

2008

Belief in a Just World, Blaming the Victim, and Hate Crime Statutes

Dharmika Dharmapala

Nuno Garoupa

Richard H. McAdams

Follow this and additional works at: https://chicagounbound.uchicago.edu/public_law_and_legal_theory

 Part of the [Law Commons](#)

Chicago Unbound includes both works in progress and final versions of articles. Please be aware that a more recent version of this article may be available on Chicago Unbound, SSRN or elsewhere.

Recommended Citation

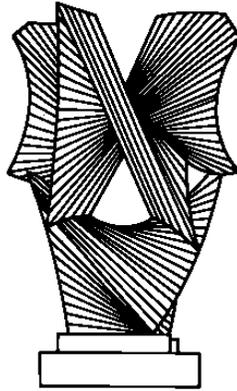
Dharmika Dharmapala, Nuno Garoupa & Richard H. McAdams, "Belief in a Just World, Blaming the Victim, and Hate Crime Statutes " (University of Chicago Public Law & Legal Theory Working Paper No. 242, 2008).

This Working Paper is brought to you for free and open access by the Working Papers at Chicago Unbound. It has been accepted for inclusion in Public Law and Legal Theory Working Papers by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

CHICAGO

JOHN M. OLIN LAW & ECONOMICS WORKING PAPER NO. 438
(2D SERIES)

PUBLIC LAW AND LEGAL THEORY WORKING PAPER NO. 242



BELIEF IN A JUST WORLD, BLAMING THE VICTIM, AND HATE CRIME STATUTES

Dhammika Dharamapala, Nuno Garoupa, and Richard H. McAdams

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

October 2008

This paper can be downloaded without charge at the John M. Olin Program in Law and Economics Working Paper Series: <http://www.law.uchicago.edu/Lawecon/index.html> and at the Public Law and Legal Theory Working Paper Series: <http://www.law.uchicago.edu/academics/publiclaw/index.html> and The Social Science Research Network Electronic Paper Collection.

Belief in a Just World, Blaming the Victim, and Hate Crime Statutes*

Dharmika Dharmapala[†]
University of Connecticut

Nuno Garoupa[‡]
University of Illinois

Richard McAdams[§]
University of Chicago

August 2008

*We thank one anonymous referee, the editor Francesco Parisi, Albert Choi, Dominique Demougin, Lee Fennell, Louis Kaplow, Friedrich Kubler, Henrik Lando, William Landes, Anup Malani, Thomas Miles, Janice Nadler, Eric Posner, Jeff Rachlinski, Jennifer Robbenolt, Chris Sanchirico, Katherine Strandburg, Lu-in Wang and participants at the 2005 American Law and Economics Association meetings and the 2006 Conference on Uncertainty, Risk and Regulation: The Behavioral Law and Economics Perspective, at Technische Universität, Berlin for valuable comments. The usual disclaimer applies. Garoupa acknowledges the financial support of the FCT, PPCDT/JUR/55752/2006, Portugal.

[†]Department of Economics, University of Connecticut, 341 Mansfield Rd., U-1063, Storrs, CT 06269-1063, USA. Email: dharmika.dharmapala@uconn.edu

[‡]University of Illinois College of Law, 504 East Pennsylvania Avenue, Champaign, IL 61820, USA. Email: ngaroupa@law.uiuc.edu

[§]University of Chicago Law School, 1111 East 60th St., Chicago, IL 60637, USA. Email: rmcadams@uchicago.edu

Abstract

The earliest economic theory of discrimination proposed the subsequently neglected idea of a “vicious circle” of discrimination (Myrdal, 1944). We draw on psychological evidence (that people derive utility from believing that the world is just) to propose a behavioral economic model in which the vicious circle envisaged by Myrdal can arise. We demonstrate the power of this approach through an application to the issue of whether and how to justify penalty enhancements for hate crimes against members of disfavored groups. The crucial assumption is that individuals engage in biased inference in order to preserve their Belief in a Just World, thus attributing the disproportionate victimization of a group to that group’s negative characteristics, rather than to the hate-motivated preferences of offenders. In a simple two-period setting, we show that disproportionate victimization of the disfavored group in the first period can lead to additional crime against that group in the second period. The reason is that potential offenders’ inferences about the victimized group’s characteristics become more negative as a consequence of disproportionate victimization, raising the net benefits of crime against that group (under the assumption that the benefits of crime depend partly on the victimized group’s perceived characteristics). Our main result is that penalty enhancements can reduce the social harm due to these extra crimes.

JEL Classification: K4.

Keywords: hate crimes, behavioral economics.

1 Introduction

Consider these incidents:

Example 1: High school students routinely assault and abuse a small clique of punk rockers. As a consequence, there is a stronger feeling in the local community that people hanging around town dressed as punks are likely to harm the community. With time, more individuals (some not even high school students) engage in assaulting punks.

Example 2: Homeless women and men are an easy target for youth gangs. Most people think that some bad behavior in the past explains the current situation of homelessness. Given their social marginalization and uneasy relationship with law enforcement agents, violence against the homeless seems to be tolerated by the general population in the area and to have spread after the initial attacks by youth gangs.

Example 3: Hate crimes in a community cause an increase in the number of assaults against its black members. Non-hater whites observe this fact and infer that the blacks have done something to deserve these assaults. Thus, whites explain the “extra” assaults as arising from the fact that blacks in the community conform to some whites’ negative stereotypes of them, as by being lazy, immoral, or prone to criminality. As a result, the perceived shame or expected guilt from victimizing blacks falls and they suffer from more assaults, robbery, and car theft.

The first two of these examples are loosely inspired by the actual incidents described in Blake (2001), while the last is a hypothetical. All three, however, share a common pattern in which an identifiable group is subject to victimization for which the group is partly blamed, which then leads to more negative perceptions of the group. These negative perceptions, in turn, lead to increased victimization. This “vicious circle” has much in common with the theory of discrimination proposed by Myrdal (1944). According to his account, an exogenous increase in discrimination against a disfavored group leads to worse outcomes for that group. Members of the dominant group then interpret these bad outcomes as evidence of the disfavored group’s intrinsic negative characteristics, which leads to more discrimination. This in turn causes even worse outcomes and more negative inferences about the disfavored group’s characteristics.

The victim-blaming in Myrdal’s vicious circle is puzzling from the perspective of standard economic models in which agents engage in rational Bayesian inference. In particular, it requires that members of the dominant group naively ignore the effects of discrimination on the disfavored group’s outcomes.

In this paper, we propose a behavioral economic model in which the vicious circle envisaged by Myrdal and the above examples can arise. We assume the existence of racial animus (Becker, 1957) and then demonstrate the significance of victim-blaming through an application to the analysis of hate crime statutes. The basis for the model lies in a psychological literature, pioneered by Lerner (1965), showing that most individuals feel a strong need to believe that they live in a world that is just, in the sense that people generally get what they deserve and deserve what they get (Bénabou and Tirole, 2006, p.700). Although these psychological experiments are potentially subject to multiple interpretations, we argue in Section 2.2. below that the phenomenon is sufficiently well-studied that an exploration of its consequences for the law is warranted. We also argue below that the Belief in a Just World can reasonably be interpreted as a bias, which we term the “just world bias” (JWB; the term “Belief in a Just World” (BJW) is used below to denote the broader concept).

