect some fundamental deficiency in our humanity” (our emphasis). One of these concerns framing. Emotional framing by “othering” an out-group is a precondition for mass atrocity to occur (Stanton, 2013).

21 Many mass atrocities are

---

21 The issue may go deeper than framing. Nastiness, for instance, may be a feature of the human mind (Abbink and Sadrieh, 2009; Pinker, 2011, ch. 8).
well-reported, even in their earliest stages, yet decisionmakers often hesitate to intervene decisively. Part of this, for instance at the level of the UN Security Council, is due to political wrangling, but another part is due to a decision frame that elevates such wrangling above humanitarian instincts. Even when these instincts remain intact, decisionmakers are subject to psychophysical numbing and compassion fade: Pouring over data, tables, charts, and analyses—a statistical frame—dulls the mind and it can become difficult to process the numbers affectively. Thus, Slovic, *et al.* (2016) recommend the use of affective tools to reframe decisionmakers’ mindsets so as to “employ intuitive thinking to activate and support deliberative processes” (p. 630), for example affective imagery or meeting atrocity victims in person. Statistics of abuse remain important, to be sure, but “storytelling” can help lay the groundwork for their absorption.

To select a second example, contrary to the assumption in standard theory that humans pay full attention to all relevant information and base decisionmaking on its unbiased processing, humans in fact are less than fully attentive and they struggle to process information. Moreover, attention is subject to salience and the number of competing stimuli faced. To cope, people employ decision heuristics (Simon, 1955). Further, due to the presence of prior beliefs, people are not always rationally neutral in regard to information received and this can bias even their inattention and result in suboptimal decision heuristics (DellaVigna, 2009). Slovic, *et al.* (2016) once more provide an example. While psychophysical numbing and compassion fade can help explain why decisionmakers fail to intervene with sufficient urgency in some mass atrocities, in others they may be quite aware of a situation and yet still fail to act adequately. What may help explain this is that decisionmakers rarely have a single issue to decide or, alternatively, that even single issues are comprised of multiple decisionmaking dimensions such as diplomatic, military, political, economic, and cultural. Unlike standard theory, which assumes that choices made
correspond to a decisionmaker’s expressed values, the *prominence effect* states that choices are made on a number of weighted criteria, and the more important or salient of these (the “prominent” ones) most influence the final decision made. One reason is that “unlike expressed values, chosen actions need to be justified, and decisions congruent with prominent attributes are inherently more defensible” (p. 622). Thus, actual decisionmaking—often the lack of intervention in ongoing mass atrocities—can systematically differ from what would be expected on the basis of standard theory but can well be illuminated with an approach grounded in behavioral economics.

3.2.2 *Nonstandard beliefs*

Atrocity victims often believe a situation to be not all that dangerous, arguing that still there is time for change, that others will come to one’s assistance, or that opportunities to escape will exist (see the numerous eyewitness accounts recorded in Totten and Parsons, 2018). Striking examples of this are provided by the families of Raphael Lemkin and Hersh Lauterpacht, respectively the originators of the legal concepts of genocide and of crimes against humanity (Sands, 2016). For example, Lemkin, a Polish Jew, fled in September 1939 owing in part to his belief that Nazi-Germany was going to unleash ever greater atrocities against Jewish people. Pleas made to many of his relatives that they, too, must flee were met with expressions like: “There is nothing new in the suffering of the Jews, especially in time of war … A Jew must wait and pray” (quoted in Power, 2002, p. 24). Tragically, 49 of Lemkin’s relatives perished during the Holocaust (Bartrop and Leonard, 2014, pp. 1301–2). Standard theory—with its assumption that individuals have unbiased, that is, on average correct, beliefs about the state of the world—cannot fully explain victims’ hesitant behavior. In contrast, nonstandard beliefs reflect empirical phenomena such as victim overconfidence (for example in ability, information, and opportunity)
and bias in projecting one’s current state, preferences, and beliefs into the future.

A relatively new research area, highly relevant to mass atrocities, is motivated reasoning and beliefs. Ariely (2013) for instance summarizes empirical literature showing that individuals act self-interestedly when given the opportunity to appear to act altruistically, or else that they delegate decisionmaking to others such that their own altruistic moral beliefs are upheld even as the delegate makes decisions in the self-interest of the delegator. Individuals place value on feeling moral rather than necessarily on being moral. Gino, Norton, and Weber (2016) quote an engineer whose work at J.A. Topf & Söhne included designing more efficient incineration facilities (cremation ovens and ventilation systems) for use in Nazi-era extermination camps: “I was a German engineer and key member of the Topf works, and I saw it as my duty to apply my specialist knowledge in this way to help Germany win the war ...” (2016, p. 190). The expressed preference (“I saw it as my duty”) is bent toward biased beliefs about one’s own good moral standing rather than about a “preference for morality itself” (p. 191). Whereas in standard theory individuals are “classical Bayesians [who] will seek out the most informative evidence and process it in an unbiased way, motivated Bayesians will also be influenced by the evidence that they encounter but will be biased both in choosing which information to acquire and in their interpretation of such information in order to facilitate beliefs in their own morality” (p. 191). Motived to “protect the psychological self” (Staub, 1989, p. 39), individuals form self-serving judgments of what is, or is not, moral, fair, or just and they alter judgments about situations’ objective qualities to make self-interested behavior appear more moral than otherwise.

Bénabou and Tirole (2011, 2016) develop a model that stipulates beliefs as assets. In particular, they posit that people invest in their own identity, hoping that the value of this investment will grow over time and yield appropriate returns (an accumulation of, and dividends
from status, reputation, and life satisfaction. Such investment is worth protecting in the face of challenges. Bénabou and Tirole describe their work as “based on a general model of identity management” (2011, p. 806), the mathematical details of which show escalating commitment to beliefs even to the point where the marginal returns no longer justify continual investment (treadmill effects), the point at which standard theory suggests to shift additional investments into alternative assets. Challenges to one’s identity must be dealt with, eventually by rejecting competing identities. Belief coherence becomes important, as do self-signaling, self-validation, self-reputation, and endowment effects. Individuals become their own reference point. Escalating commitment and treadmill effects can lead to the persistence and stubbornness often found in fanaticism and go some way toward explaining the why, or at least the continuance, of mass atrocities by perpetrators. In contrast to prior literature, Bénabou and Tirole (2011) endogenize the process of identity management. Beliefs face tradeoffs between accuracy and desirability, tradeoffs that can lead to “non-Bayesian behaviors such as not wanting to know, wishful thinking, and reality denial” even as “motivated beliefs will respond to the costs, benefits, and stakes involved in maintaining different self-views and world-views” (Bénabou and Tirole, 2016, p. 142; emphases suppressed).

Epley and Gilovich (2016) turn to the mechanics of motivated reasoning in the service of belief formation and protection. In a nice analogy, they write that “most people do not reason like impartial judges, but instead recruit evidence like attorneys, looking for evidence that supports a desired belief while trying to steer clear of evidence that refutes it” (p. 136). Biased information processing (here, confirmation bias) is associated with asymmetries: When potentially favorable, belief-supporting information comes along, people tend to ask: “Can I believe this?” When potentially unfavorable, belief-challenging information comes along, people
tend to ask: “Must I believe this?” (p. 136). The first question asks for no more than one fact to support the belief; the second asks for no more than one nonsupporting fact to dismiss the challenge. Thus, evidence needed to successfully challenge a belief is greater than that needed to maintain a belief and this can make it hard to dislodge people’s minds from an entrenched, habitual way of thinking and acting. For mass atrocities, these sorts of motivated belief mechanics are at work for all the players involved, perpetrators, bystanders, and victims (Staub, 1985, 1993; Hilberg, 1992; Power, 2002; Waller, 2007).

Individuals also take care to manage what they believe other people believe about them. Belief management is driven by a preference for belief consonance and can partly explain actions such as attempting to alter the beliefs of others or else belief modification to conform one’s own beliefs to those of others. Golman, *et al.*’s (2016) reading of the empirical evidence suggests that a preference for belief consonance can help explain the “deadliest varieties” of interpersonal and intergroup conflict (p. 165) and, especially, “the curious fact that many of the most vicious disputes occur between individuals or groups who share a broad set of beliefs (consider Shiites and Sunnis or Catholics and Protestants) and revolve around differences in beliefs that can seem minor from the perspectives of outsiders to the conflict” (p. 166).

Nonstandard beliefs are relevant to the topic of mass atrocities, then. Especially important is the idea that individuals “will respond” when tradeoffs related to beliefs are altered or when there are changes in the political, economic, or cultural environment within which tradeoff choices are made. Such alterations in tradeoffs related to beliefs are reported in virtually all case studies of mass atrocity. Equally important is the idea that nonstandard beliefs can become a social phenomenon (Bénabou and Tirole, 2016, p. 142), an idea we explore in Section 3.3.

3.2.3 Nonstandard preferences

---

One can think of nonstandard preferences in two ways. One introduces cognitive biases and other psychological phenomena into utility functions. This is the behavioral economics approach. The other brings in social context as well. This is the identity economics approach.

Start with the behavioral approach and consider two examples from the mass atrocity case study literature. The first is the sense of hyper-urgency that atrocity architects and perpetrators often display in carrying out acts of violence. Dunn, for example, reports that in the “very first days” of Indonesia’s invasion of East Timor in December 1975, “an orgy of indiscriminate killing, rape, and torture” as well as “large-scale public executions” took place (2013, p. 286). Similar reports of sudden (within a few days) and substantial outbreaks of mass atrocity have been reported for Rwanda in 1994 (Lemarchand, 2009, p. 492), Pakistan’s attack against Bangladesh in 1971 (Jahan, 2009, p. 302), and Nazi Germany’s invasion of Poland in 1939 (Browning, 2004, pp. 12–35). How to explain this urgency? Whereas standard expected utility theory assumes time-consistent discounting of the value of future utility to the present, laboratory and field data show that humans engage in time-inconsistent, hyperbolic discounting: As if viewed through a faulty telescope, the present appears disproportionately engorged (Thaler, 2016, pp. 92–3). Favoring present gratification, when deferred gratification may be the more rational course of action to take, reflects lack of self-control, literally a “de-ranged” sense of time, which in the case of mass atrocities can express itself in “orgies” of large-scale initial killing.

A second example concerns how atrocity architects deal with loss, or fear of loss. Frequently, they “dig in” to maintain their position (status quo bias) even if it were rational to change it, for instance, to surrender in the face of sure defeat or otherwise to negotiate an outcome superior to the one they are likely to experience in the absence of negotiation. Standard theory does not
allow for this kind of nonrational behavior. But suppose utility depends not only on the absolute value of one’s expected state of the world but also on differences of that value relative to a reference point (Kahneman and Tversky, 1979). As an empirical regularity, relative to that anchor, humans tend to weigh the prospect of a loss more heavily than the prospect of an equivalent gain. Applied to mass atrocities, Anderton and Brauer (2016) then show that reference dependence and loss aversion can help explain why an atrocity committing dictator may be intransigent and be harder to dislodge from a taken position than standard theory would predict. Relatedly, Midlarsky (2005, pp. 104–7, 135–53) appeals to reference dependence and loss aversion to help explain the rise of Nazi-era war and genocidal ideologies leading into World War II, tracing their emergence in the 1930s in part to drastic German losses in territory, power, and resources following World War I (see also Staub, 1989, pp. 30, 91–4).

Identity economics takes a different tack. To rehearse the context, (1) people define themselves in terms of one or more social categories they choose and/or are assigned to; (2) social status inheres in each category; (3) the set of categories determines people’s identity, their self-image and sense of belonging; (4) identities, in turn, determine how people behave in situations in which they interact with other people; (5) for each situation, people internalize behavioral norms that an ideal-type person in each category would deem appropriate to follow (this ideal may be prescribed); (6) people learn how they and others in a situation are expected to behave (the role each plays); (7) when behavior conforms to the ideal, utility gains occur; and (8) as situations, and the composition of social categories present in a situation, change, people adopt different frames and adapt their behavior accordingly (they play different roles). This captures the fundamental tenet of social psychology: People’s behavior is determined not so much by who they are but by where (that is, with whom) they are. Behavior is situational more
than it is dispositional, and identity is one way in which people frame the situations they are in or are put into. Unlike in standard theory, utility functions are situation-dependent and change when the norms underlying social categories and situations change.

Identity is important in economics (Smith, 1759; Boulding, 1956; Akerlof and Kranton, 2000), and the observation that individuals who identify with each other and cluster into groups of like-minded individuals can benefit society at large has led to a revival of the longstanding notion of social capital (Coleman, 1988; Putnam, 1993, 1995; Becker and Murphy, 2001; Durlauf and Fafchamps, 2005). Yet social capital can become antisocial capital as identity-based groups that include people can also exclude them (Acemoglu and Robinson, 2012). Beyond exclusion, Sen writes that “identity can also kill—and kill with abandon,” and more so when “violence is fomented by the imposition of singular and belligerent identities on gullible people, championed by proficient artisans of terror” (Sen, 2006, pp. 1–2). Demagogues and dictators are good at identity reduction to a binary “us” versus “them.” Technical analyses of how individuals, groups, and (at least implicit) networks interact go back to Schelling (1971), but formal analysis of agents’ personal identity came to widespread prominence only with a series of papers by Akerlof and Kranton (2000, 2002, 2005, 2008).

Applying economic methods to identity, Akerlof and Kranton (2000) propose a utility function $U_j = U_j(a_j, a_{-j}, I_j)$. Suppressing vector notation, in addition to the usual vectors of person $j$’s actions ($a_j$) and other people’s actions ($a_{-j}$), utility $U_j$ also depends on $j$’s identity ($I_j$). Identity itself is modeled with an identity function, $I_j = I_j(a_j, a_{-j}, c_j, e_j, P)$. In words, identity is the outcome of one’s own and others’ actions, of course, but also—and importantly—of “assigned social categories $c_j$” (Akerlof and Kranton, 2000, p. 719). Since social status inheres in social

---

23 Identity groups can also be formed by forcible inclusion, in which attempted internalization of collective benefits into individual’s preferences may go awry and, in time, pose a threat to the cohesion of the identity group (Ille, 2017a).
categories, “a person assigned a category with higher social status may enjoy an enhanced self-image” (p. 719) and consequently a higher level of utility. The $e_j$ term captures the degree to which person $j$ matches the ideal-type of the category, and the ideal-type itself is indicated by a set of behavioral prescriptions, $P$. In standard theory, identity and its determining factors are assumed given and unchanging and therefore analytically ignored. Akerlof and Kranton note that “because identity is fundamental to behavior, choice of identity may be the most important ‘economic’ decision people make ... [and] limits on this choice may ... be the most important determinant of an individual’s economic well-being” (2000, p. 717).

