

# Journal of Contemporary Criminal Justice

<http://ccj.sagepub.com/>

---

## Hate Crime Research: Design and Measurement Strategies for Improving Causal Inference

Donald P. Green and Amber D. Spry

*Journal of Contemporary Criminal Justice* 2014 30: 228

DOI: 10.1177/1043986214536662

The online version of this article can be found at:

<http://ccj.sagepub.com/content/30/3/228>

---

Published by:



<http://www.sagepublications.com>

Additional services and information for *Journal of Contemporary Criminal Justice* can be found at:

**Email Alerts:** <http://ccj.sagepub.com/cgi/alerts>

**Subscriptions:** <http://ccj.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://ccj.sagepub.com/content/30/3/228.refs.html>

>> [Version of Record](#) - Jul 9, 2014

[What is This?](#)

# Hate Crime Research: Design and Measurement Strategies for Improving Causal Inference

Journal of Contemporary Criminal Justice  
2014, Vol. 30(3) 228–246  
© 2014 SAGE Publications  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1043986214536662  
ccj.sagepub.com



Donald P. Green<sup>1</sup> and Amber D. Spry<sup>1</sup>

## Abstract

The credibility revolution in the social sciences has placed new emphasis on research designs that provide strong evidence of cause and effect. The next generation of hate crime research must move in this design-based direction. This essay reviews recent examples of experiments and quasi-experiments in criminology, political science, and economics that provide useful design templates for hate crime researchers. At the same time, we caution that advances in design must also be accompanied by advances in measurement if researchers are to gauge the long-term effects of interventions designed to reduce the risk of hate crime.

## Keywords

hate crime, causal inference, research design

During the 1980s and 1990s, the terms *hate crime* and *bias crime* took root in popular discourse, media coverage, and public law (Jackson, 2005; Jacobs & Potter, 1998; Messner, McHugh, & Felson, 2004). Although usage varied, the key features of this neologism were embodied in the Federal Hate Crime Statistics Act of 1990, which defined hate crime as

crimes that manifest evidence of prejudice based on race, gender and gender identity, religion, disability, sexual orientation, or ethnicity, including where appropriate the crimes of murder, non-negligent manslaughter; forcible rape; aggravated assault, simple assault, intimidation; arson; and destruction, damage or vandalism of property. (Hate Crimes Statistics Act, 1990)

---

<sup>1</sup>Columbia University, New York, NY, USA

## Corresponding Author:

Donald P. Green, Department of Political Science, Columbia University, 420 W. 118<sup>th</sup> St., New York, NY, 10027 USA.

Email: [dpg2110@columbia.edu](mailto:dpg2110@columbia.edu)

By this definition, hate crime is a behavioral manifestation of prejudice, where the behavior in question is conduct, such as assault, that would otherwise be unlawful. Although this definition of hate crime was and would remain controversial, with varying opinions about which target groups should be covered (Boyd, Berk, & Hamner, 1996; Craig, 1999; Jenness & Broad, 1997; Wang, 1994), whether the definition should include forms of political expression (Lawrence, 2009), and whether motivations can be measured reliably (Berk, 1990; McDevitt et al., 2000), the advent of hate crime as a public policy issue attracted the attention of scholars from a wide variety of disciplines. The 1990s saw the rapid proliferation of academic books and articles across an array of disciplines: law, sociology, criminology, and political science.

Perhaps because hate crime research straddled disciplinary boundaries, scholarship subsided as hate crime faded from front-page news and policy debates. Unlike prejudice, a core topic in social psychology, or ethnic conflict, a core topic in political science, hate crime never achieved the status of a topic on which courses were routinely taught. Ironically, by the turn of the century, hate crime had been upstaged by growing scholarly interest in kindred topics such as genocidal violence (Fearon & Laitin, 2003; Kaufmann, 2006; Kiernan, 2007; Madley, 2005; Peterson, 2002; Verwimp, 2005) and the automatic activation of prejudices (Dasgupta et al., 2000; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald et al., 2009; Gregg, Seibt, & Banaji, 2006). Hate crime research was also methodologically vulnerable. At a time when the social sciences were experiencing a "credibility revolution" (Angrist & Pischke, 2010) that placed new emphasis on the experimental or quasi-experimental research designs that could convincingly demonstrate cause-and-effect, hate crime research was predominantly the study of correlations. Some research sought to look at over-time aggregate relationships between hate crime and macroeconomic conditions (Green, Glaser, & Rich, 1998; McLaren, 1999) or demographic transformations (Esses, Jackson, & Armstrong, 1998; Grattet, 2009; Green, Strolovitch, & Wong, 1998; Olzak, 1989), but for the most part empirical research focused on cross-sectional correlations (Green, Strolovitch, et al., 2001; Messner et al., 2004; Nolan & Akiyama, 1999; Stotzer, 2010; Waldner & Berg, 2008).

Looking back at the rise and decline of hate crime research, one might reasonably ask whether the study of prejudice-motivated crime warrants continuing scholarly attention. In our view, the answer is yes, not only because of the substantive importance of the topic but also because this is one of the few literatures in which theories of prejudice are tested using behavioral outcomes outside the laboratory and outside the United States. But if this literature is to move forward, it must do so in step with methodological advances in social science. Specifically, researchers must place a premium on research designs that can convincingly identify causal effects. As we suggest below, this requirement puts new emphasis not only on experimental designs but also on the development of new measurement techniques for assessing outcomes.

To preview our argument, we begin by defining what we mean by experimental, quasi-experimental, and observational research designs. We next suggest a variety of untapped research opportunities for field experimentation and the investigation of naturally occurring experiments, drawing examples from other substantive domains.

At the same time, we acknowledge an important limitation of experiments that assess the effects of interventions on hate crime rates: These rates reflect the incidence, as opposed to risk of hate crime. If victimized groups relocate to more tolerant areas in the wake of an anti-hate crime intervention, it is possible for an intervention that truly reduces the risk of hate crime to nevertheless raise the incidence of hate crime in treated areas (Bowling, 1994; Ferraro, 1995; LaGrange, Ferraro, & Supancic, 1992; Weisburd et al., 2006). Interventions that actually work may appear to be counterproductive. For the hate crime literature to overcome this problem, researchers must develop alternative outcome measures that gauge the risk of victimization. We sketch out some possible ways of doing so and conclude by envisioning the next generation of hate crime research.