One interpretation of this belief is that it represents a form of self-deception that is sustained in equilibrium because there is a demand for and supply of motivated beliefs (Bénabou and Tirole, 2006). The demand arises because believing that the world is more “just” than it actually is works as a form of self-commitment that overcomes weakness of will by motivating more effort.¹

¹Bénabou and Tirole (2006) also offer a “classical” intergenerational interpretation of the just world belief that involves no irrationality. Parents tend to have different preferences from their children, especially in having a lower discount rate for the future. Parents therefore frequently desire that their children put forth greater effort and obey rules to a greater degree than is consistent with their children’s discount rates. One means of motivating additional effort and obedience is for parents to indoctrinate children to believe that the world is more just than it actually is, so that children will believe that in the long term their outcomes depend more on effort and obedience, and less on good or bad “luck” than is actually the case. The cliché that “cheaters never prosper” probably better motivates effort and obedience than would the more accurate statement that “cheaters prosper only in certain narrow circumstances.” On this view, adults then engage in Bayesian inferences, but start out with priors from childhood weighted strongly towards a belief in a just world.

If people believe the world is more just than it actually is, then they will generally infer that bad outcomes are more “deserved” by those who suffer them than is actually the case, meaning that bad outcomes are more the result of the victims negative characteristics or misbehavior than is actually the case. If so, then when discrimination causes its victims to suffer bad outcomes, observers will over-attribute the outcome to the victim’s negative characteristics and under-attribute the contribution of discrimination. This skewed belief preserves the belief in a just world. Stated differently, given an exaggerated prior about the degree to which outcomes are deserved, it is rational to infer that bad outcomes are more deserved than they actually are. Believing they are deserved can then make rational additional forms of discrimination. Although there are many potential applications, in our example, hate crimes cause more negative beliefs about the characteristics of the hate crime victims, which in turn leads to more (non-hate) crime.

Hate-motivated crimes have attracted the scholarly attention of a large legal and philosophical literature (e.g. Lawrence, 1999; Hurd and Moore, 2004) and a small but growing economic literature (Dharmapala and Garoupa, 2004; Dharmapala and McAdams, 2005; Gan, Williams and Wiseman, 2004). The central issue in this analysis is whether and how to justify penalty enhancements for hate crimes. In this paper, we derive a Myrdalian vicious circle from the Belief in a Just World, an approach that can provide a justification for hate crime penalty enhancements that does not rely on fairness considerations (as in Harel and Parchomovsky, 1999), and that applies more broadly than do the efficiency arguments in Dharmapala and Garoupa (2004).² We reach this result by extending the model of Dharmapala and

²Dharmapala and Garoupa (2004) extend the economic model of optimal enforcement to the case where the population is divided into dominant and disfavored groups (defined e.g. by ethnicity or religion). A subset of potential offenders from the dominant group is assumed to harbor racial animus (cf. Becker, 1957) in the form of discriminatory or hateful preferences, so that they derive greater benefits from committing an otherwise identical crime against members of the disfavored group than from committing the same crime against a member of their own group. If expected sanctions are independent of the victim’s group, this hate motivation gives rise to an equilibrium in which members of the disfavored group face a disproportionately high probability of victimization. A central lesson of Dharmapala and Garoupa (2004), however, is that penalty enhancements may or may not be optimal on efficiency grounds, depending on factors such as the distribution of benefits from crime and the social costs of sanctions. It appears that, within a consequentialist (economic or utilitarian) framework, there is no harm caused by disparate victimization

Garoupa (2004) to include not only the Belief in a Just World (BJW), but also a new variable representing the moralistic beliefs that a potential offender holds about the characteristics of a potential victim. Specifically, these beliefs concern the perceived intrinsic value or “moral worth” of the individual one may victimize. We assume the beliefs affect the potential offender’s costs (net benefits) of crime by affecting the expected guilt or shame of the criminal act. In particular, members of the dominant group may hold negative beliefs about the characteristics of members of the disfavored group, e.g., their dishonesty, immorality, selfishness, laziness, or lack of patriotism. The more negative these characteristics are believed to be, the less are the expected costs from committing a crime against a member of the disfavored group.

We use this framework to show that, absent the BJW, penalty enhancements do not influence Bayesian inferences about group characteristics (because the penalty enhancements themselves will convey as much information about such characteristics as will disproportionate victimization). Then, we analyze how penalty enhancements may influence inferences about victims’ characteristics in a setting where potential offenders tend to overestimate the degree to which people deserve the outcomes they receive. The general intuition underlying the model is as follows: hate crimes cause negative outcomes for its victims; those with a strong prior Belief in a Just World infer that these outcomes are caused by the victims’ negative characteristics; these inferences lower the expected cost of crime against such victims and therefore cause more crime.

More specifically, our model demonstrates that, in a setting with no penalty enhancements, both haters and nonhaters in the dominant group commit crimes against the disfavored group, which suffers from disproportionate victimization. We assume that a new cohort of nonhaters within the dominant group observe the rate of victimization and make inferences about the number of crimes attributable to haters preferences and the number attributable to non-haters’ beliefs about the victimized group’s negative characteristics. Because not all crimes are solved, those inferences are made in the face of uncertainty. Attributing an unsolved crime to a hater poses a greater conflict with a Belief in a Just World than attributing an unsolved crime to the vic-

per se.

tim’s negative characteristics.³ Observers with an excessively strong Belief in the Just World will therefore attribute fewer of the unsolved crimes to haters and more to the victim’s negative characteristics than would someone with a more accurate view of the world’s justness. Failing to account for the full extent of discriminatory preferences in the face of uncertainty, the BJW causes such observers to revise their beliefs about the victim’s characteristics in a negative direction. These revised beliefs in turn raise the net benefits from crimes against that group, which causes additional crimes against that group.⁴ Our central result is that penalty enhancements for hate crimes can reduce the social harms associated with these additional crimes.

The paper proceeds as follows. Section 2 provides a background on hate crimes and the just world bias. Then, the basic model is presented in Section 3, while Section 4 discusses the implications of our model, and concludes.

2 Background on Hate Crimes and the Just World Belief

2.1 Hate Crimes

In recent years, the Federal Bureau of Investigation (2005) has tabulated reports of “hate crimes” involving 9,000-12,000 victims per year. Most of

³Psychological evidence suggests that most people view crime victimization caused by hate motivation as more unjust than victimization caused by negative characteristics of the victims (Rayburn, Mendoza and Davison, 2003).