Caruso (2016) contains a direct application to Germany, 1933–45. National identity and its manipulation for political ends, combined with confirmation and status quo biases, can lead to a self-reinforcing mechanism in which selective uptake of news increases one’s level of utility even as society as a whole goes astray. Like Breton and Wintrobe (1986), Caruso notes the division of labor in the diffusion of responsibility for the totality of the horrible acts agents commit. Cognitive dissonance (Festinger, 1957) arising from the clash of agents’ actions with their moral sentiments (Smith, 1759) was managed, Caruso argues, by engineering moral disengagement from the effects perpetrators’ acts had on victims. He writes: “Crucial to the facilitation of large-scale moral disengagement was a hierarchical organizational structure based on the diffusion of responsibility” (p. 323). In this, he also adapts extended work by Akerlof and Kranton (2005, 2008) on additional, intermediary actors, for example, in our context, street block leaders positioned between ideologues and citizens, commanders between leaders and soldiers, or mid-level bureaucrats sandwiched between officials in super- and subordinate bureaus. The introduction of additional actors (“supervisors”) then creates new tradeoffs.\footnote{Tradeoffs of this kind are also explored in Li, Matouschek, and Powell (2017) but couched there in terms of “power” within an organization. They show that “entrenched power structures [can be] consistent with optimal...}
In Caruso’s formal adaptation of Akerlof and Kranton’s concepts to the Holocaust, he shows that the addition of identity to a model of individual choice (for an insider to participate with either high or low effort in mass atrocity) allows the perpetrating nation to get away with offering lower pecuniary rewards to the insider than predicted by standard theory. As Caruso (2016, p. 321) writes: “true Nazi individuals exert a high level of effort irrespective of monetary incentive.” Monetary earnings and identity can become substitutes, with the added benefit of relaxing the ideologue’s budget constraint. (In a model with more than two effort levels, earnings and identity can be complements.) Important here is that identity can be manipulated by changing a citizen’s self-perception, more so when a citizen or group of citizens can be isolated, that is, when their situation can be manipulated as well. Insider–outsider concepts readily apply (for example in regard to nationality and ethnicity) and lend themselves to identity-reducing “us” versus “them” thinking. Investing in defining and designing prescriptions of identity ideal-types and thus investing in identity (re)formation are prominent phenomena in virtually all mass atrocities. Indeed, the branding of others is the first of Stanton’s (2013) ten stages of genocide.

Murshed and Tadjoeddin (2016) also pick up on Akerlof and Kranton’s ideas on identity but, following Sen (2006), introduce a multiple-identity utility function wherein an individual’s total utility derives from two source utilities. The first is tied to an affiliation with a principal group with which the individual identifies; the second is tied to one or more other identities of importance to the individual as determined by a vector of factors relevant to this individual. The first is “like a club good,” the second “similar to a private good” (p. 501). The principal group-
related utility depends on an individual’s own actions, on other people’s actions, and on identity. This group-related identity itself is influenced, as in Akerlof and Kranton (2000), by one’s own actions (for example, to frequent one’s own ethnic group’s food markets or worship with one’s own group) and by one’s co-ethnic or co-religionists’ actions, but also—crucially—by a Glaeser-type (2005) hate parameter ($\mu$) that is subject to manipulation by atrocity-seeking agents such as in-group elites.

If a despot succeeds in driving up the value of $\mu$, then one’s identity aligns more closely with one’s primary group and perceived errant behavior, even by own-group members, generates disutility. It follows that one can then “afford” to inflict counterdeviant measures at greater cost to oneself than before: One becomes more willing to retaliate against deviancy, even within the in-group. Now, if (as in Glaeser, 2005) $\mu$ is viewed as a signal sent by a group leader about “others,” then members will need to judge the likely veracity of the signal. Is the leader’s signal regarding Muslims true? Is Hutu railing on RTML, a radio station, against Tutsi true? This creates search costs to appraise signal veracity. If search costs per message are high or if total search costs add up due to the frequency of hate messaging, group members may succumb to the message. This drives up identity-based affiliation with one’s own group—a “circling the wagons” effect—and prepares members to engage in and bear increased costs of retaliation against seeming deviants. Note that hate messaging can be modeled as a coordination device for group behavior in genocide (Yanigizawa-Drott, 2014).

3.3 Tributary 3: Networks and complexity analysis

Mass atrocities are individually motivated yet aggregate activities of a group, so that beyond identity as a psycho-social phenomenon per se it is crucial to develop models “link[ing]
individual behavior and group outcomes” (Durlauf and Young, 2001, p. 1). The next two
subsections draw on some of the available material and deal with variations of what we call, in
analogy to DellaVigna (2009), the nonstandard self. First, we look at the relation between
individuals and groups in social networks (Section 3.3.1). Second, we examine complex social
systems (3.3.2).

3.3.1 The nonstandard self: The networked self
The incorporation of peer interaction effects, participation thresholds, and entire social networks
into economic analysis has emerged over the last half century as a prominent research field,
along with questions about how norms emerge, cascade through society, and eventually dissipate
as other norms arise. Key questions of network economics—How do networks arise? How does
a given network structure affect agents’ choices in the network? How do networks evolve? How
do information, learning, and imitation spread across networks? How do key players affect
network outcomes? How can networks be manipulated?—matter as much for understanding the
emergence and spread of norms and network effects in cases of financial contagion, disease
edemics, peer effects in education, or the intergenerational transmission of poverty as they do
for mass atrocities. Already there exist network literatures on crime, conflict and defense, and
terrorism with obvious parallels to mass atrocity risk and prevention. Albeit characterized as an
“ensemble of different scenarios” (Dziubinski, Goyal, and Vigier, 2016, p. 217) with a
“bewildering array of different (and sometimes even contradictory) results” (Acemoglu,
Malekian, and Ozdaglar, 2016, p. 570), nevertheless this literature is highly relevant as mass

26 See Alós-Ferrer (2018) for a review of the emerging field of social neuroeconomics. He argues in favor of
bidirectional causality of the development of the human (social) brain and societal development.
27 Brock and Durlauf (2001a, 2001b), Jackson (2014), Bramoullé, Galeotti, and Rogers (2016), Goyal (2016),
28 Cunningham (1985), Baccara and Bar-Isaac (2008), Franke and Öztürk (2009), Goyal and Vigier (2014), Masys
(2014), Acemoglu, Malekian, and Ozdaglar (2016), Dziubiński, Goyal, and Vigier (2016), Kenney, Coulthart, and
atrocities are crimes perpetrated through social networks on social networks (Lemkin, 1944). For example, Feldman and Seibel (2005) emphasize that the Holocaust involved “networks of Nazi persecution” and, according to Adalian (2013, p. 126), the “chain of command that put the Armenian genocide into motion joined every link in the administration of the Ottoman state.”

Begin with the canonical participation threshold models of Schelling (1971, 1978) and Granovetter (1978). To use Granovetter’s example, will 100 people assembled in a public square riot? Suppose player A has a riot participation threshold of zero (a self-acting person), player B has a threshold of one (another player needs to riot first), player C of two, and so on up to the 100th player who has a threshold of 99. Player A may well be ignorant of this frequency distribution but, being an instigator, tests the waters and takes a riotous action such as throwing a rock through a shop window. This activates B to join, which then activates C, and so on until the entire crowd (n=100) riots. The tipping point for the action to diffuse through the crowd consists of the first player. Compare this group with another where player B has a threshold of two. Now when player A breaks the window, no second player joins the action. The riot, such as it is, consists of a lone vandal. Thus, whether and how an observed collective action cascades through a crowd depends as much on the aggregation function as on any underlying characteristics (attributes, preferences, beliefs, norms) its individual members may possess. This leads to a two-fold observation. One, as Schelling notes, is that “[i]nferences about individual motives can usually not be drawn from aggregate patterns” (1971, p. 143). Every player may have a riotous disposition but is waiting for situational clues that may or may not occur. The other is that a trivial marginal change in an initial condition—the addition or removal of a single player—can lead to significant changes in situational dynamics with dramatic consequences for the final outcome, for example whether a newsworthy riot occurs or whether a lone vandal is hauled off
by police officers and remains unbeknown to virtually all others.

Anderton and Brauer (2018) develop a model along these lines in which a village harbors a small number of neighbors ill-disposed toward an out-group. They show how under a baseline set of initial conditions acceptance of an untoward attitude toward outsiders spreads over time and space until the whole village is “infected” with an anti-out-group norm. They then study how trivial changes, for example in the initial geographic location of a single instigator or structural changes in how village members are connected to each other (the spatial distribution of neighbors), can lead to widely divergent outcomes ranging from no acceptance, to partial acceptance, to full acceptance throughout the village of antagonistic feelings toward out-group members. Moreover, their model shows how the spatial pattern of anti out-group feeling changes over time depending of which set of conditions is in place. Even a setup as simple as theirs helps address a number of puzzles genocide scholars have noted such as why some villages (societies) mass participate in atrocities while other villages (societies) with seemingly similar characteristics do not. The model is applicable to a frequently reported phenomenon in mass atrocity case studies, namely, contestation among members within an in-group over whether atrocities will be inflicted on “others” (for example, Browning, 2004, pp. 15-24; Straus, 2006, p. 65; Vági, Csösz, and Kádár, 2013, pp. lvi-lvii). Straus (2006, p. 65) reports that within some Rwandan villages during the genocide a “tipping point was reached; once that happened, mass mobilization of Hutu men swiftly followed, with devastating consequences for Tutsis.” The contagion mechanisms in the model can be generalized to include contagion of atrocity actions across villages, which Straus (2006, p. 93) characterized during the Rwandan genocide as a “cascade of tipping points.”

Jackson (2014) drills more deeply into peer interaction effects. His model can demonstrate
how a person whose own norms do not favor—indeed, oppose—genocidal behavior may nonetheless engage in such behavior if enough other people do. Explaining such seemingly puzzling behavior has vexed genocide writers repeatedly. Drumbl (2016), for instance, quotes Victor Frankl (a major writer on the Holocaust experience) who argues that no one, however constrained, is ever bereft of agency (p. 224, fn 19), meaning that one’s own agency can always override peer effects. In contrast, Drumbl writes of the reality of “tragic perpetrators”—and “imperfect victims” compelled to victimize others—who in spite of their own feelings help commit atrocities (2016, pp. 218, 245). Jackson (2014) illustrates this tension with a logistic equation, \[ \log\left[p_i/(1-p_i)\right] = \beta X_i + \lambda F_i, \] where \( p_i \) is the probability of an individual taking an action, \( X_i \) is a set of characteristics of the individual (the vector notation is suppressed), and \( F_i \) is the portion of that individual’s peers (or friends) who also take or are believed to take the action. Parameters \( \beta \) and \( \lambda \) capture strengths or “contributions” of the individual’s self-influence relative to peer influence.\(^{29}\) Standard theory assumes that \( \lambda=0 \), ignoring peer effects and, like Frankl, thus exalting individuals’ agency. But the larger is \( \lambda F_i \), the larger is the influence of one’s peers, even to the extent that it can override an agent’s own inclinations. Note the multiplicative form of \( \lambda F_i \). A large influence by a small number of peers (“my best friend is doing it”) or a small influence by many peers (“everyone is doing it”) both affect an individual’s decisionmaking. Thus, seemingly inexplicable behavior such as participation by individuals inherently opposed to mass atrocities may be due to situational dynamics in spite of participating individuals’ own preferences: Despite misgivings, one goes along with the group. In network economics, the “puzzle” of mass participation can disappear. Indeed, the “mass” in mass atrocity—perhaps even the majority—may be inherently unwilling to participate and yet goes along owing to peer

\(^{29}\)Banerjee, Chandrasekhar, Duflo, and Jackson (2013) apply this model to the diffusion of microfinance in India. We adopt the structure of the model to show its immediate applicability to mass atrocity diffusion.
influences within the network. Importantly, peer influence can be manipulated by atrocity architects, for example, through hate messaging and propaganda.

Another network model with immediate applicability is Jackson and Zenou (2015). It has the added benefit of modeling conditions under uncertainty. Specifically, agents have incomplete information about the network structure. When choosing action 0 or 1 (do not or do participate in mass atrocity) each player in the network knows her own degree of network connectedness but is uncertain about neighbors’ degrees. Likewise, each player knows her cost of choosing action 1 but not the neighbors’ costs. What emerges is a distribution of player types based on degree and cost. Assuming player neighbors are randomly chosen, and also payoff interdependence (strategic complements or substitutes) between pairs of linked players, behavior in the dynamic version of the model can lead to participation contagion. Jackson and Zenou (2015) derive the expected probability that a player’s neighbor will choose action 1 as a function of the expected probability that a player’s neighbor chose action 1 in the previous period. For strategic complements, the function monotonically increases and is generally nonlinear as shown in Figure 2.

[Figure 2 here]

The figure shows that multiple equilibria can exist \((a, b, and c)\), two of which are stable \((a\) and \(c)\). Applied to mass atrocity participation, equilibrium \(a\) implies zero probability of participation, while equilibrium \(c\) implies a high probability of participation. The figure also implies that initial conditions are critical, a routine finding in network research. If the initial population participation ratio, \(r_0\), lies to the left of \(p_1\), contagion will die out and equilibrium converges to \(p=0\) at point \(a\). If the initial ratio lies to the right of \(p_1\), however, a contagion of growing participation will play out on the network until equilibrium \(c\) is reached. Comparative
statics show that an upward shift in the $\Phi$ function will push the high participation equilibrium ($c$) up and to the right and point $b$ down and to the left (not shown) (Galeotti, et al., 2010; Jackson and Zenou, 2015). This increases the degree of participation in the network associated with the high equilibrium ($c$) but also implies that an initial condition not too far to the left of $p_1$ might now be to the right of the (new) point $b$, thus kicking off a participation contagion toward the higher equilibrium, $c$. Such outcomes due to the upward shift in the $\Phi$ function will also occur if participation costs decrease or if the expected number of degrees of each neighbor rises (Jackson and Zenou, 2015, pp. 137–8).

Models of the Schelling-Granovetter-Jackson-Zenou type have been applied to socially-condoned mass coercion against females including polygyny, footbinding, female genital cutting, and sex-selective abortions. Anderton and Brauer (2018) do the same with a focus on mass atrocity. In their work, an $S$-shaped diffusion curve of a norm of atrocity acceptance emerges from an adapted Easley-Kleinberg (2010, ch. 17) model, with the added benefit that the micro foundations of the diffusion curve are explicit in the form of agents’ reservation (demand) functions, a peer effect function, the (derived) aggregate benefits function, and the per person cost of atrocity perpetration. By manipulating the elements of the model, atrocity architects can attempt to achieve the high participation they seek. As in Jackson and Zenou (2015), when lowering the cost of atrocity participation, the equilibrium threshold (the tipping point) to be crossed will be lowered. Mathematically, this implies that the high participation equilibrium level will increase which, in turn, leads to a greater rate of participation once the (lowered) threshold is gotten over. Alternatively, atrocity architects can attempt to shift the diffusion curve upward by increasing actors’ intrinsic valuations for atrocity (e.g., by propaganda) and/or by promoting greater network benefits from atrocity. Of course, atrocity preventers want to do the

---

opposite to create an alternative diffusion. For example, Staub (1989, pp. 165-6) documents that in the French village of Le Chambon during the Holocaust, “caring was infectious” as key leaders helped “convert” police and inhabitants to helping and resisting Nazi orders to turn over Jews.

As a final example, we turn to the linear quadratic model (LQM) which despite loss of generality allows one to derive explicit solutions and incorporate agent heterogeneities with relative ease.31 One application, again, is in Anderton and Brauer (2018). Figure 3 is a stylized depiction of a social network designed to bring out-group destruction to three villages, A, B, and C. At the network’s center is player 1, the atrocity architect. Players 2, 3, and 4 are regional managers or bureau heads associated with each village. Players 5 to 10 are commanders of troops in their assigned locales. Each player i chooses an action’s intensity, $x_i \geq 0$, assumed to be harm directed against an out-group. Payoffs are given by a linear quadratic utility function $U_i = a_i x_i - \frac{b_i}{2} x_i^2 + \sum_{j \neq i} w_{ij} x_i x_j$, where $a_i \geq 0$ and $b_i > 0$ are benefit and cost scalars, respectively, and $w_{ij} \geq 0$, the peer effect, is the weight or importance that player i places on player j’s action (as in Jackson, 2008, p. 290). Each unit of action $x_i$ brings player i marginal benefits of $a_i + \sum_{j \neq i} w_{ij} x_j$ at marginal costs of $b_i x_i$. The $w_{ij}$ parameter captures strategic complementarities (when $w_{ij} > 0$) among linked agents and thus the increase in i’s wellbeing when j’s positive action interacts with i’s positive action ($x_i x_j$). Multiple reasons for strategic complementarities exist, including comradery among in-group players or information flows among networked agents that enhance the “ideological necessity” of destroying an out-group (Kühne, 2010). Player i maximizes $U_i$ by choosing $x_i$, with all other elements in the equation treated parametrically. From this, a system of

---

reaction functions can be derived and an aggregate amount of actions (harm), $X^*$, can be computed.