## Experimental and Quasi-Experimental Designs

In the social sciences, the term *experiment* refers to studies in which the units of observation are assigned by some known random process (e.g., a coin flip) to treatment and control conditions. Random assignment helps ensure that treatment and control groups differ systematically only insofar as one group receives the treatment and the other does not.<sup>1</sup> For example, imagine a study in which the objective is to estimate the average effect of an advertising campaign that uses roadside billboards to increase public awareness of hate crime laws. If the unit of analysis were the municipality and the outcomes were gauged by a survey of public awareness conducted a year after the launch of the advertising campaign, an experimental study would randomly assign each municipality to advertising or no advertising with known probability.

An observational study, by contrast, does not employ random assignment; instead, an unknown process determines whether units are treated or not. In the context of the billboards example, an observational study would compare municipalities that happened to deploy public information billboards with those that did not. Although observational studies are easier to conduct than experiments, the lack of control over the assignment process introduces uncertainty when researchers draw causal inferences from the results. If the data were to reveal differences in public awareness between treated and untreated municipalities, does this difference indicate the causal effect of the treatment, or was the deployment of billboards a marker for unmeasured differences in municipalities that are correlated with public awareness? Although researchers may strive to measure these unobserved confounders and control for them statistically, the very fact that they are unobserved implies that one can never entirely rule out threats to causal inference. This fundamental uncertainty undercuts the scientific value of much observational research (Gerber, Green, & Kaplan, 2004).

Somewhere along the continuum between experimental and observational research lies the quasi-experiment (Cook & Campbell, 1979) or natural experiment (Dunning, 2012). This intermediate category encompasses research designs that focus on naturally occurring assignments that are plausibly characterized as random. For example, if a federal government grant were to make funds available to cities of more than

100,000 residents for deploying public information billboards, one could arguably compare cities with just under 100,000 residents to cities with just over 100,000 on the grounds that this arbitrary population threshold partitions these cities in a near random way.<sup>2</sup> Like an observational study, a quasi-experiment does not use a random procedure to determine treatment assignment; however, like an experiment, a quasi-experiment makes use of a design-based identification strategy that focuses attention on a specific set of comparable cases. The persuasiveness of the evidence generated by quasi-experimentation comes down to how plausibly random the treatment versus control comparison is.

## Experimental Designs

As noted above, past research on hate crime has relied almost entirely on observational designs. What would experimental and quasi-experimental designs look like? Let's start with experimental designs. To appreciate the wide array of design options available to researchers, it is instructive to canvass recent work outside the domain of hate crime research. The following four categories of experiments offer useful templates for studies that seek to investigate the causes of hate crime.

### *Randomly Varying Crime Prevention*

The field of criminology has long conducted experiments in field settings. Some of the most ambitious studies have attempted to assess the deterrent effects of increasing police presence in randomly selected locations. An early study (Kelling, Pate, Dieckman, & Brown, 1974) paved the way for more methodologically sophisticated experiments that focus on randomly varying police presence in crime "hotspots" (Sorg, Haberman, Ratcliffe, & Groff, 2013; Weisburd et al., 2006). Outcomes in these experiments are assessed by tracking the incidence of crime in geographic units such as street blocks or neighborhoods. Complementing these area-level experiments are individual-level experiments that look at the effects of arrests (Sherman, Smith, Schmidt, & Rogan, 1992) and incarceration (Green & Winik, 2010; Martin, Annan, & Forst, 1993) on offenders' subsequent criminal activity.

In principle, both ecological and individual research paradigms could be applied to the study of hate crime. For area-level policing experiments to have ample statistical power, the number of hotspots needs to be as large as possible, which means that the study must be set in a location where hate crime incidents occur with high frequency. This requirement may rule out all but the most active epicenters of hate crime, such as inter-communal violence in Indian slums or anti-gay violence in Russia, and also necessitates reliable and ongoing measurement of hate crime in multiple locations. An alternative, discussed below, are measures of risk that may be used to monitor outcomes across locations and time; ideally, such measures would detect variation in the potential for hate crime even when the number of actual incidents is small. Similar opportunities and difficulties apply to experiments involving individual perpetrators. Where ample numbers of perpetrators can be found and where their subsequent

behavior can be monitored, random assignment allows researchers to address long-standing debates about how to assign sentences to minimize recidivism.

### ***Randomly Varying Messaging***

Although messaging experiments date back to the 1940s (Hovland, Lumsdaine, & Sheffield, 1949), fully randomized experiments have been relatively rare in media research until relatively recently (for a review, see Green, Calfano, & Aronow, 2014). An experiment carried out in Rwanda, a society with simmering ethnic tensions following a genocide and civil war, randomly assigned villages to receive tape-recordings of radio soap operas (Paluck, 2009; Paluck & Green, 2009). Half the villages listened to an HIV education soap opera, and half listened to a soap opera designed to reduce ethnic hostility. Outcomes were assessed using surveys and behavioral measures. One limitation of this study is that subjects knew that they were listening to a radio program as part of a research evaluation. Other recent experiments on mass media have used less obtrusive designs, assigning media markets to different levels of television and radio advertising and measuring outcomes in ways that do not call subjects' attention to the connection between media exposure and outcomes. Some studies have used surveys to measure opinion change in treatment and control regions, whereas others have tracked behavioral outcomes using publicly available administrative data, such as voter turnout rates (Green & Vavreck, 2008; Panagopoulos & Green, 2011).

A variety of testable messaging hypotheses may be derived from theories of hate crime. For example, if hate crime reflects perpetrators' beliefs about the public's tacit approval of their behavior (Herek, 1992; Levin & McDevitt, 2002, p. 37), an effective messaging campaign might stress broad consensus denouncing this kind of conduct. If hate crime reflects a deterrence failure whereby perpetrators ignore sentence-enhancing sanctions (McDevitt, Levin, & Bennett, 2002), an effective public messaging might stress these sanctions and the importance of reporting hate crime incidence to police for investigation. Finally, if hate crime reflects widespread intergroup hostility, hate crime might be mitigated by messaging campaigns that promote empathy and dispel negative stereotypes.