⁴It is important to emphasize that this mechanism is driven by uncertainty about the motivations of offenders. In particular, we are *not* claiming that crimes that are known to be hate-motivated will lead to negative inferences about victims, e.g. through an inference that the victims must have done something to deserve such hatred. To the contrary, if one observes what is known to be a hate-motivated crime, there is no reason to change one’s views about the characteristics of the victim; the haters’ discriminatory preferences constitute a sufficient explanation for their crimes. Rather, the effect we explore arises when crimes are unsolved and the motive is ambiguous. In this context, the strong prior Belief in a Just World causes observers to attribute a larger proportion of unsolved crimes to nonhaters than would one with a weaker prior. The central role of uncertainty has several implications, including the idea that minority groups will seek to publicize the hate preferences that motivate many of the crimes they suffer.

these crimes are based on race. Some (such as the murderous rampages described in Dharmapala and McAdams, 2005) are severe, but most are less serious, involving crimes such as vandalism, intimidation, or simple assault (Federal Bureau of Investigation, 2005). Because many crimes go unreported, victim surveys report higher numbers. The National Crime Victimization Survey shows an annual average of 210,000 hate crime victims in the United States from 2000 to 2003 (Harlow, 2005).

Most US jurisdictions define a crime as a hate crime if it is committed because of the perpetrator’s animus or hatred toward a racial or other specified group, or, more broadly, if the victim is selected because of membership in a specified group.⁵ For example, Federal sentencing guidelines provide for a “penalty enhancement” if the defendant “intentionally selected any victim or any property as the object of the offense of conviction because of the actual or perceived race, color, religion, national origin, ethnicity, gender, disability, or sexual orientation of any person.”⁶ Most states have similar laws for offenses, though the list of possible groups varies widely (Grattet, Jenness and Curry, 1998). As a result, a hate crime may be punished more harshly than the same crime absent the hatred or discriminatory motive the perpetrator had for committing the offense.⁷

2.2 Belief in a Just World

The excessive Belief in a Just World has much in common with other psychological phenomena economists have explored. For example, if one is to be overconfident about one’s outcomes, e.g., Jolls, Sunstein and Thaler (1998), then one must not only have excessive optimism about one’s own abilities, but also excessive optimism about the way the world works, that it tends to reward deserving people like oneself. Similarly, the literature on the fundamental attribution error (surveyed in Dharmapala and McAdams,

⁵The latter type of statute has been upheld by the US Supreme Court, see *Wisconsin v. Mitchell*, 508 U.S. 476 (1993).

⁶28 U.S.C. § 994 (1994).

⁷Similarly, in the UK, although “hate crime” has no specific legal meaning, hate motivation is an aggravating factor in sentencing since the Criminal and Disorder Act 1998 (CDA 1998) and incitement to racial hatred has been extended to include religious grounds by the Anti-Terrorism, Crime and Security Act 2001 (ATCSA 2001).

2005) demonstrates that people tend to over-attribute outcomes to individuals (their traits and behaviors) and to under-attribute outcomes to the constraints the individual faced. If so, then we would expect individuals to attribute some excessive blame to those who suffer bad outcomes. As explained above, one can follow Bénabou and Tirole (2006) in interpreting these skewed inferences as a product of rational but skewed intergenerational transmission of information (where parents with lower discount rates than their children indoctrinate their children in BJW to motivate them to work harder and be more obedient). But one might also follow the literature on cognitive dissonance (e.g., Akerlof and Dickens, 1982) and “motivated belief” (e.g., Bénabou and Tirole, 2002) and interpret the BJW as a means of self-deception to relieve anxiety about the arbitrariness of the world or to self-commit in the face of one’s imperfect willpower.

The evidence that people tend to believe excessively in a just world come mostly though not entirely from laboratory experiments.⁸ In the pioneering study of Lerner and Simmons (1966), subjects viewed on a television what appeared to be a contemporaneous experiment on learning, in which a subject (actually a confederate of the experimenters) received painful electric shocks for giving incorrect answers. After ten minutes, the experimenters asked subjects to evaluate this “victim.” Before making their evaluation, however, the experimenters told the subjects either (1) that they would thereafter watch the same person in another ten minute session of the same experiment (the midpoint condition) or (2) that they would thereafter anonymously vote on whether the person would continue with the negative reinforcement experiment with electric shocks or be moved to a positive reinforcement experiment with monetary rewards (the reward condition). In the latter reward condition, the result of the vote (which was always to move the victim into the reward scenario) was announced before the subjects evaluated her. The main result was that subjects evaluated the victim significantly more negatively in the midpoint condition than the reward condition. Lerner and Simmons inferred that the midpoint condition was more threatening to the subjects sense of justice than the reward condition, because only in the latter could the subjects restore justice by ending the suffering and rewarding the victim for past suffering. Without that power to correct injustice, the subjects ad-

⁸For more extensive reviews, see Lerner and Miller (1978), Lerner (1980), Maes (1998), Ross and Miller (2003), and Dharmapala, Garoupa and McAdams (2006).

justed their views of the victim negatively to make her seem more deserving of her bad outcome.⁹

Psychologists have found similar effects in experiments using a variety of more typical victims, from those suffering poverty in developing nations (Reichle and Schmitt, 1998), to cancer victims (Maes, 1994), and, most relevant for our purposes, crime victims (Wyer et al., 1985; see also Lincoln and Levinger, 1972). For example, in Wyer et al. (1985), the experimenters had subjects read vignettes about a rape after previously “priming” them with exposure to photographs of various sorts. When the priming photographs depicted aggressive acts - other apparent crimes - that challenge the idea of a just world, the experimenters found that the subjects were significantly more likely to evaluate the putative rape victim negatively and to hold her responsible for the rape (Wyer et al., 1985: 330-32, 337). Thus, the recent perception of a threat to the Belief in a Just World (BJW) causes greater adjustments of beliefs to preserve the BJW.

Although the literature is extensive, it is unfortunate that none of these experiments uses the distinctive methods of experimental economics, which test behavior in settings where subjects are provided with monetary incentives to act rationally. One might, for example, expose subjects to an individual’s victimization and then place them in a situation in which they have the opportunity to trade with the victim. We hope that such economic experiments will be carried out in the future (and that this paper highlights some of the important policy applications that would be at issue in such an exercise).

Nonetheless, the voluminous psychological literature is at least suggestive of the fact that people overestimate the justness of the world. There are other forms of evidence that support the point. For example, Hafer (2000a) avoided relying on self-reported beliefs by employing a “modified Stroop task”¹⁰ to

⁹One implication of this experiment is that, if observers are able to intervene in ways that help the victim, the negative inferences about the victim’s characteristics may be ameliorated. However, in the application to hate crimes in this paper, it appears reasonable to assume that most individuals will not be in a position to help crime victims they do not personally know or encounter. It is worth noting, nonetheless, that the existence of organizations that, for instance, accept donations to help crime victims may work to counter negative inferences about victims.

¹⁰A Stroop effect is a delay in performing an identification task; other research shows that the delay reveals psychological stress. Typically, subjects are asked to identify the color of a word flashed briefly on a screen. “[P]eople take longer to identify the color

measure the (possibly unconscious) effects of exposure to perceived injustice (a criminal who escapes punishment) on individuals ability to perform certain timed tasks. She found that perceived injustice caused a unique decline in performance and that those subjects whose timed performance suffered more were more likely to derogate the victim. In a follow-up study, this tendency to derogate victims was more pronounced among those whose survey results indicated a stronger BJW (Hafer, 2002). One famous study (Rubin and Peplau, 1973) stepped outside the laboratory and found that, in a gathering of draft-eligible men for a draft lottery in 1971, individuals with high BJW scores were more likely than others to evaluate lottery winners ahead of lottery losers, despite knowing that the lottery selected people at random. Similarly, Bénabou and Tirole (2006: 704) explain that sociologists and political scientists have conducted hundreds of detailed interviews of working-and middle-class individuals on economic concerns and found that, “much like a religion,” people “obstinately hold on to a belief that effort, hard work, good deeds will ultimately pay off,” even when evidence in their own life provides strong evidence to the contrary.