[Figure 3 here]

Numerous comparative static and other analyses are possible in the LQM. In addition to policies (for example, sanctions, threats of litigation, forceful interventions) that might affect the benefits, costs, and network synergies of perpetrating agents in the network, one can identify and remove the network’s key player (Ballester, Calvó-Armengol, and Zenou, 2006). In considering more complex networks, one could incorporate the infiltration of atrocity resisters into the network (that is, negative $w$ terms when interacting with atrocity perpetrators) or a prevention network in contestation with a perpetration network in which outcomes are determined by strategic positioning and Tullock-like contest success functions. Note, especially, that this and other network models are formulated in terms of agents’ utility functions which implies that behavioral and identity economics can be added to the network specifications.

3.3.2 The nonstandard self: The complex self

We discussed work focused on standard payoff functions (Section 3.1) enriched by concepts from behavioral and identity economics (3.2) and also work focused on behavior aggregation in social networks (3.3.1), whether built up from enriched utility functions or not. We now add another rubric: complex adaptive systems (CAS). First we discuss atrocity-relevant examples from evolutionary game theory (EGT) and then from agent-based models (ABM).

The dynamics of social interactions can determine group-level outcomes that become self-reinforcing: “Everyone conforms, everyone is expected to conform, and everyone wants to conform when they expect everyone else to conform” (Young, 2015, p. 359). Modeling social dynamics turns out to be more complex than modeling social equilibrium (Durlauf and Young,
2001). This is because complex systems “resist analytical derivation from the behavior of the individual parts” and because collective outcomes can be “emergent properties” of the system itself (Gintis, 2016, p. 110). Evolutionary models are defined by individuals interacting in a semi-autonomous and structured manner in a complex system characterized by recursive and nonlinear dynamics. In contrast to models with strategically interacting, payoff maximizing individuals under more or less strong rationality assumptions, complex adaptive models feature autonomous, heterogeneous agents following relatively simple and limited numbers of rules (heuristics) with resulting path dependency, complex unexpected behaviors, and emergent system properties. The central theme is not equilibrium but (constant) adjustment, either of agents in given systems and/or of systems with given agents (Wilson, 2016).

To illustrate models of this type, we start with evolutionary game theory (EGT). For example, Arce and Sandler (2003) apply EGT to study pairwise interactions between individuals of fundamentalist (more intolerant) and nonfundamentalist (more tolerant) subpopulations. For fundamentalists to prevail, they either must display greater initial cohesion (less variance) within their subpopulation in making demands against nonfundamentalists or else face a relatively acquiescent nonfundamentalist subpopulation. To induce nonfundamentalists to behave fundamentalistically (for example, in outward appearance or religious observance), fundamentalist demands in pairwise interactions with nonfundamentalists must also appear reasonable, not strident. If successful, this can induce among nonfundamentalists what Kuran (1989, 1995) calls “preference falsification,” the public display of behavior privately opposed (akin to Michaeli and Spiro, 2017). This “chameleon” or “counterfeit” behavior by

---

32 For a comparative review of EGT with ABM, see Adami, Schossau, and Hintze (2016). We do not address complex networks, a confluence of network analysis and complex systems. Some scholars may view machine-learning and data-mining models (for instance, Douglass, 2016, on mass killing during the Vietnam War) as “complex” also, but here we regard them as pattern extraction and, possibly, forecasting and prediction tools (covered in Section 4) more than as explanatory tools.
nonfundamentalists adapting to fundamentalist demands in order to “fit in” can lead to seemingly tolerable coexistence between subpopulations even as society as a whole drifts into harmful behaviors such as the committing of mass atrocities under a fundamentalist ethos. Once the game has evolved, a shock sufficiently sharp to change the relevant proportion of preference falsifiers among the interacting subpopulations can reset the social dynamics to evolve toward more tolerant societies (for instance, in Germany and Japan following their defeats in World War II). Peaceful contravention of fundamentalism is studied in Arce and Sandler (2009), with the key result revolving around the degree to which policy can help discourage counterfeit behavior and encourage the public display of one’s private beliefs without falsification.

Ille and Mansour (2015) present an EGT model to study the mass atrocities-rich attempt by ‘Islamic State’ (IS) to capture territories and establish itself as a recognized sovereign state in international relations. It is modeled with two payoff functions, one when an atrocity strategy is played, the other when a nonatrocity strategy is played. The atrocity function states that payoffs are positive for IS when the number of atrocities committed rises and/or when the direct return per atrocity committed rises. But the payoff also depends on the opponent’s (potential intervener’s) behavior. This constitutes the cost part of the atrocity payoff function. If the opponent’s actual level of intervention is viewed by IS as half-hearted (below a threshold), then its atrocities are in essence vindicated and thus increase the overall payoff value, and vice versa. The nonatrocity payoff function states that if IS does not commit atrocities, it bears a reputation cost from not pursuing its universal entitlement claim. For the intervener, two corresponding payoff functions are modeled. In terms of results, an interior solution is driven by the initial levels of atrocities committed and interventions done, respectively. For example, high public support for intervention on account of a low level of atrocity “acceptance” (the intervener
“tolerates” some atrocities up to an acceptance level) serves as a credible threat of intervention. This level of support keeps atrocities relatively low and, therefore, requires few interventions. Conversely, seeming apathy toward atrocities goads IS to commit more of them. But the model’s comparative statics reveal that public support can go awry. Keeping all parameters constant except for the “return to intervention” parameter, which reflects the strength of public support for intervention, can lead either to evolutionarily stable mixed strategies (ESS) or to a mutual corner ESS in which pure strategies of atrocities and interventions prevail. This outcome depends only on the initial levels of atrocities and intervention. Even with high public support, if interventionists underreact to initial atrocities, the resulting escalation equilibrium is stable. The reason is that even when IS loses from any given intervention, committing atrocities can carry net benefits inasmuch as they provoke attention to its cause. Reducing IS’ benefits from committing atrocities requires, the comparative statics reveal, not merely public support for intervention per se but also that they be effective and enduring (preventing IS from capitalizing on atrocities) and that they start off with low atrocity acceptance levels to begin with. Inadequate disgust can lead to (stable) escalation. Findings like this should give genocide scholars and intervention policymakers pause to consider more fully the requisites for and dynamics of intervention.

Other EGT applications related to mass atrocities include Anderton (2015), Gangopadhyay (2016), and Young (2015). Anderton (2015), for example, finds that policy “dithering” (intervention delay) is harmful inasmuch as it permits the gradual accumulation of more “aggressive” types until a threshold is reached beyond which the social system as a whole switches from peaceful to aggressive and future intervention may no longer be effective in preventing genocide. Gangopadhyay (2016) presents a complex group identity-based,
evolutionary threshold model with endogenous equilibria selection for heterogeneous societies in which individual in-group agents incorporate reference group considerations and may engage in killings of members of an out-group even if their own reference group opposes such action. The model captures (1) the underlying game of interactions among society’s agents, (2) the economic structure that affects payoffs from mass killings, and (3) how killing or nonkilling strategies diffuse through a population. A key result is that a population’s average propensity to carry out mass killing, or the proportion of the population “programmed” to choose a level of mass killing greater than a certain threshold, determine whether the equilibrium is a mass killing society or not. Finally, we note that Young (2015) analyzes the evolution of social norms using stochastic evolutionary game theory. Among his many applications, Young offers a brief analysis of footbinding, arguably a mass atrocity perpetrated upon females.

Like EGT, agent-based models (ABM) have been applied to numerous forms of human interactions such as peaceful mass migration, evacuation management, general crime, organized crime, terror, cyber violence, and war. Direct applications to mass atrocities have also emerged. For example, von Briesen, Bacaksizlar, and Hadzikadic (2017) built an ABM model with six agent attributes: identity, ideology, exertion of influence, susceptibility to influence, threshold to act, and radius of sight (distance to neighbors). Attributes are combined with rules regarding agent interaction, information updating, action-taking, and the passing on of parent traits to offspring. Starting from a population split equally by an identity attribute but otherwise facing the same conditions (save for random initialization) leads, in all model runs, to the stark result that genocide always is the emergent outcome! In the model, the mere concept of ‘the other’ suffices to produce this result. However, due to random initialization, the number of generations

---

it takes to arrive at the genocide outcome varies considerably across model runs. The authors observe that stable boundary regions between identity groups can persist for lengthy periods of time. Further research will need to explore extensions, for example, the introduction of a third identity group or nonrandom initial distributions, but even the base result raises important policy questions such as whether characteristics of boundary agents differ appreciably from those of nonboundary agents located elsewhere in the model space. How do the characteristics of boundary agents change over model generations? Are there characteristic system-level changes just prior to the successful breakdown of a boundary that then lead to the elimination of one group? Boundary protection or group containment control is also the subject of related research in social network models (Kan, et al., 2014) wherein groups, or leaders of groups, seek to penetrate ideological or other boundaries of another group and, simultaneously, seek to inoculate their own borders from susceptibility to invasion.

In another example, Ille (2017b) uses EGT and ABM to study competing sectarian identities. Ethnic factionalism—one way in which sectarianism can be expressed—has been taken into account by scholars but usually on an a priori basis. It is studied by assumption rather than in a co-evolutionary context. For Ille, in contrast, conflict between identity groups must evolve out of the model. The coercive nature of sects, free-riding within sects, and social inefficiency (due to sectarian conflict) can, but need not, arise from within the model. Run with various initializations and parameter combinations, the results show that “sectarian identities are not perpetual, but constructed by a process of past interactions” (p. 15). While sectarian social contracts are likely to evolve under certain conditions, sectarian affiliation per se does not cause sectarian conflict; affiliation just assigns a role within a situation. In contrast, “externally reinforced differences along sectarian fault-lines obviate the emergence of egalitarian social contracts and can lead to
sectarian conflict” (p. 16). The reason is that, in the model, external mobilization can lead to coordination failures and contests over a sect’s position within a social contract, or else that within-sect imbalances can generate cross-sectarian conflict and further imbalances within other sects. Unlike Golman, et al.’s (2016) discussion of agents’ preference for belief consonance (Section 3.2), for Ille (2017b) the apparent “narcissism of small differences” is not due to “minor ideological disparities but [to] a divergent perception of the role assigned to the sectarian marker” (p. 16). This type of modeling is important as it picks up on why some societies with diverse ethnic compositions may collapse into mass atrocities while others do not.

3.4 Tributary 4: Political economy

Political science emphasizes the study of political institutions (principles, organization, and methods of legitimate authority). Behind institutions lie the acquisition, allocation, use of, and change in power, including the use of threat and coercion. Invariably, cost-benefit considerations enter. We illustrate aspects of political economy and mass atrocity by selectively discussing types of dictatorship (Section 3.4.1), ideology and the use of hatred (3.4.2), and the struggle for societies in transition from dictatorship to democracy (3.4.3).

In models of dictatorship, the desire for power enters the utility function as an objective to be achieved; mass atrocity may be a means of achieving power. However, the objective may instead be to advance a particular ideology (or otherwise achieve infamy); power, including the possible use of mass atrocity, then is among the possible means of realizing the goal. The study of transition to democracy is important for at least two reasons. First, it may help identify necessary conditions to make the transition smooth and bloodless, more so since periods of transition are widely recognized as fraught with danger (Fein, 1995; Mann, 2005; Esteban, Morelli, and Rohner, 2015). Second, however, while desirable in many respects, it turns out that democracy is
neither necessary nor sufficient to prevent mass atrocity. Nondemocracies need not spiral into episodes of mass atrocities, whereas democracies can descend into mass slaughter (Mann, 2005; Madley, 2016).

3.4.1 The nonstandard self: The dictatorial self

Absent political competition, societies can descend into dynastic dictatorships (Olson, 1993), often characterized by unbridled rent-seeking and kleptocracy. To kleptocratic consumption, Wintrobe (1998) adds the desire for power itself to a dictator’s utility function, and power requires expenditure on repression and/or on purchasing the loyalty of subject populations. Relying wholly on standard theory extended to the political realm, Wintrobe generates a two-by-two schema reflecting four ideal-type dictatorship equilibria under high or low levels of repression and loyalty, respectively (Figure 4).

Different constellations of circumstances result in different equilibrium combinations of repression and loyalty. To deal with what Wintrobe characterizes as the Dictator’s Dilemma—“the greater the dictator’s power, the more reason he or she has to be afraid” (1998, p. 22)—both, tyrant and totalitarian seek recourse to high levels of repression. But a tyrant sees less reason to purchase the loyalty of needed retainers, or any portion of the population at large, than does the totalitarian who, for example, may need to placate members of a totalitarian party. In contrast, “tinpot” dictators and “timocrats” (from the Greek thymos, to love or respect) both seek low levels of repression. The timocrat does so because he believes that his “love for the people is reciprocated by them” (p. 14). In contrast, the tinpot dictator’s interest lies in resource

35 The dilemma, Wintrobe points out, is two-sided: “As much as dictators want to be loved, the subjects want them to believe that they are loved, for only then are the people safe from them” (1998, p. 22).
exploitation for his own kleptocratic enrichment, for which purpose little more than a palace guard will do, but not otherwise to “disturb the traditional way of life of the people” (p. 11).

Wintrobe (1998) rarely addresses issues of mass atrocity directly but his examples, especially those regarding totalitarian regimes (Hitler, Stalin), point to conditions under which mass atrocity can flourish. Verwimp (2003a), however, constructs a direct application of Wintrobe’s totalitarian dictator model to Rwanda in the 1980s and 1990s. The paper is relevant in several respects: First, it is both theoretical and empirical; second, it integrates fluctuations in international commodity markets into the analysis (in this case, the world price of coffee); and third, it shows that adverse external shocks affecting the coffee market contributed directly to then-President Habyarimana’s difficulty to continue to buy the Wintrobian loyalty of the affected population which, for Habyarimana to stay in power, then necessitated an increasingly repressive dictatorship. For a time, however, reliance on repression was mitigated by compensatory foreign aid inflows that helped stabilize Habyarimana’s regime until the genocide erupted so explosively in 1994.

Somewhat neglected in the analytical literature, the international political economy links are interesting as even a cursory reading of case studies reveals many instances of mass atrocities tied to external events. Indeed, in her review of the Rwandan genocide, Friedman (2016) goes beyond the world coffee market in adding the world tin market whose price also collapsed. As tin is mined rather than cropped, it affected different population segments. Since coffee and tin had been Rwanda’s number one and two export earners, the opportunity cost of rebellion fell, a theme also modeled and estimated in Gangopadhyay’s (2016) study of Sunni-on-Shiite mass killings in Pakistan, 1978–2012, in which declining world sugar prices play a significant role in increasing the killings. Regarding the role of foreign aid, Marriage (2016) discusses the
“compatible logics” of peace and mass killing in the nominally postconflict Democratic Republic of the Congo (DRC) in the wake of the Second Congo War (1998–2003). Whereas much of the economics-oriented literature believes that “addressing the economic functions of violence could throw its mechanism into reverse and lead to peace,” for instance by applying global economic levers to compel misbehaving dictators to change course, Marriage argues that the very opposite applies to the DRC: “… mass atrocities persisted and profits continued to accrue to political elites, those of neighboring countries, their armies and allied militias, and foreign companies [and that] a bargain was struck between northern donors and elite Congolese politicians in the peace agreement that provided unmonitored aid and facilitated the rapid, unregulated liberalization of assets”—that is, the exploitation of the country’s natural-resource wealth (pp. 356–7). The intent to reverse Marriage’s “economic function of violence” with counter-violence economics perhaps may be read as a form of “motivated belief” among the elite of global political and economic decisionmakers. If so, this may carry dire implications for postwar peace and reconstruction efforts, which frequently fail and lead to the resumption of violent conflict (Hartzell, 2017).