### ***Randomly Assigned Changes in Power Arrangements***

A rapidly growing line of field experimental research explores the effects of changing who governs. This research began by focusing on the distributive consequences of installing women as local council heads in Indian villages (Chattopadhyay & Duflo, 2004) and gradually expanded to consider the effects of leadership change on attitudes toward women and minority groups (Beaman, Duflo, Pande, & Topalova, 2012; Bhavani, 2009; Chauchard, 2014). This series of experiments was occasioned by India's 73rd and 74th constitutional amendments that required one third of all local governments to place a woman in the position of local council head. In effect, the Indian government in 1992 set in motion a vast randomized experiment in which underrepresented groups were propelled into positions of power.<sup>3</sup> In addition to

capitalizing on random allocation of leadership positions by governments, researchers have also collaborated with non-governmental organizations that have made the empowerment of women a precondition for aid programs in places like Afghanistan (Beath, Christia, & Enikolopov, 2013), the Democratic Republic of Congo (Humphreys, 2008; Humphreys, Sanchez de la Sierra, & van der Windt, 2012), and Indonesia (Olken, 2010). Of special relevance are studies that assess whether elevating women to council-level positions in male-dominated societies affects the social position of women in the eyes of their fellow male and female villagers (Beaman et al., 2012). Finally, there are studies of randomly varying contact with others, such as Clingingsmith, Khwaja, and Kremer's (2009) study of how Pakistanis' attitudes change as a result of being randomly selected to visit Mecca or McKenzie et al.'s series of studies examining how visa lotteries affect the outcomes of those who are randomly selected to travel abroad (Gibson, McKenzie, & Stillman, 2011; McKenzie, Gibson, & Stillman, 2010).

Hate crime researchers have yet to take advantage of opportunities afforded by quotas or equal-access requirements that empower subordinate groups, such as women or ethnic minorities. Theories advanced by Blumer (1958) and Bobo (1999) imply that interventions that threaten to disrupt social hierarchies precipitate backlash against those who are perceived to be outsiders or upstarts. Indeed, hate crime is sometimes characterized as an informal use of power to maintain hierarchy when formal policing breaks down (Franklin, 2000). One interesting feature of this line of research is that it pits two competing hypotheses against one another: Formal rules that affirm norms of equality are thought to promote tolerance (Franklin, 2000; Sniderman, Fletcher, Russell, Tetlock, & Gaines, 1991) while provoking extra-legal behavior among those who reject these norms.

### *Audit Studies to Measure Discrimination*

One of the largest experimental research literatures in social science assesses the extent to which people of varying ethnic, racial, gender, or class attributes receive different treatment at the hands of employers, realtors, bankers, and retailers (Fix & Struyk, 1993; Marias, Patell, & Wolfson, 1985; Banerjee, Bertrand, Datta, & Mullainathan, 2009; Ondrich, Ross, & Yinger, 2000; Riach & Rich, 2002). Many of these studies use a matched pair design whereby two putatively identical people with different racial attributes attempt to lease the same apartment (Ondrich, Ross, & Yinger, 2000) or apply for the same job (Pager, 2007). In recent years, this literature has increasingly featured communication via the Internet, as many job and housing markets have moved online (Ahmed & Hammarstedt, 2008). Starting with Bertrand and Mullainathan (2004), a burgeoning literature has tested whether employers discriminate based on race and gender, whether teachers grade based on ethnicity (Sprietsma, 2012), and whether legislators respond to putatively Black and White constituents' questions at similar rates (Butler & Brockman, 2011). Closer to the domain of hate crime, experiments have been used to discern which kinds of issues are most likely to make White supremacists advocate violence in chat rooms (Glaser, Dixit, & Green, 2002).

The use of audit experiments has many applications in the realm of hate crime. For example, working in collaboration with undercover police, researchers could assess the relative risk that gay or straight couples are harassed as they walk through a neighborhood. Below, we consider the important role of audits to measure the risks faced by target groups in various settings. Other applications involve the way in which public officials, police, or journalists respond to hate crime allegations. For example, working in collaboration with local groups that informally monitor hate crime complaints, researchers could assess how the attributes of the victim affect the speed, seriousness, and empathy with which police respond to a hate crime complaint. This design could be implemented without use of deception if the monitoring organization gathers complaints that could legitimately be construed in more than one way (e.g., incidents directed against victims who are both racial minorities and gay); the experiment would vary which identity is described initially as the basis of the incident.

## **Quasi-Experimental Designs**

Sometimes practical and ethical constraints preclude the use of randomized experiments, in which case the next best option is a study that capitalizes on an as-if random intervention, such as a discontinuity, abrupt policy change, or unforeseen event. Four types of quasi-experimental designs may serve as models for hate crime research.

### *Discontinuities*

The use of discontinuity designs to identify causal effects has become increasingly popular in the social sciences. The basic idea behind discontinuity designs is to compare observations that have all but identical background attributes but receive different treatments due to the application of rigid rule. A classic case is the comparison of students whose entrance exam scores vary by only 1 point, with some barely qualifying for a scholarship and others falling 1 point short. Because data tend to be sparse at the point of discontinuity (e.g., the cutoff score required for a scholarship), researchers use regression to leverage information from observations that lie farther away from the discontinuity (see Imbens & Lemieux, 2008).

Substantive applications of discontinuity designs are quite diverse. For example, Folke and Snyder (2012) study the policy consequences of electing a right-wing party candidate as municipal council representative, Hopkins and McCabe (2012) study the budgetary decisions of White or minority mayors in the United States, and Dunning and Nilekani (2013) study the consequences of installing lower caste village leaders in India. One could use the Folke and Snyder (2012) design to evaluate the hypothesis that office holding and hate crime are substitutes—that higher rates of hate crime occur in polities without political representation for right-wing sentiment. And, akin to experimental studies of India's random rotation of women council leaders, one could use a discontinuity design to study whether elevating a member of a subordinate group to a position of power leads to a surge of hate crimes directed at that group.



## *Interrupted Time Series*

This method focuses on the effects of a sudden shock that is hypothesized to generate a shift in outcomes. Classic examples of this methodology may be found in Campbell and Ross (1970), who use interrupted time-series analysis to evaluate the effectiveness of interventions designed to diminish traffic violations. More recent applications to criminology include evaluations of police interventions (McGarrell, Chermak, Weiss, & Wilkson, 2001; Novak, Hartman, Holsinger, & Turner, 1999) and the implementation of new laws or procedures that affect the severity of punishment or the probability of apprehension (Dube, Dube, & García-Ponce, 2013; Vásquez, Maddan, & Walker, 2008). Because these interventions are not randomly timed, researchers must be alert to threats to inference, especially regression to the mean (see Cook & Campbell, 1979).