As real world evidence of a strong desire to believe in a just world, one might consider the success of Hollywood movies in which good characters are rewarded and bad characters punished. Indeed, psychologists have found that adults are less likely even to label a narrative a “story” if it contradicts the just world account (Brewer, 1996). As real world evidence of blaming victims, surveys of Americans during World War II suggest that they became more likely over this time to view Jews as wielding too much power in the United States. “Far from evoking sympathy, the Nazi persecutions apparently sparked a rise in anti-Semitism in this country.” (Selznick and Steinberg, 1969: 63).

Supporting Bénabou and Tirole’s intergenerational interpretation of their model, the BJW is particularly strong in children. As noted by Piaget (1965:260): “A great many children think that a fall or a cut constitute punishment because their parents have said to them, It serves you right, or That will be punishment for you, or God made it happen!” This fact should come as no surprise given how frequently children’s stories and myths (e.g., Cinderella, Pinocchio, and Santa Claus) “emphasize the rewards that fol-

of words that are associated with whatever is of emotional concern (i.e., whatever is threatening) than the color of neutral words.” (Hafer, 2000a: 166).

low from virtue and the punishments that follow from misbehavior.” Rubin and Peplau (1975: 72). Indeed, some evidence links children’s ability to delay gratification with their expectation of fairness or justice; see Long and Lerner (1974) and Mischel (1974). Adults outgrow the crudest form of the heuristic (Lerner, 1998: 267), but the rule of thumb continues to influence the process of inference. Indeed, the bias may remain functional in adults by increasing the perceived value of long-term planning (Hafer, 2000b), and so may have advantages that offset (or render imperceptible) its disadvantages. Hafer (2000b: 1069) surveyed subjects to determine their general Belief in a Just World and their long-term investment orientation. Across various measures, she found a significant correlation: “the more participants had a chronic long-term focus, the stronger their BJW.”¹¹

Other evidence supports the motivated bias interpretation of the Bénabou and Tirole model. Lerner (1998) argues that individuals derive utility from believing in a just world, or suffer increasing anxiety and stress from believing the world is increasingly unjust. People are therefore willing to trade off the utility of BJW against the informational benefits from unbiased inferences about the world. There is evidence supporting the benefit of some trade-off, that a BJW correlates with stronger marital relationships (Lipkus and Bissonnette, 1996) and in coping with bereavement (e.g. Bonanno et al., 2002).

In any event, in one important respect, our model goes beyond the BJW as it is usually discussed in psychology. We focus on inferences about groups, although the BJW is typically discussed in terms of inferences about individuals. However, we believe that our extension is reasonable. In explaining how people maintain the BJW, Lerner (1998) says that “the person who derogates a victim will generate a culturally plausible basis for that condemnation.” The perceived negative characteristics of a disfavored group to which the victim belongs naturally provides such a basis. Indeed, there is evidence

¹¹Hafer “primed” her subjects by having each write an essay about his or her life, focusing either on future long-term goals (long-term focus condition) or current activities (no-long-term focus condition). She then had the subjects observe a video of a student talking to a counselor about the depression she began suffering after acquiring a sexually transmitted disease (when a condom broke during intercourse). When asked to evaluate the victim, subjects in the long-term focus condition derogated the victim significantly more than those in the no-long-term focus condition.

suggesting that people are less likely to make negative inferences about individuals who are similar to them in some salient respect (e.g. Lerner and Agar, 1972). If differences between the individual and victim are sufficient to heighten the individual’s willingness to blame the victim, then we believe it is a reasonable extension to posit that the individual will use such differences to “explain” his or her derogation of the victim.¹² The above example (rising anti-Semitism in the United States during World War II) provides support for our reasoning.

The relevance of the just world belief and related psychological phenomena for the study of hate crimes has been recognized previously by Wang (1997; 1999). Her argument, however, is primarily that victims of hate crimes suffer greater psychological harms than victims of parallel non-hate crimes, because the former victims’ Belief in a Just World is more seriously impaired. She also argues (1997: 129), as we do, that hate crimes will “promote [...] prejudice against the [victim’s] group” even among those who do not hate the group. However, we specify a precise mechanism through which this effect operates, and link the increased prejudice to increased future crime against the victimized group. In examining the effects of disproportionate victimization on the beliefs and behavior of potential offenders, our approach here is closely related to the approach of Dharmapala and McAdams (2005), which identifies the costs of hate speech by focusing on the behavior of potential offenders, rather than on the psychic harms suffered by victims.

2.3 Public Attitudes about Hate Crimes

In our model, hate crimes increase the total crimes committed against certain groups and the greater number of crimes cause non-haters to make more

¹²We also believe we are justified in ignoring the BJW in modeling the decision to offend outside the group context, even though the negative inferences apply to all crime victims. If an individual is randomly victimized, others may make negative inferences about her characteristics, and this may raise the benefits to future crimes against her. However, the probability that those who may commit those future crimes will (a) know of this particular individual’s victimization and (b) encounter this individual again in a setting where a crime may be committed, is negligible. In contrast, if negative inferences are made about the characteristics of an entire disfavored group, the probability of future criminals encountering a member of that group is relatively high. Thus, the increased crime effect could be argued to apply only to the latter case.

negative inferences about the crime victims. Our claim is that non-haters fail to attribute the correct number of these “extra” crimes to the hate preferences of certain perpetrators and instead over-attribute extra crimes to the negative attributes of the victims. Our claim rests on an empirical assertion: that non-haters view hate crimes as more unjust than the same crimes committed without hate motivation. If this were not the case, then the BJW would have no effect on the inferences made about crime rates. If otherwise equivalent hate and non-hate crimes were regarded as equally unjust, then there would be no reason to skew inferences in order to preserve one’s belief in a just world.

There is, however, psychological evidence that most people view crime victimization caused by hate motivation as more unjust than victimization caused by negative characteristics of the victims. Rayburn, Mendoza and Davison (2003 p. 1063) find that subjects exposed to a hate crime scenario view the perpetrator as being more culpable than the perpetrator in an otherwise identical non-hate crime. Similarly, even though subjects blame all crime victims to some degree (p. 1069), they rate the victim of a hate crime as less culpable than the victim of a non-hate crime (pp. 1062-63).¹³ Moreover, the study supports the importance of the inferences made in the face of crime: if observers know for certain that a crime is the product of discriminatory preferences, as the scenarios in the experiment expressly made clear, they blame the victim less. But because individuals do not know for certain the cause of most crimes (particularly if the law does not force the revelation of hate motives), the bias has scope to operate, allowing them to attribute the crime to the victim’s negative characteristics and to blame the victim.