3.4.2 The nonstandard self: The ideological self

In the foregoing, power is the objective and ideologically-driven hatred can be among the means to rile up an in-group. But what if ideology itself is the objective and power is part of the set of constraints (Bernholz, 2001)? Informal power is a means to foster the spread of the ideology until formal power is gained. Once gained, formal power can then be “used to tax and to use government revenues to finance the conversion of believers, the removal of inconvertibles and the effort to increase the probability of spreading the totalitarian regime internationally, which both may imply the use of force” (p. 36). Pushed by their beliefs, adherents are willing to expend
limited resources, and consequently sacrifice a measure of consumption, in order to gain converts and to punish inconvertibles who are viewed as enemies. This captures the possibility of utopian atrocity architects (Valentino, 2004).36

Like Wintrobe (1998), Bernholz’s (2001) technical work relies on standard tools. Unlike Wintrobe, he directly raises the issue of mass atrocities that may be “needed” to cow or convert unbelievers into submission to the one true faith, or else to eliminate them. Bernholz further argues that the nature of ideology, with its claim of universal reach, is all-embracing in outlook and imperialistic in action. A recent example would be that of ‘Islamic State’. It follows that an ideology either does become universal, or else must peter out geographically, its domestic aims having been reached. In the latter case, the ideology “matures” and may become an territorially confined ideocracy more than it becomes a totalitarian regime, hence perhaps less reliant on terror and repression—another distinction between Bernholz and Wintrobe.

Regardless of whether power or ideology is the objective, atrocities require on-the-ground perpetrators, and instilling envy, spite, malice, or hate are among the tools that can be employed to motivate them (Fehr and Schmidt, 1999; Choi and Bowles, 2007; Hillman, 2010; Gershman, 2013; Levin, 2014). In a foundational model, Glaeser (2005) examines the instrumental creation and use of ideologized hate for political ends. He posits that when audiences are willing to listen to hate (demand), then political leaders can create hate (supply). For politicians, the relevant calculus revolves around whether any designated out-group is important to the political issue at stake and whether the group is sufficiently isolated for it to serve as an effective casus belli against which to rally one’s own adherents. For listeners, the relevant calculus revolves around

---

36 In contrast to Bernholz (1997), Bernholz (2001) no longer requires lexicographically ordered “supreme values.” Note that ideology may be a stated goal when the real objective may be power. For example, Ferrero (2017b) studies the “stated goal of ethnic cleansing” in the Serb-Bosian war in the 1990s. Alternatively, Azam and Ferrero (2017) develop a model in which a quest for infamy, more than ideology, drives mass atrocity.
the cost of investigating the veracity of politicians’ hate stories set against any private benefit derived from learning the truth about such stories. Hate will not spread when listeners have sufficient private incentive to seek out the truth. This incentive is subject to manipulation by politicians. If a listener’s private benefit can be lowered, or if the cost of learning the truth can be increased, hate may spread. Incessant repetition and unending variation in hate stories, for example, increase the cost of verification. Glaeser applies the model to three cases, all of which relate to mass atrocities: (1) anti-Black hatred in the American South; (2) political anti-Semitism in nineteenth and twentieth-century Europe; and (3) modern anti-Americanism in the Middle East. For instance, in a counter-example to help make his point, Glaeser writes that in contrast to much of Europe “political anti-Semitism was generally absent from Italy from 1860 to 1935 … Like Germany, Italy was unified by a king whose supporters were the postunification right, but Italy’s right wing was implacably opposed by the Church because the unification of Italy in 1871 had involved the expropriation of Papal property. Pius IX excommunicated the King and pretty much anyone else involved in Italian politics. Since both the right and left were anticlerical, Jews were spread across the aisle and were politically irrelevant. As the model suggests, when the outgroup does not differ in a policy-relevant way, hatred serves no purpose” (p. 75). In Germany, of course, things were different. Whereas Glaeser (2005) eschews adopting any notion of identity for his model, Murshed and Tadjoeddin (2016) meld Glaeser’s economics of hate model with Akerlof and Kranton’s (2000) and Sen’s (2006) ideas concerning (multiple) identities to formulate an extended model then applied to mass atrocity in Indonesia in the mid-1960s.

3.4.3 The nonstandard self: Democracy and the institutional self

Unlike Wintrobe (1998) or Bernholz (2001), for Acemoglu and Robinson (2006) the concern lies less with varieties of ideologies and dictatorships than with how democracies may arise (or fail
to arise) out of forms of dictatorship. In this they resemble Olson (1993). Their tools, however, are game-theoretic, aimed at elucidating the strategic nature of the interaction between rulers and ruled and anchored around the question of whether and how rulers can credibly commit not to abuse whatever political privilege the ruled may grant. Their work helps to illuminate the conditions under which mass atrocity can arise, operate, and dissipate. At issue are distributional conflicts, over power as well as over income, between two idealized groups. Enriched by assumptions on the relative size of the populations and the nature of the political power one or the other (sub)group may possess, numerous model permutations and associated findings emerge. For example, if despite a numerical disadvantage, the “rich” wield political power sufficient to sway even an otherwise democratic-looking, median-voter setup, then, in the extreme, the (re)distributional outcome may not differ at all between democracies and nondemocracies (p. 116). Democracy alone does not guarantee the absence of mass violence (see also Mann (2005) who, in the words of Bloxham and Moses (2010, p. 9) “debunks the spurious notion that democracies do not engage in genocide, and thus the idea that the spread of (capitalistic) democracy is the antidote to genocide as well as war.”). Acemoglu and Robinson (2006) stay within the territorial boundaries of a political entity, usually conceptualized as a nation-state. Mass atrocities, however, can be carried out across such boundaries, for example, imperial Japanese atrocities committed in Nanjing, China.

Verwimp (2004a) may be read as an applied example of Acemoglu and Robinson (2006). Discussing Rwanda’s polity prior to the 1994 genocide, he sets up a model of nested games where each game’s payoff depends on the other games’ payoffs. From 1990–3, Rwanda’s autocratic regime (the “dictator” player) faced domestic opposition in a transition game over democratization and, simultaneously, faced a rebel group in a civil war game. Payoffs in each are
affected by a fourth player, the “international community.” In the transition game, the dictator’s preference ordering is $B > B - C_T > B/2 > 0$, where $B$ is benefits (control over the national budget) and $C_T$ is the cost of buying the loyalty of, or else repressing, the opposition. So long as $C_T < B/2$, this is better than sharing budget resources ($B/2$) with the opposition. For the opposition, the preference ordering is $B > B/2 > R > 0$, where $R$ is the cost of resisting the dictator (with $R < B/2$). As it turns out, the Nash equilibrium is an unsatisfactory outcome for both players and not Pareto-efficient. To permit any other outcome, either the preference orderings or the game structure has to change, a point to which we return momentarily. Meanwhile, in the civil war game, the dictator’s initial preference ordering is $B > B/2 > B - C_W > 0$, where $C_W$ reflects the war costs. For the rebels (the Rwandan Patriotic Front, RPF), $B > B/2 > M > 0$, where $M$ reflects RPF’s gain under mutual warfare. Here, a prisoners’ dilemma results with war as the outcome. As the dictator plays two separate, nonnested games, he examines the combined payoff function (which can include a term to weigh the relative importance to the dictator of the transition and war games).

However, the games can interact in that the rebel group’s actions (the RPF gained ground from 1990–3), and even its very presence, can spur the opposition to increase its resistance to the dictator which, in turn, can increase the dictator’s cost of buying off the opposition or else spend more to repress it. Still looking to maximize the payoff from the combined payoff function, this can lead the dictator’s preference ordering in the transition game to switch such that the dictator faces prisoners’ dilemma equilibria in both of the nested games. Nesting, or linkage, changes the games. An otherwise noncooperative dictator in nonnested games may feel compelled to cooperate in one or more games when they are nested, presumably an outcome the dictator

---

37 In the model, the dictator is a unitary actor. Divisions, and hence pressures, within a dictator’s party, can foster intransigence as discussed in Section 3.2. Also see Anderton and Carter’s (2019) textbook treatment on nonunitary elites.
dislikes. The dictator therefore possesses an incentive to change the game’s very structure, which might involve the removal of a player in one game thus collapsing two games into one remaining (institutional) game. Since rebels are war-like, it is natural to focus on eliminating the nonwar-like opposition player in the transition game. This, Verwimp states, is what happened in Rwanda, where “elimination” meant splitting the opposition into two camps, one aligned with the dictator (Habyarimana) and one that joined the rebels (p. 19). This involved killing “moderate Hutu,” demonizing all Tutsi as “rebels,” and creating the impression of an “ethnicity”-based conflict. By 1994, this also meant killing Tutsi altogether, so that “the genocide of the Tutsi minority can be understood to deal with another institution or rule of the game, namely its infinity” (Verwimp, 2004a, p. 37). If one kills moderate Hutu, one collapses two games into one; if one kills all Tutsi, one no longer has to worry about playing any game at all (p. 37). The threat of a transition to democracy (sharing B with the opposition) can hasten mass atrocity.

3.5 A summative statement

Standard theory contains rich concepts and tools with which to (re)interpret instances of mass atrocities across a wide range of cases and, as such, can be helpful in comparative studies. Within well-established theoretical frameworks, it facilitates reflection upon perpetrators’ and victims’ constraints, substitution possibilities, strategic behavior, the potential to incentivize (mis)behavior, and the potential for and limits to effective intervention. Behavioral and identity economics add to the standard toolkit, for example, by exploring the nature and role of personal and group identities in preference (re)formation and how they may open up, or constrain, an

---

38 Verwimp then adds foreign aid to the model, relaxing the dictator’s budget constraint or, due to fungibility, reducing the dictator’s costs in either game. If aid is of sufficient size then a “peaceful” dictatorship can result, more so when aid is unconditional in which case the games revert to two nested games, and the dictator no longer need choose cooperation in either, thus obviating the incentive to redesign the games. In Rwanda’s case, unconditional aid was large. In its absence, Verwimp speculates, the collapse of export earnings might have compelled the regime to cooperate. But with aid flowing, despite ample evidence of initial instances of mass atrocities occurring, the regime wagered on ending the game altogether: genocide.
actor’s conception of what beliefs and decision choices are possible. Knowledge about psychosocial biases, motivated reasoning, and bounds to rationality all can help to explore case studies of mass atrocity with theoretically grounded tools. As these advances still seek recourse to standard theory’s core of utility and production functions, the issue is not “rational” versus “nonrational” actors, but to incorporate both possibilities within a net payoff-maximization framework. While it surprises how few applications one finds of behavioral economics to mass atrocities (Anderton, 2014b), perhaps the greater promise lies with the line of inquiry begun by Akerlof and Kranton (2000) which carries numerous implications, for instance in regard to the nature, role, and power of intermediaries in groups and the manipulation and management of identity.

This leads to the modeling of social networks and complex systems. Inherent and manufactured identities, motivated beliefs, and cost-benefit-driven individual adoption of roles in social situations can aggregate in networks which, in turn, can lead to unpredictable feedback effects on sets of peers in local neighborhoods and the individuals therein. Virtually any observed mass atrocity outcome is, in principle, amenable to modeling founded on network analysis and can help explain a large number of puzzles scholars of mass atrocities have struggled with. We highlight three of them. First, while social psychologists long have known that it can be frightfully easy to recruit relatively small numbers of ordinary people into “becoming evil” (Waller, 2007; Roth, 2010), network models additionally explain mass participation in mass atrocity. Second, these models can explain why mass atrocity spreads in location (or society) A yet stays contained in location B even if individuals in both locations share identical characteristics. Third, the models can explain why people who are strongly opposed to acts of mass atrocity nonetheless help commit them. Network theory can provide

39 See Anderton and Brauer (2018) for more.
compelling insights into these behaviors without appealing to wildly changing or bizarre preferences or circumstances (Gintis, 2009). The puzzle no longer is why mass atrocities can arise; instead, valid questions remain regarding how they arise and play out in specific cases. This is a considerable advance in scholarship and begins to make thoughts regarding timely and effective intervention, or even prevention, potentially tractable. Finally, perspectives of political economy, usually also built upon the core of standard theory, can help to investigate mass atrocities across the spectrum from dictatorship to democracy in theory-grounded, formal ways. Among other findings are the caution not to fetishize democracy for its own sake and to consider nested games played by the various vested interests involved.

We conclude this section with five summary observations. First, despite their diversity, the surveyed approaches are unified in their focus on the cost-benefit (or net payoff or fitness) calculus implied by Robbins’ (1937) choice-theoretic view. Second, standard theory alone cannot adequately explain mass atrocities; broader, more comprehensive explanations require the whole gamut of approaches. Third, scholars’ predominant interest lies in the actions of perpetrators, certainly more so than in the (in)actions of bystanders (including potential mass atrocity interveners) and one finds an almost complete absence of studying the feasible choice set of (potential) victims and how their constraints may be lessened. Fourth, tightening constraints on perpetrators need not symmetrically loosen those of victims, and may worsen them (unintended “backfire” effects of intervention). Fifth, social networks and entire social systems can take on a life of their own and spiral out of control. Even when no one in particular intends any great harm to anyone else, system effects can take over and “direct” the future course of events. Thus, personalized theories that focus on individual culpability or victimhood in mass atrocity have to be complemented by theories of behavior in social contexts.
We now turn to a discussion of empirical evidence.

4. Actors and Their Choices: Empirical Evidence

This section surveys (Section 4.1) and summarizes (4.2) empirical evidence on risk factors and seriousness of mass atrocities. This work spans five tributaries: (1) large-\(n\) cross-country studies of the correlates of high-level mass atrocities (samples typically include cases with 1,000+ fatalities), (2) large-\(n\) cross-country studies of the correlates of “low-level” violence against civilians (VAC) (cases or events with <1,000 fatalities), (3) empirical determinants of atrocity seriousness (for example, number of fatalities or number of attacks), (4) forecasting, and (5) micro-level studies that emphasize individual participation in atrocities and/or the spread of VAC across locales for specific atrocity cases.\(^4\)

4.1 Survey of empirical evidence

4.1.1 Large-\(n\) studies of cross-country evidence of high-level mass atrocity risks

The first generation of cross-country empirical studies emerged in the work of Rummel (1995), Krain (1997), and Harff (2003) and focused on the risks of government perpetration of mass atrocities. Rummel (1995) used factor analysis to ascertain risk factors for democide, “the intentional killing of people by government” (p. 4), which includes genocide and politicide. Krain and Harff, however, used a logit specification to assess risk factors for genocide and politicide and their dependent variable is an indicator of the onset of either event in country \(i\) in year \(t\). Each includes \(k\) lagged independent variables given by vector \(X_{it-1}\) proxying political, etc.\(^4\)

\(^4\) A comprehensive survey of the rapidly growing empirical atrocity literature may require a book-length treatment, similar to Geller and Singer (1998) for the literature on interstate conflict and Collier and Sambanis (2005, vols. 1 and 2) and Mason and Mitchell (2016) for the literature on civil war. In our limited survey here, we focus upon major risk factors for high- and low-level atrocities and atrocity seriousness across wide swathes of literatures without focusing heavily on how results vary across types of actors (for example, governments, rebels, militias) or atrocities (for example, genocides, politicides, mass killings, VAC).
social, and economic conditions that each anticipates will affect the onset probability,

\[
ONSET_{it} = \Lambda(\beta_0 + X'_{it-1}\beta_k + \varepsilon_{it}).
\]

Based on her own dataset of genocides specifically (rather than mass atrocities broadly) that occurred since 1955, Harff notes that almost all occur contemporaneously with or shortly after a civil conflict or regime collapse. The inverse does not follow: Many civil conflicts and regime changes do not involve genocides. Given civil conflict or severe regime change, Harff finds six factors that significantly elevate genocide risk: (1) large political upheaval, (2) autocratic regime, (3) political elite holds an exclusionary ideology, (4) the ruling elite is from an ethnic minority, (5) history of prior genocide, and (6) a country is less open to economic trade. From a baseline probability of about three percent, Harff calculates that each of the six factors increases genocide risk between two and six percentage points. When all six factors occur together, genocide risk rises to 90 percent. Krain meanwhile finds that a particular form of political upheaval—civil war—is a significant predictor of genocide risk across all of his regressions, whereas other factors such as ethnic fractionalization, government concentration of power, and a country’s percentage of world trade are not.