This type of approach has occasionally been applied to the study of hate crime. Testing whether a terrorist attack provokes retaliatory hate crime against people who are linked (perhaps mistakenly) to the terrorist group, Hanes and Machin (2014) studied the consequences of the 7/7 attacks in London and the 9/11 attacks in the United States on subsequent anti-Arab hate crime. In this example, the authors construct a time series of hate crime attacks against a set of target groups and look for a short-term bump in hate crimes after the terrorist event; the persuasiveness of this approach is bolstered by placebo tests that show no corresponding jump in hate crimes directed at other groups. Similarly, King and Sutton (2013) used an interrupted time-series design to test whether particular types of hate crimes increased after antecedent events such as the Rodney King and O. J. Simpson verdicts, the 9/11 attacks, and court rulings that legalized same-sex marriage.

The research task going forward is to assess how general this pattern is—Does it hold cross-nationally? Does the response in terms of hate crime depend on whether the attackers are perceived to have acted alone or whether they are linked to a broader political movement? A related line of work is the study of local events that serve as a flashpoint for hate crime, such as the opening of a gay bar or Muslim community center. So long as researchers obtain reliable measurements of hate crime before and after the event for both affected and unaffected areas, a difference-in-differences estimator (Dunning, 2012) may be used to isolate the causal effect of the flashpoint event. Again, caution is required when interpreting the results, especially when the timing of the intervention is non-random and selection of comparison groups is post hoc (Bertrand, Duflo, & Mullainathan, 2004).

## *Recurrent Cycles*

Another source of plausibly exogenous variation in treatments is the set of events that recur at regular periods. For example, Levitt (1997) uses the timing of municipal elections in the United States as a source of exogenous variation in the size of police forces, which in turn is used as a factor affecting crime rates (for a correction of Levitt, 1997, see McCrary, 2002). Election cycles have also been linked to surges in ethnic

violence in India (Wilkinson, 2004, but see Varshney & Gubler, 2012). Spikes in summer heat and global climate change have been used as a source of interpersonal aggression (Anderson, 2012; Kenrick & MacFarlane, 1986; Miguel, Satyanath, & Sergenti, 2004). Rotation in which countries' representatives head international institutions has been linked to foreign aid patterns, which in turn are thought to affect the frequency of human rights abuses (Aronow, Carnegie, & Marinov, 2012).

### *Natural Experiments*

Researchers typically use the term *natural experiment* to refer to the study of a broadly similar collection of observations that differ only with respect to exposure to an intervention (Rosenzweig & Wolpin, 2000). These designs often make use of natural disasters or fluctuations in weather to identify the effects of scarcity on ethnic conflict (Nel & Righarts, 2008) or election outcomes on policy (Gomez, Hansford, & Krause, 2007). Some applications seek out instances where observations receive almost identical "treatments." For example, Jones and Olken (2009) analyze the effects of assassinations on democratization by comparing instances in which national leaders were attacked and killed with those in which national leaders were attacked but survived.

Because what constitutes a closely comparable set of observations is a matter of opinion, applying the term natural experiment to an observational study is sometimes controversial. An arguable example of a natural experiment in the context of hate crime research is Krueger and Pischke's (1997) study of hate crime in post-reunification Germany, which experienced an abrupt change in its immigrant settlement policy and shortly thereafter saw a surge in anti-immigrant hate crime. Among the challenges of studying the effects of changes in hate crime law are the endogenous timing of adoption of new laws (the concern is that adoption may be precipitated by a spike in hate crime) and the fact that the measurement of hate crime itself tends to become more systematic after the adoption of hate crime laws, which may render before-after comparisons uninterpretable.

### **Further Challenges: Measurement of Risk**

Suppose that researchers in years ahead were to assemble a series of well-designed experiments and quasi-experiments. There is no question that such a development would advance the understanding of cause and effect as it pertains to hate crime. However, even if stronger experimental and quasi-experimental designs were adopted, there remain barriers to inference that cannot be overcome without new methods for measuring hate crime. Some of these measurement issues, such as the need for comparable measurement standards before and after the implementation of a new policy or program, have been mentioned above. Below we highlight an additional measurement issue concerning the distinction between incidence and risk.

To appreciate this subtle measurement issue, consider the following scenario. Suppose researchers in a large jurisdiction were to conduct a randomized experiment designed to reduce the number of hate crimes occurring in certain hotspots. The

specific intervention could be any of the ideas discussed in the previous section; for concreteness, suppose that the intervention featured increased police surveillance in an effort to deter hate crime. Suppose further that the study were conducted impeccably: The allocation of police surveillance faithfully follows the random assignment, and the daily incidence of hate crime is recorded accurately in every geographic location for the duration of the study. A study of this sort would surely contribute valuable insights regarding the short-term effects of increased policing: We might obtain highly credible evidence that police presence does indeed reduce the number of hate crime incidents that occur during periods of heightened surveillance.

The limitations of this type of experiment become apparent when we consider the long-term effects of this intervention.<sup>4</sup> If a hate crime hotspot were to become substantially safer for potential victims, they might visit these areas with greater frequency. Perhaps they might become more likely to select a former hotspot as a place to live or work. If our only measure of outcomes were the number of hate crime incidents, we might observe a perverse pattern over the long run: As areas become safer for potential victims, the number of hate crimes directed against them increases. This scenario is familiar to those who study the correlation between hate crimes and demographic profiles. Relatively tolerant areas, such as New York's Greenwich Village, attract large numbers of gay men and lesbians; these areas also tend to record the greatest number of hate crimes directed at gay men and lesbians (Bailey, 1999; Green, Strolovitch, et al., 2001).

The problem is that the number of hate crimes does not necessarily provide an accurate assessment of the *risk* of victimization. In the preceding example, one could imagine the incidence of hate crimes rising in an effectively treated hotspot precisely because potential victims perceived a decrease in risk. If each potential victim is exposed to half the risk of hate crime in the wake of the intervention, *ceteris paribus*, the expected number of hate crimes will increase if the number of potential victims more than doubles. The same logic applies to forms of behavior that vary in risk: If in the wake of the intervention, certain behaviors become safer (e.g., same-sex couples holding hands in public), the number of people engaging in those behaviors might also increase, possibly leading to an increase in the number of hate crimes. In effect, if one imagines hate crime to be a probabilistic process in which potential attackers are occasionally confronted with behaviors they find intolerable, even interventions that make hate crime more costly to perpetrators might nonetheless lead to more hate crime because the number and salience of provocations increases.