Note that this point also explains why hate crime penalty enhancements are often politically feasible. The public will tend to support hate crime enhancements because citizens view the hate motivation as making the perpetrator more culpable and therefore more deserving of punishment. Given the JWB, the public will of course underestimate how commonly the enhance-

¹³Even so, the study does not prove that the subset of the dominant group that commits hate crimes shares this conception of justice. Thus, our model assumes that only those individuals in the dominant group who are not hate-motivated view disparate victimization as unjust (and are therefore subject to the just world bias). This assumption does not mean that haters are more “rational” in making inferences, but simply emphasizes that the results do not depend on haters viewing disparate victimization as unjust.

ments will be used, but they will not necessarily object to their enactment.¹⁴

3 Model

3.1 Basic Assumptions

Our model is based on the standard economic theory of law enforcement (as reviewed in Garoupa, 1997 and Polinsky and Shavell, 2000). We extend this setup to a two-period framework, where the set of potential offenders in period one is replaced by a new “generation” or cohort of potential offenders who enter the model in period two. Each of these (risk-neutral) potential offenders has the opportunity to commit up to one crime against each of two groups of equal size¹⁵ – the X ’s and the Y ’s. For simplicity, we assume that all potential offenders in both periods are X ’s, the same approach as in Dharmapala and Garoupa (2004).¹⁶ The illegal gains from committing the crime are represented by the variable b (as is conventional in the economic theory of enforcement, each individual’s b is unobservable, although the distribution is known). The distribution of offenders (X ’s) over b is given by the cumulative distribution function (cdf) $F(b)$ (note that $F(b)$ is the fraction of X ’s who receive benefits less than b from committing the crime).

Let the probability of detection of a crime against a member of group j be denoted by p_j , where $j = X, Y$. Assume also that this probability does not change over time. The sanction imposed in period t for a crime against a

¹⁴Similarly, given that the finding that individuals tend to blame crime victims less when the crime was hate motivated, there is no reason to believe that the public will make negative inferences about groups from the mere existence of hate crime legislation. Also, unlike affirmative action, which some may perceive as providing a “special privilege” for certain groups, hate crime penalties are neutral regarding the relevant criteria, e.g., race or religion, because one can incur an enhancement for crimes against members of the majority race or religion as well as against minority races or religions. Indeed, much of the actual enforcement of these laws involves the prosecution of minorities for crimes against majorities.

¹⁵Different group sizes would have a scale effect, but would make no qualitative difference to the results of the model.

¹⁶Introducing crimes committed by Y ’s would complicate the model, but would not affect the basic results.

member of group j is denoted s_i^j . The complete set of sanctions is denoted by $\{s_1^X, s_1^Y, s_2^X, s_2^Y\}$.

3.2 Rational Learning

In period 1, suppose that p_X and p_Y are both known to all potential offenders. Then, (risk-neutral) potential offenders commit the crime against X 's if $b \geq p_X s_1^X$. Bearing in mind that the groups are of equal size, the rate of victimization of X 's is:

$$r_{X1} = 1 - F(p_X s_1^X) \quad (1)$$

Potential offenders commit the crime against Y 's if $b \geq p_Y s_1^Y$. The fraction of X 's who commit the crime against Y 's is thus $(1 - F(p_Y s_1^Y))$ and the rate of victimization of Y 's is:

$$r_{Y1} = 1 - F(p_Y s_1^Y) \quad (2)$$

Suppose that there is no penalty enhancement, and sanctions are uniform (i.e. $s_1^X = s_1^Y$), but that Y 's are more "vulnerable," in the sense that $p_Y < p_X$ (for example, the police are less likely to investigate crimes against Y 's, or Y 's are less likely to report crimes to the authorities). Then, it follows straightforwardly that disparate victimization occurs (i.e. $r_{Y1} > r_{X1}$).

In period 2, suppose that a new cohort of potential offenders knows p_X but not p_Y . Potential offenders commit the crime against X 's if $b \geq p_X s_2^X$, so that the rate of victimization of X 's is:

$$r_{X2} = 1 - F(p_X s_2^X) \quad (3)$$

Let \widehat{p}_Y be the probability of detection inferred by the new cohort of potential offenders (who can observe the sanctions and rates of victimization that occurred in period 1). The crime level against Y 's depends on the inferred \widehat{p}_Y ; from Equation (2), this inference problem can be implicitly characterized by:

$$r_{Y1} = 1 - F(\widehat{p}_Y s_1^Y) \quad (4)$$

This straightforwardly implies that $\widehat{p}_Y = p_Y$. Thus, potential offenders commit the crime against Y 's if $b \geq p_Y s_2^Y$, and the rate of victimization is:

$$r_{Y2} = 1 - F(p_Y s_2^Y) \quad (5)$$

In this setup, rational learning implies that a higher rate of victimization of Y 's in period 1 will lead *ceteris paribus* to a lower inferred \widehat{p}_Y and hence to a higher period-2 rate of victimization of Y 's. That is, crime in period 1 generates additional crime in period 2. However, penalty enhancement in period 1 (i.e. setting $s_1^X < s_1^Y$) cannot address this problem, as long as period-1 sanctions are observable to the new period-2 cohort.¹⁷ For instance, suppose s_1^Y were increased, holding everything else fixed. Then, r_{Y1} would fall, but the inference problem (Eq. 4 above) would lead to the same inference $\widehat{p}_Y = p_Y$. In essence, period-2 offenders realize that the reason the victimization rate for Y 's is low is because of the penalty enhancement, and they incorporate this into their inference. It follows that penalty enhancements for hate crimes cannot be explained by a dynamic rational learning setup of this kind (although penalty enhancements for vulnerable groups may still be optimal for purely static reasons - i.e. to reduce the harms from crimes in period 1).

3.3 Introducing Hate Motivation

Now, assume that members of group X can be partitioned into two (exhaustive and mutually exclusive) subgroups: those with hate motivation (in a sense defined more precisely below), denoted by X^H , and those without hate motivation, denoted by X^N .¹⁸ In each period, the former constitutes a fraction $\alpha \in (0, 1)$ of all X 's, and the latter a fraction $(1 - \alpha)$ (these proportions are assumed to be common knowledge). The illegal gains from committing the crime, represented by the variable b , differ across the two subsets of X . The distribution of non-hate-motivated offenders (X^N 's) over b is given by the cumulative distribution function (cdf) $F_N(b)$, independently of whether the victim of the crime is an X or a Y (note that $F_N(b)$ is the fraction of X^N 's who receive benefits less than b from committing the crime). The distribution of hate-motivated offenders (X^H 's) over the benefits from crimes against other X 's is also given by $F_N(b)$. However, the distribution

¹⁷This can be generalized to the case where sanctions are not observable, as long as no systematic errors are made by potential offenders in inferring the sanctions.

¹⁸For the sake of additional realism we introduce the assumption that not all X 's are haters. However we also assume that haters do not think that hate crimes are unjust; our results would only be strengthened if we assumed otherwise.

of X^H 's over the benefits from crimes against Y 's is given by the cdf $F_H(b)$. The notion of hate motivation is captured by assuming that X^H 's derive greater benefits, *ceteris paribus*, from crimes against Y 's. Specifically, following Dharmapala and Garoupa (2004: 190), we assume that for a given b in the relevant range (in particular, high enough values of b such that the crime may be committed):

Assumption 1a: $F_N(b) > F_H(b)$,

i.e., the fraction of X^H 's who receive benefits less than b from crimes against X 's is greater than the corresponding fraction for crimes against Y 's. We also make the following assumptions about the probability density functions (pdf's):

Assumption 1b: For low values of b , $f_N(b) > f_H(b)$; for high values of b , $f_N(b) < f_H(b)$.