Many of the cross-country risk studies that followed Krain and Harff also employed a logit model (or probit or rare events logit), but with different samples of countries, time periods, and sets and/or measures of explanatory variables. In addition, some scholars employed alternative measures of the dependent variable including genocide occurrence as distinct from onset or mass atrocity onset or occurrence as distinct from genocide. Moreover, a few scholars employed

41 To economize, we use “genocide” here to include politicides. This does not mean that we view genocides and politicides as conceptually the same. They are distinguished in the Political Instability Task Force Geno-politicide (or PITF-G) dataset and in some empirical studies (for example, Uzonyi, 2015, 2018).
hazard models rather than or in addition to logit (for instance, Wayman and Tago, 2010; Anderton and Carter, 2015). One study, importantly, assesses the risk factors for mass atrocities conducted by nonstate actors (rebel groups and pro-government militias) (Loyle, 2018). Such studies generally consider most or all of four intersecting classes of independent variables: (1) upheaval and threat such as threats to leaders’ political or territorial control or opportunities to eliminate rivals, (2) state political regime and institutions such as autocracy/democracy or constraints on executive leaders, (3) social fragmentation such as polarization/fractionalization or political and economic discrimination, and (4) economic conditions such as development, natural resources, and trade.\(^\text{44}\) We now summarize each of these.

First, \textit{upheaval or threat}. The most robust finding to date is that, all else equal, proxies for upheaval and threat are usually significantly associated with greater mass atrocity risk.\(^\text{45}\) This includes proxies for civil wars (or interstate wars or wars more generally),\(^\text{46}\) number of years of prior upheaval (Hazlett, 2011), types of leaders that come into power such as revolutionary leaders or irregular leadership changes including coups and assassinations (Brehm, 2017a; Kim, 2018), degree of civilian support for rebels (Valentino, Huth, and Balch-Lindsay, 2004; Kim, 2010), new state status (Anderton and Carter, 2015), and interstate rivalry (Uzonyi, 2018).

Second, almost all studies include a measure of countries’ \textit{political regime and/or institutional constraints} on political leaders. Among the many that include political regime level,
about half find that democracy correlates to low, and/or autocracy to high, mass atrocity risk while the other half find no significant effect. Four studies (Fein, 1995; Hazlett, 2011; Anderton and Carter, 2015; Kim, 2018) find that mixed political regimes (anocracies) are at greater risk than either autocracies or democracies. Importantly, Fein (1995) and Esteban, Morrelli, and Rohner (2015) find that transition from nondemocracy to democracy elevates risk even when controlling for political regime level. Meanwhile, Colaresi and Carey (2008) and Brehm (2017a) find that constraints on executive leaders reduce mass atrocity risk (see also Hazlett, 2011).

Third, social fragmentation exists when groups within a society form parallel economic, political, and cultural structures with relatively little interaction between them and in which some groups are closed off to opportunities available to others (Waller, 2016, p. 181). Various theoretical rationales have led to numerous variable measures for social fragmentation in empirical studies including ethnolinguistic fractionalization (ELF), ethnic polarization, ruling elite ethnicity and/or exclusionary ideology, political and/or economic marginalization or discrimination, and identity- (ethnic, religious) or ideology-based conflict. Among studies that include ELF (or close variants thereof), most find that the coefficient estimate is generally not significant, although there are exceptions. Easterly, Gatti, and Kurlat (2006) is the only empirical study of which we are aware that considers a nonlinear effect of ELF on mass atrocity risk, finding that low ELF corresponds to high risk and vice versa, all else equal. Their result ties nicely to empirical tests of ethnic polarization because very low levels of ELF (that is, low diversity) correspond to high levels of ethnic polarization and growing values of ELF correspond to low indices of polarization in theoretical models (Montalvo and Reynal-Querol, 2008, p.

47 For the counter-example, social cohesion, and some conditions therein, see Jha (2007).
Empirical results on the effects of ethnic polarization on mass atrocity risk are mixed.\textsuperscript{49} Mixed results have also been reported in studies that include a proxy for exclusionary ideology and/or elite ethnicity\textsuperscript{50} or the presence of identity (ethnic or ideological) conflict.\textsuperscript{51} Of the few studies that include political and/or social marginalization, all find evidence of positive and significant effects on mass atrocity risk including for measures of political marginalization or discrimination (Kim, 2010; Brehm, 2017a), economic discrimination (Ulfelder and Valentino, 2008; Anderton and Carter, 2015), general (political and economic) marginalization or discrimination (Anderton and Carter, 2015; Brehm, 2017a), and exclusionary ideology (Harff 2003, Uzonyi 2018).

Fourth, regarding \textit{economic conditions}, most studies include a measure of real GDP per capita to proxy level of economic development, state strength, and/or the opportunity cost of participating in violent conflict.\textsuperscript{52} But only Scully (1997), Easterly, Gatti, and Kurlat (2006), Anderton and Carter (2015), and Uzonyi (2018) report coefficient estimates that are generally negative and significant. Interestingly, Easterly, Gatti, and Kurlat (2006) find that intermediate levels of GDP per capita are associated with greater mass atrocity risk. But when restricting their sample to twentieth century cases, they find a linear effect in which greater GDP per capita correlates to lower risk. They conclude that mass atrocities in the nineteenth century were more likely to occur in relatively rich countries; in the twentieth century the risk was greater in poorer countries. Scully (1997), meanwhile, interprets the significant negative relationship between real


\textsuperscript{50} Positive and significant: Hazlett (2011), Brehm (2017a); insignificant: Wayman and Tago (2010), Krcmaric (2018).

\textsuperscript{51} Positive and significant: Quinn (2015); mixed: Valentino, Huth, and Balch-Lindsay (2004).

GDP per capita and democide as evidence that atrocity perpetrators obey the law of demand: The larger is real GDP per capita, the greater the opportunity cost or “price of the people being killed” (Scully, 1997, p. 81).

Still regarding economic conditions, several studies consider the effect of natural resources on mass atrocity risk. Consistent with their theoretical model, Esteban, Morelli, and Rohner (2015) find that oil production as a fraction of GDP and resource rents from energy, minerals, and timber flows increase risk. Querido (2009) finds that some measures of natural resources (concentrations of diamonds, onshore oil production, and opium production) each increase risk, but Brehm (2017a) finds that the presence of diamond mines and oil production do not significantly affect such risk. Other resource measures used in risk studies do not find significant effects, including primary commodity exports as a share of GDP (Montalvo and Reynal-Querol, 2008) and rebels’ access to natural resources (Kim, 2010).

Other economic variables do not generally show significant effects, including economic crisis (Brehm, 2017a), income inequality (Kim, 2010), and economic growth or internet use per capita (Anderton and Carter, 2015). Effects of various trade measures (trade openness, trade as a percent of world trade, etc.) on mass atrocity risk are mixed. Some studies find trade correlates to significantly lower risk, but others find mixed or insignificant results.53

In a word, of the four variable rubrics studies consider in large-n cross-country studies of mass atrocity (1,000+ fatalities) onset or occurrence, the only strongly consistent empirical result regards political upheaval and/or threat.

4.1.2 Large-n studies of risks of “low-level” violence against civilians

Mass atrocities rarely begin with killing 1,000 or more people; they usually begin with smaller-scale violence against civilians (VAC) by which the literature means attacks in which estimated fatalities are below (and generally well below) the 1,000+ level of fatalities often associated with mass atrocities. As such, VAC risk studies may be helpful for empirical inquiry into how atrocities may scale up to cases that are “massive,” although to date such empirical inquiry is generally lacking (but see Anderton and Ryan, 2016; also see Section 3.3 for theoretical work that may bridge this span.) Another difference between large-n high-level and large-n “low-level” atrocity risk studies concerns perpetrating actors. The high-level studies generally focus on governments as perpetrators. The VAC studies also consider governments (Tir and Jasinski, 2008; DeMeritt, 2012; Machain and Rosenberg, 2018), but some study civilian attacks by rebel groups (Pospieszna and DeRouen, 2017), militias (Stanton, 2015), or combinations of state and nonstate actors (Eck and Hultman, 2007; Hicks, et al., 2011b; Anderton and Ryan, 2016).

Despite a wide diversity in empirical materials such as dependent variable data sources, independent variables, perpetrating actors, and so on, most VAC empirical studies find that measures of threat or upheaval usually correlate to a significantly greater atrocity risk by state and/or nonstate actors. In contrast, empirical results for political regime are mixed. A few VAC risk studies include a measure for social fragmentation. For example, Tir and Jasinski (2008) find that minority groups that face political discrimination are more likely to be targeted by governments, all else equal. Machain and Rosenberg (2018) report mixed results on the effects of political discrimination on the risk of minority group rebellion and subsequent risk of VAC. Meanwhile, Anderton and Ryan (2016) report that their coefficient estimates for ELF are

---


55 Some find no significant effect for democracy (Eck and Hultman, 2007; Tir and Jasinski; 2008; Pospieszna and DeRouen, 2017); one reports a significant negative effect for democracy (Machain and Rosenberg, 2018); one finds that anocracies are at greater risk for VAC events than either democracies or autocracies (DeMeritt, 2012).
generally positive with about half being significant. Turning to *economic variables*, Tir and Jasinski (2008) find that greater economic growth and higher GDP per capita each significantly reduces VAC risk by governments. Machain and Rosenberg (2018), however, find that greater economic growth significantly increases the risk that minority groups will rebel against the state and that the state will violently repress groups. They interpret their results as consistent with a diversionary theory of attacks by states against internal (as distinct from external) enemies (see also Tir and Jasinski, 2008). Specifically, when economic conditions are poor (low economic growth), minority groups know that they are more likely to be targets of diversionary attacks by the government and so they are less likely to rebel and thus less likely to be violently attacked. Regarding natural resources, Pospieszna and DeRouen (2017) find that rebel access to contraband financing (illegal drug trafficking) and the presence of gemstones in the conflict zone significantly increase the risk of rebel-perpetrated atrocities. Finally, we note that foreign aid receipts (DeMeritt, 2012) and trade openness (Eck and Hultman, 2007; DeMeritt, 2015a) did not have significant impacts on VAC risk.

An important additional aspect of the VAC literature is research that empirically assesses third-party efforts to stem atrocity. For example, Pospieszna and DeRouen (2017) hypothesize that third-party *mediation* during civil wars can be perceived as threatening to rebels, making rebels more likely to resort to VAC. After empirically modeling the likelihood that mediators will show up in civil wars, they find significant support for their hypothesis. DeMeritt (2012) considers “*naming and shaming*” actions of human rights organizations (HROs) and of the UN Human Rights Council (UNHRC) and finds that for both naming and shaming significantly decreases the risk that governments will initiate violence. Fjelde, Hultman, and Nilsson (2018) focus on effects of UN *peacekeeping troops* on VAC risk in various locales in Africa.
Controlling for the likelihood of peacekeepers being deployed, they find that peacekeepers significantly reduce the risk of civilian targeting by rebels, but not by governments. The authors maintain that governments are better able than rebel groups to resist the aims of the peacekeepers in regard to civilian protection.

In sum, as for the high-fatality (1,000+) studies, the atrocity-increasing risk effect of political upheaval and/or threat emerges as a consistent finding. Other results are mixed and/or too few in number to venture a statement.

4.1.3 Atrocity seriousness

Atrocity seriousness studies measure the dependent variable based on the magnitude of atrocity as distinct from onset or occurrence. As such, they employ count or categorical estimation methods such as negative binomial, Poisson, ordered logit or probit, or Tobit, although a few use OLS, GMM, or LSDV.

Most of these studies find, again, that measures of threat or upheaval generally correlate to significantly greater civilian atrocities perpetrated by state and/or nonstate actors.\(^{56}\) Regarding political regimes, few general patterns emerge. Some find that democracy has a significant negative effect on atrocity seriousness of state and/or nonstate actors, but it is insignificant in others.\(^{57}\) Koren and Bagozzi (2017) find that anocracies are associated with the greatest seriousness by state and nonstate (rebels and militias) actors, but Eck and Hultman (2007) find the opposite. Relatedly, some find that the number of civilians killed by rebels is significantly


greater under nonautocratic regimes (anocracies and/or democracies) relative to autocracies.\footnote{Eck and Hultman (2007), Wood (2010), Wood, Kathman, and Gent (2012), and Hultman (2012).}

Interestingly, Jo and Simmons (2016, 2017) find that strength of a domestic statute dealing with international crimes and the strength of a country’s rule of law each exert a significant negative effect on the number of civilians killed by governments and rebels.

Empirical results for social fragmentation measures also are disparate. For example, Wood (2014b) finds that ELF significantly increases the magnitude of rebel VAC, and Anderton and Ryan (2016) report similar results for governments, rebels, and militias. Others find that ELF is not significant or has a significant negative impact on atrocity seriousness.\footnote{Azam and Hoeffler (2002), Krain (2005), Easterly, Gatti, and Kurlat (2006), Querido (2009), Quinn (2015).} Fjelde and Hultman (2014) and Ottmann (2015), however, employ finer-grained measures of ethnicity. Fjelde and Hultman’s (2014) dependent variable is the number of civilians killed by governments or rebels in Africa per spatial/temporal grid, finding that governments kill significantly more civilians in grids in which civilians are ethnically aligned with rebel groups, and vice versa. Ottmann (2015) generally finds that greater ethnic fractionalization and greater ethnic polarization within rebel groups each significantly increases the magnitude of rebel VAC.