An analogous issue arises in other contexts. Improvement in brakes reduces the risk of traffic accidents at a given level of speed; however, truck and taxi drivers respond to improved brakes by increasing their average speed (Aschenbrenner & Biehl, 1987; Vaa, 2013). The net effect is to leave the number of accidents unchanged. Clearly, it would be incorrect to conclude from the over-time incidence of taxicab accidents that improved brakes were ineffective; rather, they produced two countervailing effects that offset one another, faster speeds and reduced risk at a given speed. In the context of hate crime, we would ordinarily applaud innovations (such as increased police surveillance) that

reduce risk even if it means that the number of incidents remains constant because more potential victims were ultimately exposed to risk.

For hate crime researchers, the conundrum is that the lack of apparent long-term effects on the number of hate crimes is consistent with two very different hypotheses. The first is the one described above, whereby diminished risk increases the number and type of risky behaviors. The second is that the intervention has no long-term effects whatsoever. One way to adjudicate between these hypotheses is to measure risk directly. If the intervention were to significantly reduce risk while leaving incidence unchanged, researchers would have a clearer sense of the intervention's immediate and long-term repercussions.

How would one construct a systematic measurement of risk? The measurement protocol resembles an audit experiment in which confederates periodically engage in behaviors that are believed to provoke hate crime. For example, an interracial couple might hold hands while walking down a street, or a person dressed in the garb of a religious minority might attempt to deliver flyers to passersby. Outcome measures might range from open expressions of hostility to more subtle nonverbal behaviors indicating aversion or disdain. If such auditing behaviors were deployed at randomly selected times and locations in sampling units corresponding to experimental units (e.g., neighborhoods), one could use the rate of each type of behavior as an outcome measure in an experimental study. Moreover, one could supplement this measurement approach with other modes of data collection, such as a probability survey of local residents that asks respondents about the perceived risk of hate crime associated with assorted vignettes (see Institute for Social Research, 1995). Recognizing that no measure of hate crime risk is free from flaws, the researcher nevertheless attempts to apply systematic procedures redundantly and symmetrically to treatment and control units. Redundancy helps insulate the conclusions from bias associated with a particular measure's flaws; symmetry ensures that whatever the flaws may be, both the treatment and control group are equally exposed to them so that the difference between treatment and control, in expectation, reflects the treatment effect.

Of course, undertaking a systematic risk assessment only adds to the already perplexing measurement problems confronting the study of hate crime. Audit studies are potentially dangerous and require some degree of collaboration with police to limit the ill effects of an encounter with an attacker. Although policing in other domains (e.g., robbery) sometimes involves undercover work whereby plainclothes officers pose as potential victims, this type of work is seldom if ever done according to a random sampling scheme that covers both high- and low-risk areas. To implement an audit therefore requires introducing undercover activity to the domain of hate crime while at the same time introducing the concept of unobtrusive measurement as an activity distinct from the apprehension of perpetrators. This is a tall order; however, the past decade has witnessed increasingly ambitious measurement strategies for assessing hard-to-assess outcomes such as inter-ethnic trust (Habyarimana, Humphreys, Posner, & Weinstein, 2011), prejudice (Kremer, Barry, & McNally, 1986), or deference to authority in post-conflict societies (Paluck & Green, 2009).

## Conclusion

The literature on hate crime has lost some momentum since its apogee in the 1990s. However, like a society that industrializes late, it may ultimately profit from the passage of time. Social science has become more methodologically sophisticated in the interim, and the study of hate crime stands to benefit from the many developments in the design and application of experiments and quasi-experiments. By reviewing some recent exemplars, this essay has sought to encourage hate crime scholars to look for well-identified and theoretically informative research opportunities.

It should be stressed that the goal going forward is not merely to improve the methodological quality of this research literature but rather to conduct well-crafted research that investigates substantively important aspects of hate crime. A frequently voiced concern about identification-oriented social science is that it privileges well-executed but substantively trivial research. Although we think this concern is unlikely to apply to identification-oriented studies of hate crime, we nevertheless conclude by suggesting three broad research agendas that both lend themselves to credible causal inference and speak to long-standing theoretical and policy questions.

One research agenda is the causal link between political institutions and hate crime. For example, do systems of representation that allow extremist parties provide an outlet for xenophobic sentiment and thereby reduce the expression of hate crime? Or do such parties legitimate and increase the risk of hate crime? This kind of research topic could be addressed through the analysis of interrupted time series or discontinuities, using changes in rules of representation or the narrow victories/defeats of extremist candidates. A related question concerns state capacity, or the ability of formal institutions to apprehend, prosecute, and punish offenders. Over-time changes in technology, training, and transportation that expand state capacity set the stage for researchers to investigate downstream effects on intergroup violence.

A second agenda concerns the relationship between macroeconomic conditions and hate crime. Many theories of intergroup conflict argue that competition and resentment precipitate out-group hostility and provide incentives for mass media and political leaders to channel this sentiment into xenophobic forms of collective action. Cross-regional differences in economic performance may be fraught with unobserved heterogeneity, but the over-time study of changing economic conditions, especially in the wake of external shocks, may be informative. Moreover, the many community-based aid experiments that target significant financial resources to villages in developing countries make it possible to investigate the basic question of whether exogenous shocks to local economic conditions affect inter-ethnic attitudes or behaviors.

Third, the role of social norms lends itself to well-identified and substantively significant research. It is often argued that what people do in everyday life is heavily shaped by their perceptions about what others think. In particular, people are sensitive to standards—real or imagined—of socially acceptable attitudes and behaviors. This theoretical perspective implies that interventions that clarify or dramatize social norms may have a profound effect on the manner in which out-groups are regarded and treated, especially when it comes to extra-legal behaviors such as hate crime. This

proposition lends itself to experimental investigation in schools, via the mass media, or through any channel that can convey social norms about tolerance and respect for the rights of others in an authoritative and convincing manner. To be sure, this abbreviated list of research topics in no way exhausts the range of interesting questions to be asked. Our point is simply that methodologically sophisticated and substantively telling research is feasible given ample research effort.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### Notes

1. See Gerber and Green (2012) for a more detailed discussion of the statistical properties of experiments and the conditions under which they render unbiased estimates of causal parameters.
2. The assumption of an as-if random breakpoint would be jeopardized if some municipalities contrived to raise their populations above 100,000 to extract more federal funds. On the rationale for discontinuity designs and the assumptions on which the analysis of discontinuities depend, see Imbens and Lemieux (2008).
3. Similar randomized quota policies have been enacted in places such as Lesotho.
4. In the parlance of economics, these long-term consequences are termed general equilibrium effects. See Heckman, Lochner, and Taber (1998).