To ensure that the cdf's $F_H(b)$ and $F_N(b)$ are strictly monotonically increasing, it is also assumed that:

Assumption 2: For all b in the relevant range, $f_j(b) > 0$ for $j = N, H$.

Assumption 1a implies that (in the absence of penalty enhancements) Y 's will suffer disproportionate victimization as a result of the discriminatory preferences of X^H 's. This paper introduces another distinct source of disparate victimization - negative beliefs held by X^N 's about the characteristics of Y 's. These may induce X^N 's to disproportionately target Y 's, even in the absence of any intrinsic hate motivation. We extend the standard framework by assuming that the net benefit from a crime committed by a member of X^N depends not only on b , but also on group Y 's perceived "characteristics" (denoted by $c \geq 0$).¹⁹ These characteristics are assumed to bear on the (perceived) moral inappropriateness of the individual being criminally victimized, and correspond to the amount of guilt and shame the perpetrator expects to feel. The higher is c , the worse the group's characteristics are perceived to be, so the *net* benefit from the crime to a potential offender can be characterized as $(b + c)$ (i.e. a larger c for a particular group implies a greater net benefit of attacking a member of that group).²⁰

¹⁹For simplicity, it is assumed that X^N 's only have beliefs about group Y 's characteristics.

²⁰It is intuitive to think of c as the variation in the guilt or shame from committing

For the sake of simplicity, this variable c is assumed to only influence the behavior of X^N 's, and not that of X^H 's. The latter's hatred of Y 's is presumed to be independent of these perceived characteristics (or alternatively, the perception of Y 's characteristics among X^H 's can be viewed as being implicitly incorporated into the specification of the $F_H(b)$ function, and to be fixed across periods).

Without loss of generality, we normalize c in period 1 to zero, so that X^N 's in period 1 make their decisions about whether to commit the crime on the basis that $c = 0$. It is assumed that this first-period value of the characteristics is not observed by the new cohort of X^N 's who enter the model in period 2. In order to make their crime decisions, these individuals must infer the value of c (more generally, they must update some prior about c). Our focus is on inferences about the victimized group's intrinsic characteristics. In practice, it is possible that the inference from disparate victimization may instead be that members of the targeted group take greater risks (such as carrying large amounts of cash). This amounts, in effect, to an inference that the probability of detection (and hence the expected sanction) is lower when attacking such groups. In our analysis, we assume that the probability of detection and the expected sanction are common knowledge, so no updating of beliefs about these variables takes place. Even if the inference from disparate victimization is that members of the disfavored group take greater risks, this will lead to a higher level of crime, and so operate in the same direction as the effect we identify. We denote this inferred value by \hat{c} .²¹

Define $p = p_X = p_Y < 1$ to be the probability of detection, which is the offense. However, since there is no formal cost term in the model, it is easier, and analytically equivalent, to place c on the benefit side. The motivation for this assumption is that potential offenders sometimes expect to incur guilt or shame from committing a crime, where the precise amount depends on the perceived characteristics of the victim. The offender expects to incur less psychological aversion (guilt) or social disapproval (shame) – i.e. less cost – to commit an offense against a person perceived to have negative characteristics (low moral worth) than against a person perceived to have positive characteristics (high moral worth). For example, defrauding a “liar” and assaulting a “bully” are less costly than committing the same crimes against a person without those negative characteristics. It would be still more costly to commit the crimes against one perceived to be the moral “pillar of the community.”

²¹This type of updating could also take place within a cohort. For simplicity, however, we assume that all members of a given cohort are symmetrically informed, and that updating only occurs across cohorts.

assumed to be time invariant and independent of the victim's group. The sanction for a crime against a member of group j in period t is denoted by s_t^j where $j = X, Y$ and $t = 1, 2$. In addition, it is assumed that hate motivation is perfectly observable *ex post* by courts, so the sanction for a crime against a Y can differ depending on whether or not the perpetrator was motivated by hate. Given the assumptions made above, all X^H 's who have sufficiently large b that they commit the crime in equilibrium derive greater benefits from targeting Y 's. In this sense, all crimes by X^H 's against Y 's that are detected are revealed to be hate crimes (and subject to penalty enhancements, if they exist). Specifically, if an X^H commits a crime against a Y in period 1 and it is detected, then the sanction is s_1^{YH} . The complete set of sanctions is denoted by $\{s_1^X, s_1^{YH}, s_1^{YN}, s_2^X, s_2^{YH}, s_2^{YN}\}$. Note that the penalty enhancement for a hate crime in period 1 (the policy variable on which the analysis focuses) is $(s_1^{YH} - s_1^{YN})$.²²

The harm to an individual victim of a crime is denoted by $h > 0$ (and is independent of the victim's group and the perpetrator's motivation). This assumption, consistent with Dharmapala and Garoupa (2004), entails that the private harm to an individual victim from a hate crime is identical to that from an equivalent non-hate crime, so that we can *endogenously* derive disparities in the *social* harms that result from a pattern of discriminatory selection of victims.

3.4 Outcomes with Bayesian Inference

In this subsection, we briefly characterize the outcomes when individuals engage in Bayesian inference. In period 1, (risk-neutral) potential offenders commit the crime against X 's if $b \geq ps_1^X$. Bearing in mind that the groups are of equal size, the rate of victimization of X 's is:

$$r_{X1} = 1 - F_N(ps_1^X) \quad (6)$$

X^H 's commit the crime against Y 's if $b \geq ps_1^{YH}$, and X^N 's commit the crime against Y 's if $b \geq ps_1^{YN}$. The fraction of X 's who commit the crime

²²This formulation is sufficiently general to include the case where penalty enhancements do not exist. The analysis below begins with the assumption of uniform sanctions ($s_1^X = s_1^{YH} = s_1^{YN}$ in period 1), before proceeding to analyze the consequences of penalty enhancements.

against Y 's is thus $\alpha(1 - F_N(ps_1^{YH})) + (1 - \alpha)(1 - F_N(ps_1^{YN}))$ and the rate of victimization of Y 's is:

$$r_{Y1} = 1 - \alpha F_H(ps_1^{YH}) - (1 - \alpha)F_N(ps_1^{YN}) \quad (7)$$

Suppose that there is no penalty enhancement, and sanctions are uniform (i.e. $s_1^X = s_1^{YH} = s_1^{YN}$). Then, by Assumption 1a, it follows that disparate victimization occurs (i.e. $r_{Y1} > r_{X1}$). Disparate victimization, under these assumptions, involves a higher rate of victimization being suffered by Y 's than by X 's. However, a more general (and more analytically relevant) interpretation of the disparity is that Y 's suffer a higher rate of victimization than they would in the absence of hate motivation among those who commit offenses against them. In this setup, the two interpretations are equivalent (because of the assumption that, but for hate motivation, X 's and Y 's would be victimized at the same rates, given the same penalties), but the latter generalizes more readily to cases where multiple groups may experience hate-motivated victimization.²³

In period 2, potential offenders commit the crime against X 's if $b \geq ps_2^X$, while X^H 's commit the crime against Y 's if $b \geq ps_2^{YH}$. The crime level of X^N 's depends on their inferred \hat{c} ; from Equation (2), this inference problem can be implicitly characterized by:

$$1 - r_{Y1} = \alpha F_H(ps_1^{YH}) + (1 - \alpha)F_N(ps_1^{YN} - \hat{c}) \quad (8)$$

This straightforwardly implies that $\hat{c} = 0$. Thus, X^N 's commit the crime against Y 's if $b \geq ps_2^{YN}$. The rates of victimization in period 2 are $r_{X2} = 1 - F_N(ps_2^X)$ and $r_{Y2} = 1 - \alpha F_H(ps_2^{YH}) - (1 - \alpha)(1 - F_N(ps_2^{YN}))$.