Turning to economic variables, almost all of the studies that include real GDP per capita find that it usually has a significant negative effect on atrocity seriousness of governments and/or nonstate actors.\footnote{Besançon (2005), Querido (2009), Wood (2010), Kathman and Wood (2011), Hultman (2012), Wood, Kathman, and Gent (2012), Cohen and Nordås (2015), Quinn (2015), Koren and Bagozzi (2017), Fisk (2018); see Azam and Hoeffler (2002) for an exception.} Easterly, Gatti, and Kurlat (2006) find that intermediate levels of income are associated with significantly greater magnitudes of civilian killing. Based on disaggregated data of geographic cells experiencing armed conflict in Africa, Fjelde and Hultman (2014) find that atrocity intensity is significantly lower in high income cells. Among studies that include a natural resource measure, a majority find that it exerts a positive and significant effect, although a few
report mixed results or no significant effect.\textsuperscript{61} Several studies find that greater external material support available to rebel groups usually significantly increases rebel killing of civilians.\textsuperscript{62} Most others find that foreign aid (as distinct from targeted material support for rebels) does not have a significant effect on the seriousness of government atrocities,\textsuperscript{63} but see Azam and Hoeffler (2002) for an exception. Among the few studies that include a trade measure, some find no significant effect on atrocity seriousness, but others find that almost all of their coefficient estimates for the effects of trade are negative and some are significant.\textsuperscript{64}

Regarding third-party interventions into ongoing atrocities, several studies ask if intervention can reduce seriousness and, if so, what types of interventions work. First, Hultman (2010) considers the effect of UN peacekeeping operations (PKO) on the number civilians killed by governments, rebels, and in total in armed conflicts, 1989–2006, and finds that their mere presence does not reduce the magnitude of government atrocities. Disconcertingly, she finds that PKO presence significantly increases the seriousness of civilian killing by rebels. One moral-hazard type of explanation for this result is that rebels use VAC to establish a stronger bargaining position vis-à-vis governments in the presence of PKOs. Another is that PKOs can spur factions to split off from the main rebel group which in turn leads to more VAC (Hultman, 2010, p. 39). Hultman does find, however, that when PKOs, in addition to presence, have a mandate to protect civilians, they exert a significant negative effect on atrocity seriousness. Hultman, Kathman, and Shannon (2013) and Kathman and Wood (2016) evaluate the effects of the composition of PKOs (the number of UN troops, police, and observers) on VAC, with the former focusing on periods

\begin{footnotesize}
\end{footnotesize}
of ongoing civil war and the latter on postconflict environments. Both studies account for endogeneity, that is, the propensity of the UN to select some mission areas over others, and both find that the number of UN *troops* is significantly correlated with fewer civilian killings in total, by governments, and by rebels (and also by militias in Kathman and Wood, 2016). Regarding the number of *police*, Hultman, Kathman, and Shannon (2013) find a negative and significant effect, while Kathman and Wood’s (2016) results are mixed. Importantly, both studies find that the number of UN *observers* correlates to significantly more civilian killing, all else equal. The authors of both studies indicate that timid UN efforts are insufficient to deter civilian killing and may even signal actors that the UN is not (yet) serious about protecting civilians.

Second, Kathman and Wood (2011), Krain (2005, 2014, 2017), and DeMeritt (2015a) consider the effects of third-party *military interventions* on the seriousness of civilian atrocities. The studies by Kathman and Wood and Krain use the same datasets to code genocide seriousness (Political Instability Task Force Geno-politicide dataset or PITF-G) and military interventions (International Military Interventions dataset), similar key intervention measures (pro-perpetrator, anti-target, anti-perpetrator, pro-target, and impartial interventions), and the same estimation method (ordered logit) but for different sample periods. One crucial difference is that Kathman and Wood (2011) record the number of years that interventions are in place as distinct from the number of interventions lagged one year. Distinguishing the intervention types in the short- and long term, they find that impartial and anti-perpetrator military interventions significantly worsen civilian killing in the short term. Longer term, enduring anti-perpetrator military interventions significantly worsen genocide seriousness; likewise, longer term impartial intervention has a significant negative effect. Results reported in Krain’s three studies are decidedly different, finding that only anti-perpetrator and pro-target interventions generally significantly reduce
genocide seriousness. Similarly, DeMeritt (2015a) finds that military interventions against (in favor of) the state significantly decreases (increases) the magnitude of civilian deaths, all else equal.

Third, Krain (2012, 2014, 2017) and DeMeritt (2012) evaluate the effects of naming and shaming efforts on the magnitude of government atrocities. Krain (2012) takes a rational expectations perspective, namely, that government perpetrators would expect some level of condemnation from the international community for their actions (as formally modeled in DeMeritt, 2015b and Esteban, Morelli, and Rohner, 2015, 2016). Nevertheless, he writes, “perpetrators may not have accurately assessed the level of scrutiny and condemnation that their actions would incur” (Krain, 2012, p. 3). In his three articles, Krain tests the efficacy of naming and shaming on the degree of fatalities in genocides, 1976–2008. Krain (2012) considers several measures including media reports, Amnesty International (AI) background reports, and shaming efforts by the United Nations Human Rights Council (HRC), while Krain (2014, 2017) uses the number of AI reports only. Krain (2012, 2014, 2017) generally finds that naming and shaming efforts significantly reduce genocide severity. DeMeritt (2012), in turn, considers both very low-level events (with as “few” as five civilian fatalities) as well as larger-scale killings. She considers three naming and shaming measure per country year: (1) number of efforts by 432 Human Rights Organizations (HROs), (2) average number of reports by The Economist and Newsweek, and (3) severity of HRC punishment (also in Krain’s work). Taken one at a time across three regressions, she finds that the coefficient estimate for each measure is negative, and all but media shaming are significant. Both Krain and DeMeritt conclude that shaming can alter the behavior of murderous regimes and save lives during genocides and episodes of lower-level
of violence against civilians, respectively. This is important policy-wise because naming and shaming is noncoercive—and rather inexpensive.

Fourth, to the best of our knowledge, Krain (2014, 2017) represents the only published articles on the effects of diplomatic and economic sanctions on genocide seriousness. Including controls for military interventions and naming and shaming efforts, he effectively pits three classes of interventions against one another. His “regression competition” finds that neither diplomatic nor economic sanctions significantly reduce the seriousness of ongoing genocides. Only anti-perpetrator/pro-target military interventions and naming and shaming have salutary effects.

In sum, threat and upheaval once more emerges as a fairly clear risk factor for atrocity seriousness, and decent economic conditions may perhaps mitigate this risk. Among types of helpful third-party interventions we see somewhat consistent results emerge for the use of UN troops and for naming and shaming efforts.

4.1.4 Forecasting


A common method of calibrating the predictive accuracy of forecasting models is to use the Receiver-Operating-Characteristic (ROC) and associated area under the curve (AUC) measure. In our context, AUC is the probability that a randomly selected positive case (a country-year
with an onset of mass atrocity) will have a higher predicted probability than a randomly selected negative case (a country-year without onset) (Ulfelder, 2013, p. 12). AUC ranges from 0.5 to one, with higher values representing better predictive power. In our view, the AUC scores reported in the forecasting literature are generally good. For example, based upon logit methods, Hazlett (2011) reports AUC scores of 0.90 and 0.93 for two in-sample genocide forecasting models. Goldsmith, et al.’s (2013) out-of-sample genocide forecasts achieve a score of 0.90. Similarly, Rost’s (2013) in-sample models for forecasting genocide and genocide conditional on civil war achieved mean scores of 0.90 and 0.86, respectively. Koren’s (2015) out-of-sample forecasts for mass atrocity conditional on internal threats achieve a mean score of 0.86 across multiple runs.

Other models use unweighted, weighted, and Bayesian combinations of models (multimodel ensembles), which include standard logit (or rare events logit), models conditional and not conditional on civil wars or other serious upheavals, K-nearest neighbor (KNN) discriminant analysis, machine-learning methods, and random forest models (Ulfelder, 2012, 2013, 2014). Ulfelder’s multimodel methods lead to AUC scores of 0.75 or greater. Analyses of the Atrocity Forecasting Project’s (AFP) (2018) multimodel methods of genocide forecasting are also encouraging. Goldsmith, et al. (2013) report that the AFP’s forecasts achieved scores for out-of-sample forecasting of 0.89 for 1988–2013 and 0.92 for 1990–2010. Goldsmith and Butcher (2018) report that AFP achieves scores of 0.86 for UN warnings, 0.81 for Genocide Watch onsets, and 0.96 for PITF-G onsets for the 2011–15 period.

Few economic variables seem to improve forecasting results. For example, only the infant mortality rate (which Harff treated as a proxy of economic development) “survives the cut” in Goldsmith and Butcher’s (2018) genocide forecasting model. The only economic variable
retained in Rost’s (2013) model is real GDP per capita (performing better than the infant mortality rate). When Rost turned to forecasting genocide conditional on civil war, the only economic variable that improved predictive power was the degree of economic discrimination. In Ulfelder (2012), the only economic variable that improves mass atrocity forecasting is real GDP per capita. Hazlett (2011), however, retained two economic variables—trade openness and iron and steel production—in his genocide forecasting.

Butcher and Goldsmith (2016) provide two explanations for the nonpredictive performances of most economic variables in genocide forecasting, and we believe their insights would also apply to mass atrocity forecasting. First, most economic measures widely available in datasets such as GDP per capita, trade, natural resource stocks or production, income inequality, and even commodity prices change relatively slowly over time. Even in cases of sudden economic crisis, more proximate drivers may be at play such as political upheaval or emerging severe discrimination against out-groups. Whereas economic variables can significantly affect the risk or represent causal elements in mass atrocity over long time periods, near-term prediction is improved by relying on measures of more immediate political and social conditions. Second, some economic variables potentially useful for atrocity forecasting have not been tried. These include horizontal inequality (which may be partially captured in measures of economic discrimination), dramatic economic decline over five or ten year periods, sudden changes in the financial markets of at-risk countries (for example, a sudden fall in bond ratings or the sudden selling of government bonds to recruit troops), and nonlinear and nonmonotonic relationships between economic variables and atrocity (Butcher and Goldsmith, 2016, pp. 584–5). We also note that trade measures widely used in risk factor studies of wars like total trade or trade openness have flaws that network centrality measures of trade can improve upon (Kinne, 2012).
To date, measures of trade networks of states have not been incorporated in mass atrocity risk or forecasting studies. Finally, the aggregation of economic activities inherent in broad measures (for instance, GDP per capita, trade openness, society-wide inequality) and in large-\textit{n} cross-country studies may “hide” shocks to particular sectors critical to a country such as the aforementioned coffee, tea, and tin economies prior to the 1994 Rwandan genocide.

4.1.5 \textit{Micro-level evidence in atrocity case studies}

peasants”) and those with large landholdings (“landlords”) can “be found in their respective relation to the land and labour markets” (p. 319). His tests provide evidence that Hutu landlords had something to defend (their privileged position in society) and the quasi-landless something to gain (land of murdered Tutsi) from genocide perpetration. Rezaedaryakenari, Landis, and Thies (2017) find that rising food prices across administrative districts in sub-Saharan Africa significantly increase the risk and seriousness of atrocities by rebel groups. They also find that cultivated regions in sub-Saharan Africa are more likely to be attacked by rebels during periods of food insecurity. Meanwhile, Brehm (2017b) finds that communes in Rwanda with higher levels of formal sector employment had significantly fewer killings during the genocide, all else equal.

Some of these studies recognize social networks, even if not explicitly modeled using formal network theory. For example, McDoom (2014a, 2014b) draws on various concepts from network economics to guide his empirical inquiries into genocide acceptance and propagation in Rwanda in 1994. McDoom (2014a) distinguishes between extremist Hutu, who favored genocide, and moderate Hutu who were opposed. Extremist and moderate Hutu underwent an intragroup competition for elite control in many areas in Rwanda once the genocide began. Based on 1994 data for a sample of 145 communes in Rwanda, he finds that communes controlled politically and/or militarily by extremists succumbed relatively soon to anti-Tutsi violence, all else equal. He also finds that communes with high segregation experienced anti-Tutsi violence sooner relative to better-integrated communes.

McDoom (2014b) analyzes factors that drew individuals from the Tare sector of Butare prefecture in southeastern Rwanda into perpetration during the 1994 genocide. Theorizing that social networks can foster or dampen individual participation in genocide based upon
mechanisms operating among connected individuals, his sample encompasses survey responses of 116 males 14 years old and older from Tare’s 647 households who were present in Tare in April 1994. He finds that an individual was significantly more likely to participate in genocide the larger his social network and the larger the number of his connections to participants. Among many types of ties, McDoom finds that kinship and spatial connections significantly affect participation risk. Importantly, McDoom (2014b) does not focus on personal characteristics of individuals, but on “the importance of social structure and social interaction for [understanding individual] participation in collective violence” (McDoom, 2014b, p. 865; our emphases) (also see McDoom, 2013).

Enablers of social networking include communications and transportation networks, which are emerging in empirical micro studies of atrocity. For the Rwandan genocide, Yanagizawa-Drott (2014) studies the potential role of hate radio (Radio Télévision Libre des Mille Collines, RTLM) as a coordination device to incite violence (also see McDoom, 2012). Hate radio can have two effects, one direct (prompting listener participation in mass atrocities), the other indirect (prompting nonlisteners to participate as well). To arrive at statistically credible results, Yanagizawa-Drott exploits the quasi-random distribution of hills, flatlands, and valleys that affects the quality of radio signal reception. Holding other factors constant, communities in reach of a clear line of signal showed a significant (and quantitatively large) increase in genocide participation. So did neighboring communities, due to spatial spillovers on social networks, suggesting triggering and coordination effects facilitated by hate radio, even to the extent that the aggregate spillover effects exceed the direct effects. Yanagizawa-Drott (2014) estimates that about 10 percent of the participation in the violence is attributable to RTLM, translating to an
additional 51,000 perpetrators. Adena, et al. (2015) use a similar Irregular Terrain Model (ITM) design to study the effect of radio in Germany before and after Hitler’s ascent to the chancellorship. Not only do they confirm the qualitative findings in Yanagizawa-Drott (2014), but they are able to establish that the reinforcing effect of radio propaganda depends on predispositions to accept hate messages and that pro-hate propaganda can backfire in communities with higher dispositions of tolerance toward out-group members. Such studies begin to link theoretical models of identity, networks, and hate to empirical applications.

Relatedly, a recent study documents how Israeli attacks in the Palestinian Occupied Territories appear causally related to time periods when U.S. media divert attention to large-scale sporting events or natural catastrophes (Durante and Zhuravskaya, 2018).

Micro empirical studies of atrocity also are incorporating the influence of transportation networks on atrocity risk and seriousness (Verpoorten, 2012; Rogall, 2015; Zhukov, 2016, 2017; Brehm 2017b). For example, as noted previously, for the period 1945–59 Zhukov (2016) geocodes 2.65 million arrest records for Stalin’s population control campaigns and finds that Stalin was less willing to force remotely located civilians to move (including to Gulag camps) the further away they lived from a railway station and the greater the distance to a camp.

Studying atrocities committed by Nazi forces in German-occupied Belarus (1941–5) (specifically, the burning of 8,526 villages and the killing of hundreds of thousands of civilians), Zhukov (2017) finds that when resistance groups attacked German military targets (personnel and buildings), indiscriminate retaliations against civilians were swift and severe. But when the opposition attacked German transport networks, particularly railroads, retaliations against civilians were much lower. Whether intending to or not, by targeting transport the opposition

---

66 See also Straus (2007b), who finds little evidence that RTLM had widespread direct effects on participation during the Rwandan genocide. He does report, however, that some key agents were catalyzed by RTLM broadcasts into genocide participation in some locations.
created a demand for the use of local labor to repair the damage, thus reducing killings. As one German official in Belarus stated: “We needed that population to clear the railroad tracks” (quoted in Zhukov, 2017, p. 68).

4.2 Summative evaluation and key issues for future empirical work

Thriving streams of empirical mass atrocity research have emerged. Measures of threat or upheaval generally correlate strongly to atrocities in both analytical and forecasting studies. Other variable measures such as political regime level (not transition), ethnic diversity and polarization, trade openness, and natural resources show mixed results. For the large-\(n\) studies, low real GDP per capita and high economic discrimination seem to be the only economic variables that correlate to greater atrocity risk or seriousness, but the results are not as consistent as they are for threat/upheaval. Moreover, economic discrimination (as distinct from general economic and political discrimination and other measures of social fragmentation) has been tested in very few large-\(n\) empirical atrocity studies. Only in the micro-level case studies do we see economic conditions, particularly those related to land and agricultural and to communications, transportation, and social networks, as statistically significant factors in explaining atrocity risk and seriousness. Regarding third-party efforts, only three clear patterns have emerged. First, UN commitment of soldiers into ongoing atrocities significantly reduces atrocity seriousness; commitment of (mere) observers significantly increases seriousness. Second, most naming and shaming efforts by organizations significantly reduce atrocity seriousness. Third, neither diplomatic nor economic sanctions significantly reduce the seriousness of ongoing genocides. Empirical results on various types of military interventions into ongoing genocides are mixed. Some types of military interventions seem to worsen atrocity
seriousness, and “norm interventions” by the international community have the potential to backfire (Esteban, Morelli, and Rohner, 2015).