### References

- Ahmed, A. H., & Hammarstedt, M. (2008). Discrimination in the rental housing market: A field experiment on the internet. *Journal of Urban Economics*, 64, 362-372.
- Anderson, C. A. (2012). Climate change and violence. In D. J. Christie (Ed.), *The encyclopedia of peace psychology* (pp. 128-132). West Sussex, UK: John Wiley & Sons.
- Angrist, J., & Pischke, J. S. (2010). *The credibility revolution in empirical economics: How better research design is taking the con out of econometrics* (No. w15794). National Bureau of Economic Research.
- Aronow, P. M., Carnegie, A. S., & Marinov, N. (2012). *The effects of aid on rights and governance: Evidence from a natural experiment*. Retrieved from <http://ssrn.com/abstract=2199131>
- Aschenbrenner, K. M., & Biehl, B. (1987). Impact of risk compensation on the effectiveness of traffic safety measures in the case of ABS. *Series of Accident and Safety Research of Road Traffic*, 63, 65-70. Bergisch Gladbach, Germany: Federal Institute for Road Engineering.
- Bailey, R. W. (1999). *Gay politics, urban politics: Identity and economics in the urban setting*. New York, NY: Columbia University Press.
- Banerjee, A., Bertrand, M., Datta, S., & Mullainathan, S. (2009). Labor market discrimination in Delhi: Evidence from a field experiment. *Journal of Comparative Economics*, 37(1), 14-27.

- Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, *335*, 582-586.
- Beath, A., Christia, F., & Enikolopov, R. (2013). Empowering women through development aid: Evidence from a field experiment in Afghanistan. *American Political Science Review*, *107*, 540-557.
- Berk, R. A. (1990). Thinking about hate-motivated crimes. *Journal of Interpersonal Violence*, *5*, 334-349.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*, 249-275.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*, 991-1013.
- Bhavani, R. R. (2009). Do electoral quotas work after they are withdrawn? Evidence from a natural experiment in India. *American Political Science Review*, *103*, 23-35.
- Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review*, *1*, 3-7.
- Bobo, L. D. (1999). Prejudice as group position: Microfoundations of a sociological approach to racism and race relations. *Journal of Social Issues*, *55*, 445-472.
- Bowling, B. (1994). Racial harassment in East London. In M. S. Hamm (Ed.), *Hate crime: International perspectives on causes and control* (pp. 1-36). Cincinnati, OH: Anderson.
- Boyd, E. A., Berk, R. A., & Hamner, K. M. (1996). "Motivated by hatred or prejudice": Categorization of hate-motivated crimes in two police divisions. *Law & Society Review*, *30*, 819-850.
- Butler, D. M., & Broockman, D. E. (2011). Do politicians racially discriminate against constituents? A field experiment on state legislatures. *American Journal of Political Science*, *55*, 463-477.
- Campbell, D. T., & Ross, H. L. (1970). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. In E. R. Tufts (Ed.), *The quantitative analysis of social problems* (pp. 110-118). Reading, MA: Addison-Wesley.
- Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, *72*, 1409-1443.
- Chauchard, S. (2014). Can Descriptive Representation Change Beliefs about a Stigmatized Group? Evidence from Rural India. *American Political Science Review*, 1-20.
- Clingingsmith, D., Khwaja, A. I., & Kremer, M. R. (2009). Estimating the impact of the Hajj: Religion and tolerance in Islam's global gathering. *The Quarterly Journal of Economics*, *124*, 1133-1170.
- Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings* (pp. 19-21). Boston: Houghton Mifflin.
- Craig, K. M. (1999). Retaliation, fear, or rage: An investigation of African-American and white reactions to racist hate crimes. *Journal of Interpersonal Violence*, *14*, 138-151.
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*(3), 316-328.
- Dube, A., Dube, O., & Garcia-Ponce, O. (2013). Cross-border spillover: U.S. gun laws and violence in Mexico. *American Political Science Review*, *107*, 397-417.
- Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. New York, NY: Cambridge University Press.
- Dunning, T., & Nilekani, J. (2013). Ethnic quotas and political mobilization: Caste, parties and distribution in Indian village councils. *American Political Science Review*, *107*, 35-56.

- Esses, V. M., Jackson, L. M., & Armstrong, T. L. (1998). Intergroup competition and attitudes toward immigrants and immigration: An instrumental model of group conflict. *Journal of Sociology, 54*, 699-724.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline. *Journal of Personality and Social and Psychology, 69*, 1013-1027.
- Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review, 97*(1), 75-90.
- Ferraro, K. F. (1995). *Fear of crime: Interpreting victimization risk*. Albany: State University of New York Press.
- Fix, M., & Struyk, R. J. (Eds.). (1993). *Clear and convincing evidence: Measurement of discrimination in America*. Washington, DC: Urban Institute Press.
- Folke, O., & Snyder, J. M. (2012). Gubernatorial midterm slumps. *American Journal of Political Science, 56*, 931-948.
- Franklin, K. (2000). Antigay behaviors among young adults: Prevalence, patterns, and motivators in a noncriminal population. *Journal of Interpersonal Violence, 15*, 339-362.
- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. New York, NY: W.W. Norton.
- Gerber, A. S., Green, D. P., & Kaplan, E. H. (2004). The illusion of learning from observational research. In I. Shapiro, R. Smith, & E. Masoud (Eds.), *Problems and methods in the study of politics* (p. 251). Cambridge, UK: Cambridge University Press.
- Gibson, J., McKenzie, D., & Stillman, S. (2011). The impacts of migration on remaining household members: Omnibus results from a migration lottery program. *Review of Economics and Statistics, 93*, 1297-1317.
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues, 58*, 177-193.
- Gomez, B. T., Hansford, T. G., & Krause, G. A. (2007). The republicans should pray for rain: Weather, turnout, and voting in U.S. Presidential elections. *Journal of Politics, 69*, 649-663.
- Grattet, R. (2009). The urban ecology of bias crime: A study of disorganized and defended neighborhoods. *Social Problems, 56*, 132-150.
- Green, D. P., Calfano, B. R., & Aronow, P. M. (2014). Field experimental designs for the study of media effects. *Political Communication, 31*, 168-180.
- Green, D. P., Glaser, J., & Rich, A. (1998). From lynching to gay bashing: The elusive connection between economic conditions and hate crime. *Journal of Personality and Social Psychology, 75*, 82-92.
- Green, D. P., McFalls, L. H., & Smith, J. K. (2001). Hate crime: An emergent research agenda. *Annual Review of Sociology, 27*, 479-504.
- Green, D. P., Strolovitch, D. Z., & Wong, J. S. (1998). Defended neighborhoods, integration, and racially motivated crime. *American Journal of Sociology, 104*, 372-403.
- Green, D. P., Strolovitch, D. Z., Wong, J. S., & Bailey, R. W. (2001). Measuring gay populations and anti-gay hate crime. *Social Science Quarterly, 82*, 281-296.
- Green, D. P., & Vavreck, L. (2008). Analysis of cluster-randomized field experiments: A comparison of alternative estimation approaches. *Political Analysis, 16*, 138-152.
- Green, D. P., & Winik, D. (2010). Using Random Judge Assignments to Estimate the Effects of Incarceration and Probation on Recidivism Among Drug Offenders. *Criminology, May*.
- Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009). Implicit race attitudes predicted vote in the 2008 US presidential election. *Analyses of Social Issues and Public Policy, 9*(1), 241-253.



- Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of personality and social psychology*, *90*(1), 1.
- Habyarimana, J., Humphreys, M., Posner, D., & Weinstein, J. M. (2011). *Coethnicity: Diversity and the dilemmas of collective action*. New York, NY: Russell Sage Foundation.
- Hanes, E., & Machin, S. (2014). *Hate crime in the wake of terror attacks: Evidence from 7/7 and 9/11*. *Journal of Contemporary Criminal Justice* *30*(3), 247-269.
- Hate Crimes Statistics Act, 28 U.S.C. § 534 (1990).
- Heckman, J. J., Lochner, L., & Taber, C. (1998). General-equilibrium treatment effects: A study of tuition policy. *American Economic Review*, *88*, 381-386.
- Herek, G. M. (1992). Psychological heterosexism and antigay violence: The social psychology of bigotry and bashing. In G. M. Herek & K. T. Berrill (Eds.), *Hate crimes: Confronting violence against lesbians and gay men* (pp. 149-169). Thousand Oaks, CA: Sage.
- Hopkins, D. J., & McCabe, K. T. (2012). After It's Too Late Estimating the Policy Impacts of Black Mayoralties in US Cities. *American Politics Research*, *40*(4), 665-700.
- Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. (1949). *Experiments on mass communication*. Princeton, NJ: Princeton University Press.
- Humphreys, M. (2008). *Community-Driven Reconstruction in the Democratic Republic of Congo: Baseline Report*. Unpublished Manuscript. Columbia University. Available at <http://www.columbia.edu/~mh2245/DRC/report.pdf>
- Humphreys, M., De La Sierra, R. S., & Van Der Windt, P. (2012). *Social and Economic Impacts of Tuungane: Final Report on the Effects of a Community Driven Reconstruction Program in Eastern Democratic Republic of Congo*. Unpublished, Department of Political Science, Columbia University.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615-635.
- Institute for Social Research. (1995). *The Carolina poll of the School of Journalism and the Institute for Research in the Social Sciences*. Chapel Hill: University of North Carolina Press.
- Jackson, J. (2005). Validating new measures of the fear of crime. *International Journal of Social Research Methodology*, *8*, 297-315.
- Jacobs, J. B., & Potter, K. A. (1998). *Hate crimes: Criminal law and identity politics*. New York, NY: Cambridge University Press.
- Jenness, V., & Broad, K. (1997). *Hate crimes: New social movements and the politics of violence*. New York, NY: Gruyter.
- Jones, B. F., & Olken, B. A. (2009). Hit or miss? The effect of assassinations on institutions and war. *American Economic Journal*, *1*(2), 55-87.
- Kaufman, S. (2006). Symbolic politics or rational choice? Testing theories of extreme ethnic violence. *International Security*, *30*(4), 45-86.
- Kelling, G. L., Pate, T., Dieckman, D., & Brown, C. E. (1974). *The Kansas city preventive patrol experiment: A summary report*. The Police Foundation. Retrieved from <https://www.ncjrs.gov/pdffiles1/Digitization/42537NCJRS.pdf>.
- Kenrick, D. T., & MacFarlane, S. W. (1986). Ambient temperature and horn honking: A field study of the heat/aggression relationship. *Environment & Behavior*, *18*, 179-191.
- Kiernan, B. (2007). *Blood and soil: A world history of genocide and extermination from Sparta to Darfur*. Yale University Press.
- King, R. D., & Sutton, G. M. (2013). High times for hate crimes: Explaining the temporal clustering of hate-motivated offending. *Criminology*, *51*, 871-894.