Note that Equation (8) implies that the new cohort of X^N 's in period 2 infers \hat{c} only on the basis of the observed behavior of the previous cohort of X^N 's, not on that of X^H 's. This assumption captures the intuition that, in learning about the characteristics of the outgroup, non-hate-motivated members of the dominant group will not be influenced by the behavior of those who are known to hate the outgroup. Since X^H 's are known to hate Y 's,

²³Also, the empirical evidence cited above (e.g. Rayburn, Mendoza and Davison, 2003) seems to suggest that hate-motivated victimization, rather than different victimization rates across different groups *per se* are seen as unjust; thus, the former would be more likely than the latter to trigger the just world bias.

their attacks are correctly interpreted as stemming from discriminatory preferences, and so do not convey any new information about the characteristics of Y 's.

3.5 Introducing Biased Inference

In the previous subsection, it was assumed that the new cohort of X^N 's in period 2 observes the first-period outcomes (in particular, the rates of victimization) and policies (in particular, the expected sanctions), and knows the preferences (i.e. the distributions $F_N(b)$ and $F_H(b)$). That is, these individuals are aware that X^H 's in period 1 have discriminatory preferences, and correctly attribute the higher victimization rate of Y 's to this hate motivation. However, this extra victimization of Y 's through no fault of their own may come into conflict with the desire of X^N 's to believe in a just world.

As noted above, Rayburn, Mendoza and Davison (2003) offer empirical support for this claim, finding that subjects view a hate crime perpetrator as being more culpable than the perpetrator in an otherwise identical non-hate crime. Similarly, even though subjects blame all crime victims to some degree (p. 1069), they rate the victim of a hate crime as less culpable than the victim of a non-hate crime (pp. 1062-63). Because individuals do not know for certain the cause of most crimes (particularly if the law does not force the revelation of hate motives), the bias has scope to operate, allowing them to attribute the crime to the victim's negative characteristics and to blame the victim. By underestimating the role of discriminatory preferences in the face of uncertainty, observers preserve their belief in the basic justness of the world.

In this subsection, we introduce the assumption that X^N 's engage in biased inference, thereby generating the just world bias (JWB) discussed above. In the first period, the outcomes are identical to those characterized in Section 3.4 above. In the second period, however, the JWB will affect the inferences, and hence behavior, of the new cohort of X^N 's. Specifically, we formulate the JWB as follows: X^N 's underestimate the extent of hate motivation on the part of X^H 's in period 1. An extreme way to capture this idea is to assume that the hate motivation is completely ignored, so that F_N is used

instead of F_H in inferring \hat{c} .²⁴ Intuitively, all crime against Y 's in period 1 is attributed to nonhaters. Then, the inference problem is:

$$1 - r_{Y1} = \alpha F_N(ps_1^{YH}) + (1 - \alpha)F_N(ps_1^{YN} - \hat{c}) \quad (9)$$

i.e.

$$\alpha F_H(ps_1^{YH}) + (1 - \alpha)F_N(ps_1^{YN}) = \alpha F_N(ps_1^{YH}) + (1 - \alpha)F_N(ps_1^{YN} - \hat{c}) \quad (10)$$

The solution to this inference problem, and the main results that follow, are presented in the following subsection.

3.6 Results

Given Equation (10) above, it follows that:

Remark 1: The assumptions above imply biased inference (i.e. the just world bias); in particular, $\hat{c} > 0$.

Proof: Rearranging Equation (10):

$$F_N(ps_1^{YN} - \hat{c}) - F_N(ps_1^{YN}) = \frac{\alpha}{1 - \alpha}(F_H(ps_1^{YH}) - F_N(ps_1^{YH})) \quad (11)$$

Suppose that $\hat{c} \leq 0$. Then, $F_N(ps_1^{YN} - \hat{c}) \geq F_N(ps_1^{YN})$ (as F_N is strictly monotonically increasing), so that:

$$\frac{\alpha}{1 - \alpha}(F_H(ps_1^{YH}) - F_N(ps_1^{YH})) \geq 0 \quad (12)$$

But, this contradicts Assumption 1a (that $F_N(b) > F_H(b)$). So, $\hat{c} > 0$.

Thus, beliefs about group Y 's characteristics are more negative in period 2 than in period 1, as a result of the just world bias and the disproportion-

²⁴Our assumption corresponds to a situation of complete *naiveté* in the model of Benabou and Tirole (2006). Similar results would hold even if the underestimation of the extent of hatred were only partial. Partial *naiveté* raises the issue of how much the update of beliefs might be manipulated by eliminating evidence, avoiding certain social interactions, reassuring information, repression and other mechanisms discussed by Benabou and Tirole (2006).

ate victimization suffered by Y 's in period 1.²⁵ As the Bayesian inference is $\hat{c} = 0$, the extent of this bias can be measured straightforwardly by the magnitude of \hat{c} . Importantly, this measure of the extent of the bias depends on the period-1 sanction imposed on hate crimes against Y 's. An important caveat concerns the observability of the crimes that occur in period 1. The analysis assumes that the crime rates suffered by each group (r_{X^1} and r_{Y^1}) are observable, and that courts can perfectly observe *ex post* whether crimes are hate-motivated or not (i.e. whether they are committed by X^H 's or X^N 's). Thus, if the outcomes of all trials were observable to period 2 X^N 's, they would be able to directly observe the number of hate crimes (committed by X^H 's) and the number of non-hate-motivated crimes (committed by X^N 's) that occurred in period 1. Under such circumstances, it would appear that the underestimation of hate motivation involved in the just world bias requires not just biased inference, but also that observers ignore available information that contradicts their inferences. This point would be reinforced if courts could observe not merely whether a perpetrator is an X^H or X^N , but also her b . This would enable courts to publicize these b 's (perhaps by calibrating the penalty enhancement to the degree of hate motivation). If all crimes were solved, this would make the true distribution $F_H(b)$ directly observable. However, as long as there are some unsolved crimes (i.e. $p < 1$, as seems likely in reality), the result in Proposition 1 will continue to hold, even when the perpetrators of the solved crimes are known.²⁶ For unsolved

²⁵These revised beliefs about the disfavored group's characteristics may potentially be costly to those who engage in biased inference. For instance, these beliefs may induce suboptimally low interaction or trade with the disfavored group. In the experiments cited above, this was not an issue (the person who was evaluated by the subjects was a stranger whom they would never meet again), but it may be important in real-world settings. We do not explicitly model these costs. The implicit underlying assumption is that the desire to believe in a just world is traded off against these costs. As long as some utility is derived from the belief in a just world, then the biased inference and increased crime that we identify would continue to hold. Note, however, that because beliefs represent a direct source of utility in this setup (as e.g. in Akerlof and Dickens, 1982), we cannot impose the usual equilibrium condition that beliefs are correct in equilibrium. These wider conceptual issues are not addressed here, but would be an interesting subject for future research.