By way of reflection, we highlight five challenges as empirical research progresses. First, the literature has led to few cumulative results and has made little headway in sorting out competing theories and predictions. For example, most empirical studies include a linear or loglinear measure of political regime level; very few consider nonlinearities. Only two risk studies of which we are aware consider political regime transition, as distinct from level (Fein, 1995; Esteban, Morelli, and Rohner, 2015); only one considers the differences between the nineteenth and twentieth centuries (Easterly, Gatti, and Kurlat, 2006; but see Taylor, Pevehouse, and Straus, 2017, on low-level VAC and elections in sub-Saharan Africa.) Insufficient theorizing and empirical analysis also concerns the potential effect of trade openness on atrocity risk or seriousness. Inclusion of the variable is usually predicated on the untested idea that atrocity presumably leads to trade disruption which, in turn, is costly economically or politically. The interstate and civil war literatures long ago tested the proposition that war disrupts trade (Bayer and Rupert, 2004; Glick and Taylor, 2010). But in the atrocity literature, we do not know whether atrocities (controlling for other forms of violence such as wars) disrupt trade. Finally, we are surprised by how small is the empirical literature on third-party intervention. We need to learn what interventions work and why, especially regarding noncoercive and coercive (military) interventions.

Second, few empirical studies are guided by formal theoretical models.\footnote{Exceptions include Azam and Hoeffler (2002), DeMeritt (2015a, 2015b); Esteban, Morelli, and Rohner (2015); Anderton and Ryan (2016); Gangopadhyay (2016); Vargas (2016).} Formal models help distinguish between competing accounts of phenomena and guide in selecting the best measures of theoretical variables. For example, Esteban, Morelli, and Rohner’s (2015) theoretical model of
genocide risk requires that they incorporate a measure of political regime transition (as distinct from level) and oil production or flow (as distinct from natural resource abundance or stocks often found in other studies). DeMeritt’s (2015a, 2015b) principal-agent model of government delegation of atrocity actions to perpetrator “troops finds both theoretically and empirically that the decision of a government to kill civilians and, if to kill, how many civilians to target depends on the type of military intervention directed to the government (either against or in favor of the government) and on the nature of the government-perpetrator (principal-agent) interaction (see also DeMeritt, 2012; Krain, 2012; Machain and Rosenberg, 2018). The mass atrocity empirical literature could stand refinement of measures of key explanatory and control variables as guided by theory and, from this development, push toward discerning which theoretical perspectives hold up well and finding out whether any cumulative results emerge.

Third, even with current progress in empirical research, we note striking gaps. For example, to our knowledge, no forecasting studies address atrocities committed by nonstate groups, even as they seem to be as frequent as those perpetrated by states in datasets in which the two actor classes overlap (Section 2). We also see relatively few empirical studies of the transition from relatively low to high levels of atrocity (exceptions include DeMeritt, 2015a; Quinn, 2015; Anderton and Ryan, 2016). Another concern regards the strong focus on perpetrators (states and nonstate actors) as distinct from studies of actions of bystanders and victims (exceptions include studies of third-party interventions). Brauer and Caruso (2016) provide theoretical guidance that could aid such empirical inquiry. Further, we note that large-\( n \) cross section studies focus upon countries, regions, or locales in which the dependent variable (atrocity onset or occurrence) is 1 or 0, but (to our knowledge) studies do not consider why some groups are targeted (1) and others not (0) within and across countries. Finally, also lacking are theoretical and empirical inquiries
into the connections between structural conditions that elevate the risk of atrocities and catalytic events or “triggers” that can tip a risky situation into atrocity onset. Straus (2015) provides guidance on how triggers might be defined and better integrated into future empirical work.

Fourth, it is often assumed, implicitly (and occasionally explicitly), that genocides and other mass atrocities occur during or in the late stage or immediate aftermath of wars (civil, interstate, and colonial). This is reflected in many empirical models by the presence of a lagged measure of civil (and/or other) war. Certainly, many mass atrocities occur after a war has started, but a large number of exceptions exist as documented earlier in Section 2.4. Why wars and mass atrocities are sometimes interdependent (but with varied timings of interdependence) and sometimes not has not been seriously considered in formal theoretical models and empirical inquiry. This is true of both the mass atrocity and the civil war literatures.68

Fifth, we conclude this subsection by summarizing Douglass’ (2016) work. Important in its own right, it also points to critical data issues. He studies the U.S. Phoenix Program, 1968–72, run during the Vietnam War. This included a targeting database of 73,712 civilians suspected of supporting North Vietnamese forces, 15,438 of whom were eventually killed. Applying machine learning to the data, patterns of missing values, meanings of different variables, and heterogeneity of observations can be questioned.69 Douglass finds numerous problems with the data that needed to be dealt with prior to empirical analysis. He concludes that such detail

---

68 We make two additional observations. First, scholars date the onset of the prototypical mass atrocity of the twentieth century—the Holocaust—to 1933, whereas the mass killing began with the onset of World War II in 1939. More broadly, case study research of twentieth century genocides by Bartrop (2002) leads him to conclude that “genocide does not equate with war, and the two should not be considered synonymous; in some cases there is a relationship of interdependence between the two, but this is by no means a given in every situation; [and] military conflict does not have to be present for a genocide to occur” (p. 531). Second, to our knowledge, Schneider, Bussmann, and Ruhe’s (2012) micro-study of civilian violence in the Bosnian civil war is the only article that theoretical justifies and empirically supports a hypothesis that perpetrators commit fewer atrocity acts during or in the immediate aftermath of territorial conquests (see their H3, p. 448).
69 Basuchoudhary, et al. (2018) apply machine learning as well, although to prediction of civil conflict rather than of mass atrocities.
“typically get[s] relegated to an error term” (p. 561) and that an analyst “could reach dramatically different conclusions about outcomes by truncating the sample just to killings, by omitting information about the suspect’s position [e.g., rank in opposition forces] or the government actor committing the violence [e.g., which branch of government or foreign forces], or by missing important details about how the [data-capturing] institution[s] created records and aggregated them into a final dataset” (p. 562). As increasing numbers of local datasets become available from mass atrocity research, it may be appropriate to initiate pre-theoretical, machine-driven queries to discover data dimensionality and structure prior to building out theory and empirical testing of hypotheses.

5. Toward Prevention: Resetting Context, Reconditioning Choices, Reshaping Consequences

The literatures we have surveyed in many ways echo Scott Straus’ caution in his review of atrocity prevention tools and strategies: “… atrocity prevention is difficult” (Straus, 2016, p. viii). A survey and assessment of mass atrocity prevention efforts and tools through an economics lens remains to be written. That said, in the preceding pages, we indicated in many places how a view informed by economic theory can provide new insights into understanding and, possibly, preventing mass atrocities, for instance, the relative ease of (yet limits to) atrocity input substitution and of intervention backfiring conditions (from standard theory), the role of motivated beliefs and (self)-narratives (behavioral economics), of identity manipulation, preference falsification, and atrocity up-scaling (identity and social network economics), and of

---

the institutional design of atrocity games (political economy). To illustrate with just one more example, Esteban, Morelli, and Rohner (2015, 2016) highlight an underappreciated rationale for killing people for economic reasons. If an in-group views people of an out-group as chattel—an economic resource to be exploited—then imposing conditions that reduce this economic value to the in-group can backfire and lead to quicker and harsher atrocities committed against the out-group.

In what follows, we indicate briefly what we believe are two novel economics-informed avenues regarding atrocity prevention, one cautionary, one perhaps more hopeful. First, we expand on the theme that democracy is neither a necessary nor sufficient antidote to prevent atrocities (Section 5.1). Second, lack of “political will” to prevent mass atrocities is often bemoaned (Straus, 2016) yet concepts involving the design of (global) public goods seem readily applicable to explain this lack of will. One of the global goods in question is international law. By way of illustration, we discuss the economics of the design of international law (5.2).

5.1 Democracy, institutions, and law

Democracy, to first stay within the nation-state realm, is frequently mentioned as an atrocity prevention tool. A series of papers by Acemoglu, Egorov, and Sonin (2015; with the other papers cited there) explore democracy and nondemocracy as political institutions. For instance, their 2012 paper generates two insights: First, that a Pareto inefficient “social arrangement is made stable by the instability of alternative arrangements” and, second that “efficiency-enhancing changes may be resisted because of further changes they will engender” (p. 1446). The fear of future instability—even if post-instability efficiency were enhanced—can keep an inefficient system of governance stable. What counts is regime stability, not its form as democratic or nondemocratic. “Nudging” nondemocratic regimes in the democratic direction can backfire.
Threats of transition create “openings in the political opportunity structure” (Krain, 1997). Once achieved, democracy may well be desirable, but the path to get there can be bloody indeed.

In this regard, Myerson (2011, 2016) argues that the United States was mistaken in postwar Iraq to impose, by means of constitutional law, a centralized regime on the country, a nudge in the wrong direction as it de-incentivized the local and regional proving grounds of future national-level leaders. Myerson argues that it is at the local level that future leaders first can demonstrate their ability to manage public affairs in constituents’ best interest. If leaders pass the local test, they may in the future be elected to head up a region-wide office and, in time, perhaps advance to the national level. If, in contrast, the constitutional setup favors the national level, then—to ensure their continuing hold on national power—potential leaders will want to create regional and local vassals. Intermediate-level holders of power, instead of being responsible to local voters, become beholden to the national leader. Political theorist Stephen Krasner (2011) makes a related point, arguing that intellectual preoccupations in international relations theory have forgotten the extent to which external powers can influence a state’s internal power arrangements, possibly the most drastic example of which is the writing of a new constitution for a defeated or an occupied state as in Myerson’s concern over postwar Iraq.

Taking an economic look at comparative constitutional law, post-mass atrocity, may be a useful exercise (for example, for such wide-ranging cases such as Colombia, East Timor, Germany, Iraq, Japan, and Nepal). Constitutions are, fundamentally, about setting, or changing, the rules of the political game to create preferred political structures. Verwimp’s (2004a) study of nested games in Rwanda in the early 1990s is a powerful example of this: When the structure of the existing transition and civil war games proved inconvenient for then-dictator Habyarimana, the games first were collapsed into a single game and, eventually, into no game at
all with the elimination of the Tutsi population. In the world of business, modern management theory boils this down to the dictum of “strategy before structure.”

5.2 Global public goods and the economics of international law

If “strategy before structure” is one way to think about how actors can shackle (or unshackle) themselves to forestall (or permit) mass atrocities, then we enter the realm of law, broadly conceived, and of the economics of law. This includes the literatures on global public goods in general and the design of international human rights or other treaties in particular (Arce, 2000; Sandler, 1997; Sandler, 2004; Peinhardt and Sandler, 2015). For instance, the “too little, too late” phenomenon routinely observed at United Nations Security Council deliberations regarding sanctions and/or interventions in acute cases of mass atrocity can be explained, in part, by the difficulty that would-be intervening states experience to capture benefits commensurate with their budgetary and nonbudgetary (for example, reputational) costs. For intervening countries, these costs are highly specific and local to their political decisionmakers, taxpayers, and voters, yet any benefits obtained can be hard to pin down. In this regard, Richard Posner’s *Economic Analysis of Law* textbook (2014) contains an illuminating, if slim, chapter on Comparative and International Law, with even slimmer sections on the economic analysis of international treaties and domestic law such as the U.S. Alien Tort Statute of 1789 which may be relevant to addressing human rights violations occurring elsewhere. Somewhat expanded coverage, with sections on the use of force, the conduct of war, human rights, and international criminal law, is offered in *Economic Foundations of International Law* (E. Posner and Sykes, 2013). An economics and law overview with applications drawn by analogy from property, contract, tort, and criminal law, and from international law to mass atrocities is in Brauer, Anderton, and Schap (2016). The crux of the matter revolves around the costs and benefits of negotiating binding and
enforceable international treaties in the absence of supranational authorities. R. Posner notes, for instance, that while negotiating a multilateral treaty is less costly than negotiating a large number of bilateral treaties, these costs still are substantial and among the reasons that relevant treaties often are single-issue treaties with relatively low performance thresholds. To bargain for a larger set of benefits, the costs might exceed what nation-states might be willing to commit in present and future expected costs.

R. Posner also notes that the absence of supranational authorities need not be a problem if treaties (a form of contract) are self-enforcing, for example via reputation effects or when current or future benefits are alienable (revocable). Thus, the European Union for instance already links accession by additional states to “a certain level of adherence to human rights norms and law” (Brauer, Anderton, and Schap, 2016, p. 657), a big bone of contention in the on-and-off talks between the EU and Turkey for example. Moreover, the EU challenges and threatens punishment when existing members violate norms, as it did in 2015 in regard to Hungary’s immigration policy regarding refugees from the Syrian mass atrocity or in late 2017 after Poland issued a reform law threatening the independence of its judiciary to the dislike of the other 27 EU members who then issued an ultimatum on 20 December 2017, under Article 7 of the 1999 Treaty of Amsterdam, to compel Poland to rescind, or sufficiently revise, the reform law or else face sanctions such as loss of EU voting rights.

Further, R. Posner hints at why nation-states may be quite reputable in regard to some international treaties yet less so in regard to others. He explains that “[s]ince the costs and benefits of compliance vary across commitments, violation of one may be weak or even no evidence of the likelihood of violating another, where the benefits of compliance may be much greater” (2014, p. 996). In a similar vein, Shaffer (2012) notes that international laws as
instances of global public goods can be rivalrous to multiple national laws and to each other.

Combined with Hirshleifer’s (1983) insight that some (global) public goods are best-shot, weakest-link, or aggregate effort goods or services and the insight that atrocity “bads” come in various forms (genocide, for instance, is a “club bad” whereas a random shooting of some among an assembled group of villagers is a “common-resource pool bad” and poison gassing is an unmitigated “public bad”), Shaffer (2012) argues that differently designed international treaties are needed to address different specific situations. For example, if the United States were the sole (super)power able to effectively intervene in a mass atrocity occurring elsewhere, then prevention or mitigation of the harm done would be subject to cost-benefit considerations of the United States alone. The world would be held hostage to the self-interest of a single state. But were the law designed as an instance of global administrative law, one or more actors could be “deputized” on a case by case basis to address an imminent or ongoing atrocity and be held responsible for its (in)actions.71 Similarly, weakest-link global public goods might best be dealt with under legal pluralism rather than universal law. Only aggregate effort global public goods to prevent certain types of mass atrocities might require a global constitutionalist approach, Shaffer argues.

Given the vastness and substantive depth of law and economics scholarship, and given its corresponding paucity of application to the pressing issue of mass atrocity prevention, we suspect that great advances are possible in this arena of scholarship.