- Kremer, J., Barry, R., & McNally, A. (1986). The misdirected letter and the quasi-questionnaire: Unobtrusive measures of prejudice in Northern Ireland. *Journal of Applied Social Psychology, 16*, 303-309.
- Krueger, A. B., & Pischke, J. (1997). A statistical analysis of crime against foreigners in Germany. *Journal of Human Resources, 32*, 182-209.
- LaGrange, R. L., Ferraro, K. F., & Supancic, M. (1992). Perceived risk and fear of crime: Role of social and physical incivilities. *Journal of Research in Crime & Delinquency, 29*, 311-334.
- Lawrence, F. (1999). *Punishing hate: Bias crimes under American law*. Cambridge, MA: Harvard University Press.
- Levin, J., & McDevitt, J. (2002). *Hate crimes revisited: America's war on those who are different*. New York, NY: Basic Books.
- Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review, 87*, 270-290.
- Madley, B. (2005). From Africa to Auschwitz: How German South West Africa incubated ideas and methods adopted and developed by the Nazis in Eastern Europe. *European History Quarterly, 35*(3), 429-464.
- Marias, M. L., Patell, J. M., & Wolfson, M. A. (1985). The experimental design of classification tests: The case of commercial bank loan classifications. *Journal of Accounting Research, 22*, 87-118.
- Martin, S. E., Annan, S., & Forst, B. (1993). The special deterrent effects of a jail sanction on first-time drunk drivers: A quasi-experimental study. *Accident Analysis & Prevention, 25*(5), 561-568.
- McCrary, J. (2002). Do electoral cycles in police hiring really help us estimate the effect of police on crime? Comment. *American Economic Review, 92*, 1236-1243.
- McDevitt, J., Balboni, J. M., Bennett, S., Weiss, J. C., Orchowsky, S., & Walbolt, L. (2000). *Improving the quality and accuracy of bias crime statistics nationally: An assessment of the first ten years of bias crime data collection*. Washington, DC: Bureau of Justice Statistics.
- McDevitt, J., Levin, J., & Bennett, S. (2002). Hate crimes offenders: An expanded typology. *Journal of Social Issues, 58*, 303-317.
- McGarrell, E. F., Chermak, S., Weiss, A., & Wilkson, J. (2001). Reducing firearms violence through directed police patrol. *Criminology & Public Policy, 1*, 119-148.
- McKenzie, D., Gibson, J., & Stillman, S. (2010). How important is selection? Experimental vs non-experimental measures of the income gains from migration. *Journal of the European Economic Association, 8*, 913-945.
- McLaren, L. M. (1999). Explaining right-wing violence in Germany: a time series analysis. *Social Science Quarterly, 80*(1), 166-180.
- Messner, S. F., McHugh, S., & Felson, R. (2004). Distinctive characteristics of assaults motivated by bias. *Criminology, 42*, 585-618.
- Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy, 112*, 725-753.
- Nel, P., & Righarts, M. (2008). Natural disasters and the risk of violent civil conflict. *International Studies Quarterly, 52*, 159-185.
- Nolan, J. J., & Akiyama, Y. (1999). An analysis of factors that affect law enforcement participation in hate crime reporting. *Journal of Contemporary Criminal Justice, 15*, 111-127.
- Novak, K. J., Hartman, J. L., Holsinger, A. M., & Turner, M. G. (1999). The effects of aggressive policing of disorder on serious crime. *Policing: An International Journal of Police Strategies and Management, 22*, 171-194.

- Olken, B. A. (2010). Direct democracy and local public goods: Evidence from a field experiment in Indonesia. *American Political Science Review*, *104*, 243-367.
- Olzak, S. (1989). Labor unrest, immigration, and ethnic conflict in urban America, 1880-1914. *American Journal of Sociology*, *94*, 1303-1333.
- Ondrich, J., Ross, S. L., & Yinger, J. (2000). How common is housing discrimination? Improving on traditional measures. *Journal of Urban Economics*, *47*, 470-500.
- Pager, D. (2007). The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *The ANNALS of the American Academy of Political and Social Science*, *609*, 104-133.
- Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology*, *96*, 574-587.
- Paluck, E. L., & Green, D. P. (2009). Deference, dissent, and dispute resolution: A field experiment on a mass media intervention in Rwanda. *American Political Science Review*, *103*, 622-644.
- Panagopoulos, C., & Green, D. P. (2011). Spanish-language radio advertisements and Latino voter turnout in the 2006 Congressional elections: Field experimental evidence. *Political Research Quarterly*, *64*, 588-599.
- Petersen, R. D. (2002). *Understanding ethnic violence: Fear, hatred, and resentment in twentieth-century Eastern Europe*. Cambridge University Press.
- Riach, P. A., & Rich, J. (2002). Field experiments of discrimination in the market Place. *The Economic Journal*, *112*(483), F480-F518.
- Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature*, *38*, 827-874.
- Sherman, L. W., Smith, D. A., Schmidt, J. D., & Rogan, D. P. (1992). Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review*, *57*, 680-690.
- Sniderman, P. M., Fletcher, J. F., Russell, P. H., Tetlock, P. E., & Gaines, B. J. (1991). The fallacy of democratic elitism: Elite competition and commitment to civil liberties. *British Journal of Political Science*, *21*, 349-370.
- Sorg, E. T., Haberman, C. P., Ratcliffe, J. H., & Groff, E. R. (2013). Foot patrol in violent crime hot spots: The longitudinal impact of deterrence and posttreatment effects of displacement. *Criminology*, *51*, 65-101.
- Spruijsma, M. (2012). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, *45*, 523-538.
- Stotzer, R. L. (2010). Sexual orientation-based hate crimes on campus: The impact of policy on reporting rates. *Sexuality Research and Social Policy: Journal of NSRC*, *7*, 147-154.
- Vaa, T. (2013). Psychology of behavioural adaptation. In C. Rudin-Brown & S. Jamson (Eds.), *Behavioural adaptation and road safety* (pp. 207-224). Boca Raton, FL: CRC Press.
- Varshney, A., & Gubler, J. R. (2012). Does the state promote communal violence for electoral reasons? *India Review*, *11*, 191-199.
- Vásquez, B. E., Maddan, S., & Walker, J. T. (2008). The influence of sex offender registration and notification laws in the United States: A time-series analysis. *Crime & Delinquency*, *54*, 175-192.
- Verwimp, P. (2005). An economic profile of peasant perpetrators of genocide: Micro-level evidence from Rwanda. *Journal of Development Economics*, *77*(2), 297-323.
- Waldner, L. K., & Berg, J. (2008). Explaining antigay violence using target congruence: An application of revised routine activities theory. *Violence and Victims*, *23*(3), 267-287.

- Wang, L. (1994). *Hate crime laws*. New York, NY: Clark Boardman Callaghan.
- Weisburd, D., Wyckoff, L. A., Ready, J., Eck, J. E., Hinkle, J. C., & Gajewski, F. (2006). Does crime just move around the corner? A controlled study of special displacement and diffusion of crime control benefits. *Criminology*, 44, 549-591.
- Wilkinson, S. I. (2004). *Votes and violence: Ethnic riots in India*. New York, NY: Cambridge University Press.

### Author Biographies

**Donald P. Green** is Professor of Political Science at Columbia University. The author of four books and more than 100 essays, his research spans a wide array of topics: hate crime, prejudice, voting behavior, partisanship, and campaign finance. He has also written extensively on social science research methods and is co-author of the textbook *Field Experiments: Design, Analysis, and Interpretation* (W.W. Norton Press, 2012).

**Amber D. Spry** is a doctoral candidate in the Department of Political Science at Columbia University whose research examines group identity and political behavior in American politics. Amber is a recipient of the National Science Foundation Graduate Research Fellowship. Prior to Columbia, Amber conducted social and economic policy research at the Urban Institute in Washington, D.C.