6. Conclusion

We conclude with nine observations. First, it seems fair to say that across the tributaries of

---

71 “States … developed administrative law accountability mechanisms to apply to agencies, given that legislatures were unable to oversee them sufficiently. International institutions can be viewed analogously to national government agencies, in that both involve a delegation of power to an unelected body” (Shaffer, 2012, pp. 687–8).
economic theory laid out in this article many high-quality contributions have been made, albeit from angles so dispersed and diverse that a distinct ‘economics of mass atrocities’ had not, as yet, come into focus. Second, empirical work is not always as explicitly and formally grounded in theory as economists should like and few consensus results have emerged. Conversely, much of the work in theory is in need of empirical testing. Our ignorance still is great. Third, both theory and empirics frequently equate mass atrocities with genocides (including politicides), missing the large universe of cases that are not genocides. Moreover, fourth, theoretical and the empirical work predominantly address just the most egregious type of genocidal atrocities—killings—overlooking other types of genocides such as enslavement for economic gain or forced assimilation policies to destroy a people-group’s cultural coherence and identity. The dual focus on genocides and killings is reinforced by databases that in their majority still tend to collate information on “massive mass atrocities” (1,000+ people killed). Fifth, the purpose and design of current databases, then, seems in need of debate. It appears, for example, that many “within-nation” atrocities—from Afghanistan to Colombia and India and from Pakistan to Venezuela and Zimbabwe—for which local-based, local-language data either exist or can be found are not included in the dominant, cross-national, English-language databases empirical researchers tend to access. Valuable as they are, extant studies tend to default to “convenience datasets,” potentially skewing derived knowledge. Difficult as the task will be, new, or newly expanded, databases should cover years before 1900, should cover all atrocity perpetrators (not just modern nation-states), and should cover all sizes, types of, and motives for, mass atrocity.

Sixth, inasmuch as economics makes universal claims about how economic laws hold across time, space, and agents, the preponderance of theoretical and empirical mass atrocity studies devoted to the post-Holocaust era, while important, is lacking in imagination and scope. If, for
instance, social networks are important in today’s cases, surely they were important in, say, medieval times. The explanatory and predictive power of theory lies, in part, in its universal reach. By expanding studies more broadly across time, space, and actors, possible anomalies to theory may be discovered, and corrected, more quickly than otherwise. Seventh, theoretical work has emphasized the nature and role of atrocity perpetrators and much less so of types of bystanders and victims; similarly, the empirical work has emphasized the study of atrocity onset, seriousness, and (less so) of intervention as one type of bystander activity but far less so the study of perpetrators and, again, of victims. Victims are not without agency, especially in the early stages of victimization; we might learn much from studying their behavior. Eighth, it seems to us that many economic specialties can usefully contribute to the study of mass atrocities. As illustrated in Section 5, a ready pool of scholarship exists in the field of law and economics that can be transported to the study of mass atrocity and its prevention. Similarly, the economics of (organized and other) crime surely contains insights and tools of use to the study of mass atrocity. Likewise, knowledge gained in the economics of culture can be grafted onto the study of the looting and the destruction of tangible cultural artifacts not just in wartime but as a way of collapsing a people-group’s self-understanding and self-identification. Further, a large, and often ingenious, literature in experimental economics, including conflict-linked experiments, may well be adaptable to tease out more knowledge regarding mass atrocity-related behavior. Ninth, going beyond personalized theories of behavior, far more research is needed drawing on social context and network models. As Mueller (1990) and others show, much prevention comes in the form of changes in culture-wide attitudes and institutional reforms that help shape individual behavior. Tortures such as quartering, crucifying, and drowning, and generalized violence on account of race, religion, sex, age, or mental capacity all are becoming less thinkable and rarer than in times
past, not because human nature has changed but because institutional and social control mechanisms have changed (Pinker, 2011). As such, studying mass atrocities not merely as instances of an economics of war, violence, and insecurity but studying their prevention as exemplars toward the creation of stable, irreversible structures of peace (Boulding, 1978) seems appropriate.

Perpetrated by constantly changing types and compositions of actors, mass atrocities take on many forms and occur with some frequency and regularity even today, in war- and in peacetime. They cause enormous damage to human and economic development and with adverse effects often counted over generations. The economics profession—with its particular brand of logic, mathematics, and statistics and its broad sweep of theories from which to draw—is well positioned to make further valuable contributions to the study of mass atrocities and, we trust, to their mitigation and prevention.
REFERENCES


Atrocity Forecasting Project. 2018. “Atrocity Forecasting Project.”


Cribb, Robert. 2010. “Political Genocides in Postcolonial Asia.” In *The Oxford Handbook of*


DeMeritt, Jacqueline H. R. 2015b. “Principal-Agent Model of Delegating Death.”


Routledge.


“Female Genital Cutting is Not a Social Coordination Norm.” *Science* 349(6255): 1446–7.


Fein, Helen. 1995. “More Murder in the Middle: Life-Integrity Violations and Democracy in the


Goyal, Sanjeev. 2016. “Networks in Economics: A Perspective on the Literature.” In *The Oxford
Handbook of the Economics of Networks, edited by Yann Bramoullé, Andrea Galeotti, and Brian Rogers, 47–70. New York: Oxford University Press.


Press.


Kinne, Brandon J. 2012. “Multilateral Trade and Militarized Conflict: Centrality, Openness, and


Linstead, Stephen, Garance Maréchal, and Ricky W. Griffin. 2014. “Theorizing and Researching...


Masys, Anthony J., ed. 2014. *Networks and Network Analysis for Defence and Security*. Berlin:
Springer.


Novak, Lindsey. 2016. “Persistent Norms and Tipping Points: Female Genital Cutting in Burkina


Soudis, Dimitrios, Robert Inklaar, and Robbert Maseland. 2016. “The Macroeconomic Toll of


Zech, Steven T., and Michael Gabbay. 2016. “Social Network Analysis in the Study of


Figure 1: Mass Atrocities Perpetrated By States, Nonstate Groups, and Unidentified Actors

- State-perpetrated mass atrocity (SMA)
- Nonstate-perpetrated mass atrocity (NMA)
- Unidentified mass atrocity (UMA)

Note: Mass atrocities (estimated fatalities) plotted at onset year.
Data Sources: Authors’ compilation. See Appendix.
Figure 2: Diffusion Curve for Genocide Contagion

\[ p^t \]

\[ p^{t-1} \]

Source: Adapted from Jackson and Zenou (2015, p. 136).
Figure 3: Networked Tentacles of Atrocity Perpetration

Source: Anderton and Brauer (2018).
Figure 4: Types of Dictator Equilibria in Wintrobe’s Loyalty-Repression Model

<table>
<thead>
<tr>
<th>Loyalty</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Tyrant</td>
<td>Totalitarian</td>
</tr>
<tr>
<td>Low</td>
<td>Timpot</td>
<td>Timocrat</td>
</tr>
</tbody>
</table>

Source: Adapted from Wintrobe (1998, p. 81).
Appendix

Datasets and Protocols To Identify Mass Atrocity Cases and Fatalities per Case, 1990–2017

I. Civilian Atrocity Datasets

The following civilian atrocity datasets and one war dataset were used to (1) create the scatterplot of mass atrocity cases (Figure 1), (2) present mass atrocity fatality estimates, and (3) identify the start and end dates of mass atrocity cases in order to analyze state-perpetrated mass atrocity overlaps with wars.


5. Political Instability Task Force Worldwide Atrocities Dataset (PITF-W) to code “unidentified perpetrator mass atrocity cases” only, which covered the period 1995–2017 (http://eventdata.parusanalytics.com/data.dir/atrocities.html).

6. Armed Conflict and Location Event Dataset (ACLED) to code unidentified perpetrator mass atrocity cases only, which covered Africa for the period 1997–2017, the Middle East for the period 2016–7, and parts of Asia for the period 2010–7 (Armed Conflict Location & Event
Data Project (ACLED) (acleddata.com); Raleigh, Linke, Hegre, and Karlsen, 2010).


II. Coding Protocols for State-Perpetrated Mass Atrocity Cases

1. State-Perpetrated Mass Atrocities, 1900–44

We identified state-perpetrated mass atrocity cases for the 1900–44 period from EGK for cases in which at least 1,000 civilians were reported killed. If EGK reported a range of civilians killed, we estimated civilian deaths as the average of the two numbers. Atrocity cases in EGK in which fewer than 1,000 civilians were killed or no civilian fatalities were reported were excluded from our data analysis. Atrocity cases in EGK that had onset years prior to 1900 were excluded from our analysis even if the atrocities continued into 1900 or later. One case in EGK (China 1920s–49) did not have a point estimate for the onset year and was coded in the scatterplot with an onset year of 1925.

2. State-Perpetrated Mass Atrocities, 1945–54

We identified state-perpetrated mass atrocity cases for the 1945–54 period from UV. All cases in UV involve the intentional killing of at least 1,000 noncombatants from a discrete group (Ulfelder and Valentino, 2008, p. 2). For each case, UV report low and a high fatality estimates. We took the average of the two estimates to arrive at our measure of estimated fatalities.


We identified state-perpetrated mass atrocity cases for the 1955–2006 period from UV and PITF-G. When the two datasets had overlapping cases, we recorded only one case in order to avoid double counting. The case that was chosen was the one that provided the higher fatality estimate. For UV, the fatality estimate was determined by the averaging protocol described in II.2. For
PITF-G, death magnitude indexes were provided for each year of each case. We used the midpoint of PITF-G’s death magnitude index range to determine the fatality estimate for each year. For example, the death magnitude index for Sudan for 2002 was 2.5, indicating that estimated civilian fatalities that year ranged from 8,000–16,000, which we recorded as 12,000 in estimated fatalities. A complication arose in several PITF-G cases in which a year had a death magnitude index of 5, which signified estimated fatalities of 256,000 or more. Since a midpoint estimate could not be determined for index=5 years, we turned to Harff (2003, p. 60), a precursor of the PITF-G data, to estimate fatalities for cases in which at least one year contained an index of 5 for death magnitude.\(^{72}\)


We identified state-perpetrated mass atrocity cases for the 2007–17 period from the PITF-G dataset and by applying the UV coding protocol for mass killings to the UCDP-V data on civilian killings by governments. The UV coding protocol embodies the following:

A mass killing is “any event in which the actions of state agents result in the intentional death of at least 1,000 noncombatants from a discrete group in a period of sustained violence” (Ulfelder and Valentino, 2008, p. 2, emphasis removed);

and

“Mass killing events were considered to have begun in the first year in which at least 100 intentional noncombatant fatalities occurred. If fewer than 100 total fatalities are recorded annually for any three consecutive years during the event, the event was considered to have ended during the first year within that three-year period in which fatalities dropped below 100 per year (even if killing continues at lower levels in later years)” (Ulfelder and

---

\(^{72}\) This led to the following cases and fatality estimates from Harff (2003) included in our set of mass atrocity cases: Cambodia 1975-79, 2,700,000; Pakistan 1971, 2,000,000; and Rwanda 1995, 750,000. Also included is one case -- Sudan 1983-2002, 2,012,000 – that combined Harff (2003) data for 1983-2001 and PITF-G data for 2002.
When the cases from PITF-G and the UV protocols UCDP-V overlapped, we recorded only one case in order to avoid double counting. The case that was chosen was the one that provided the higher fatality estimate. For UCDP-V, the fatality estimate was the dataset’s “best” estimate rather than an average of its low and high estimates.


We applied the UV coding protocol in II.4 above to the UCDP-V data on civilian killings by nonstate actors (NSAs) to identify NSA mass atrocity cases and estimated fatalities per case. The fatality estimate for each NSA case was UCDP-V’s “best” estimate rather than an average of its low and high estimates. PITF-G also provides several cases of NSA-perpetrated genocides.73

When PITF-G and UCDP-V-created NSA cases overlapped, we recorded only one case in order to avoid double counting. The case that was chosen was the one that provided the higher fatality estimate.


Identifying mass atrocity cases perpetrated by “unidentified actors” is fraught with problems. If the actors in civilian killings are unidentified, they cannot be tied to a unified decisionmaking process or party responsible. As such, it is difficult to claim that civilian fatalities caused by unidentified actors cohere as a case of mass atrocity. Nevertheless, the growth of atrocity acts perpetrated by unknown actors has become a palpable empirical phenomenon, so we take preliminary steps in our article to track such cases. Two datasets that track atrocities by unknown

---

73 The nonstate cases from PITF-G include the Islamic State in Iraq and Syria from 2014-17 and Seleka and anti-Balaka militias in the Central African Republic from 2013-17. A complication arose in the PITF-G dataset in that two of the cases indicated that genocidal activity was perpetrated by the Angolan government and National Union for the Independence of Angola (UNITA) rebels (the two cases spanned the years 1975-1994 and 1998-2002). In these two cases, we assigned half of the estimated fatalities to the state (Angola) and the other half to the NSA (UNITA). This procedure generated two additional NSA mass atrocity cases in the scatterplot (i.e., UNITA 1975 and 1998) based on the PITF-G data.
actors are PITF-W and ACLED.

We applied the following protocols to the PITF-W dataset to generate mass atrocity cases perpetrated by unidentified actors: (1) The case must meet the criteria in the UV mass atrocity protocol noted in section II.4 above. (2) Fatalities from unidentified attacks are counted on a country-year basis. (3) Suicide attacks are generally recorded as perpetrator unknown, but the PITF-W codebook indicates that the source of suicide attacks is often known. As such, we excluded data points in which the main source of attack was a suicide attack. (4) When estimated fatalities spanned a range, we used the low number of the range; when estimated fatalities was “many,” we coded it as 5 fatalities (the low end of fatalities per attack tracked by PITF-W); when estimated fatalities was “dozens,” we coded it as 24. These procedures led to eight unidentified actor mass atrocity cases from the PITF-W dataset from 1995–2017.

We applied protocols 1, 2, and the first part of 4 in the previous paragraph to the ACLED data. The suicide attack issue and estimated fatalities of “many” and “dozens” were not applicable to the ACLED data. These procedures led to six unidentified actor mass atrocity cases from the ACLED dataset from 1997–2017. These cases did not overlap with the unidentified cases from the PITF-W dataset, leading to 14 unidentified cases in the scatterplot in Figure 1.

V. State-Perpetrated Mass Atrocities Overlaps with Wars

A substantial portion of the empirical literature on mass atrocity risks includes a lagged measure of war (inter-, intra-, and/or extra-state) or other form of armed conflict as a right side explanatory or control variable. Our data analysis, however, finds that less than half of mass atrocities occur after war onset and that a nonnegligible number of mass atrocities begin before war onset or occur outside the context of war (see article Section 2). Our coding protocols to determine the overlap, or lack of overlap, between mass atrocity and war are as follows.
1. We use the COW dataset’s onset and end years for inter-, intra-, and extra-state wars. All such wars involve at least one state.

2. We use our compiled dataset of state-perpetrated mass atrocities using the onset and end years available in the data sources we use.

3. We line up each mass atrocity in our dataset with any war, even if it is obvious that the war has nothing to do with the mass atrocity. Note that this protocol is “conservative,” that is, it creates a mass atrocity/war overlap even if there is not an obvious mass atrocity/war overlap.

4. If a mass atrocity overlaps with many wars, we use the notes from the data sources for mass atrocities (EGK, UV, or PITF-G) to identify which war the mass atrocity lines up with and go with that one.

5. If the previous protocol does not give rise to an obvious war in which to make the alignment, we choose “conservatively,” that is, we go with the war that is least advantageous to our claim that mass atrocities do not generally occur contemporaneously with or in the later stages of war.

VI. Comparison of Mass Atrocity Fatalities with Intrastate War Fatalities

Our estimates in the article of 347 intrastate wars and 4,893,230 associated battle-related fatalities for the period 1900–2015 are based on intrastate war data from the Correlates of War Project (http://www.correlatesofwar.org/data-sets/COW-war) for the years 1900–2007 and the Uppsala Armed Conflict dataset (http://ucdp.uu.se/downloads/) for civil and internationalized civil wars from 2008–15.

Appendix References