²⁶The focus on unsolved crimes is intended to isolate the impact of the JWB per se by abstracting away from all other forms of irrationality. If observers were assumed to, for instance, wrongly attribute to nonhaters crimes known to have been committed by haters, then this would reinforce the effect of the JWB. However, this type of assumption is stronger than required for the argument here.

crimes, the identity of the perpetrator is unknown, and an observer subject to the just world bias can underattribute these crimes to X^H 's (and hence infer a positive \hat{c}) without contradicting any directly observable information. Thus, if the motivations behind at least some subset of crimes is unknown, then there is scope for the just world bias to operate (i.e. for observers to underestimate the role of hate motivation and to overestimate the role of the victimized group's negative characteristics).²⁷

Proposition 1 For marginal changes in s_1^{YH} , \hat{c} is decreasing in s_1^{YH} .

Proof: Bearing in mind that the inferred \hat{c} adjusts to maintain the equality in Equation (10), it can be expressed as an identity:

$$\alpha F_H(p s_1^{YH}) + (1 - \alpha) F_N(p s_1^{YN}) - \alpha F_N(p s_1^{YH}) - (1 - \alpha) F_N(p s_1^{YN} - \hat{c}) \equiv 0 \quad (13)$$

Using the implicit-function rule:

$$\frac{d\hat{c}}{ds_1^{YH}} = -\frac{\alpha p (f_H(p s_1^{YH}) - f_N(p s_1^{YH}))}{(1 - \alpha) f_N(p s_1^{YN} - \hat{c})} < 0 \quad (14)$$

by Assumption 1b.

Thus, the larger the sanction imposed on hate crimes against Y 's in period 1, the less pronounced is the tendency to attribute negative characteristics to Y 's. Note that as the argument in Proposition 1 involves holding all other variables fixed, a marginal increase in s_1^{YH} is equivalent to a marginal increase in the penalty enhancement ($s_1^{YH} - s_1^{YN}$).

²⁷Note why the normalization $c = 0$ is not important for the results. First, the way we define c , the larger is c the more negative the perceived characteristics. Assume observers start out with $c < 0$ (i.e. a belief that the outgroup has positive characteristics), then they will expect a very low crime rate against that group. The JWB (i.e. ignoring the existence of haters) will make them expect an even lower crime rate, and so make their surprise (and subsequent revision of c to something less negative or even positive) more pronounced. However, there is nothing fundamentally different here from the case where the prior is $c = 0$. Take now the opposite case, a prior of $c > 0$. This means that observers start with a prior belief that the outgroup has negative characteristics. Then, they would expect a very high level of crime against that group, and (when they observe some lower level of crime) would update their prior such that they end up with more favorable beliefs about the outgroup's characteristics. With the JWB, all of the crime is attributed to nonhaters, so the inferred c is going to be more positive (i.e. inferred characteristics more negative) than if only some of the crime were attributed to nonhaters (as would happen without the JWB).

Equation (10) suggests a straightforward intuition for why this result holds. Consider a given \hat{c} , and suppose that s_1^{YH} (or equivalently the penalty enhancement $(s_1^{YH} - s_1^{YN})$) is increased marginally. The observer (a period-2 member of X^N who is subject to the just world bias) expects this change to lead to a relatively small decrease in crimes against Y 's (specifically, a fall of $f_N(ps_1^{YH})$). However, because the observer has underestimated the extent of hatred among X^H 's, she is "surprised" by the size of the fall in crime against Y 's (given by $f_H(ps_1^{YH})$, where $f_H(ps_1^{YH}) > f_N(ps_1^{YH})$ by Assumption 1b). Intuitively, this surprise occurs because the upper tail of the distribution of X^H 's is thicker than the observer believes. The observed change in the level of crime in response to the increase in the penalty enhancement is thus inconsistent with the given value of \hat{c} . This forces the observer to revise her estimate of \hat{c} downwards (i.e. the characteristics of the outgroup are viewed as being less negative).

Proposition 1 establishes that the bias due to the belief in a just world can be reduced by increases in the penalty enhancement for hate crimes. The importance of this bias stems from its role in influencing behavior, and in particular the crime rates in period 2. It can be shown that the just world bias generates additional crime in period 2 (i.e. crimes that would not have been committed under the Bayesian inference that $\hat{c} = 0$). The additional harm (or decrease in social welfare) from these extra crimes is denoted by A , and can be characterized as:

$$A = h(1 - \alpha)(F_N(ps_2^{YN}) - F_N(ps_2^{YN} - \hat{c})) \quad (15)$$

It follows that:

Proposition 2: A is strictly positive when the just world bias exists (i.e. when $\hat{c} > 0$), and is strictly increasing in \hat{c} (the extent of the just world bias).

Proof: Straightforwardly, $\hat{c} > 0$ implies $A > 0$, and

$$\frac{\partial A}{\partial \hat{c}} = h(1 - \alpha)f_N(ps_2^{YN} - \hat{c}) > 0 \quad (16)$$

Consequently, there is some amount of crime against Y 's in period 2 that is directly attributable to the just world bias. In this sense, the just world bias leads to disproportionate victimization *per se* being associated with harms (extra crimes) that even a consequentialist would recognize. Moreover, combining Propositions 1 and 2, it follows that the social harm from these extra

crimes is increasing in the extent of the bias, and hence decreasing in the penalty enhancement for hate crimes. Thus, increasing the sanction imposed on hate-motivated crimes in period 1 reduces the extent to which observers in period 2 draw negative inferences about the disfavored group's characteristics. This, in turn, reduces the level of crime against the disfavored group in period 2.²⁸

There are a number of factors that may reinforce or counteract the basic effects we identify in Propositions 1 and 2. For example, recall the assumption above that X^H 's do not view disproportionate victimization of Y 's as unjust, and hence are not subject to the just world bias. While the focus has been entirely on the inferences made by X^N 's, the basic results would only be strengthened if X^H 's also committed additional crimes in period 2 due to the just world bias. Similarly, the results would be reinforced if the period-2 X^N 's underestimate the proportion of haters in the population, as well as the extent of their hatred.

The formulation in Equation (9) assumes that the new generation of X^N 's in period 2 correctly observes the sanctions that were in place in period 1. Thus, the observer recognizes that X^H 's faced a sanction s_1^{YH} (which potentially includes a penalty enhancement), even while underestimating the hate motivation of X^H 's. If period 2 X^N 's underestimate the sanction imposed on hate crimes in period 1, then this may reduce the degree of surprise occasioned by a marginal increase in the penalty enhancement. It is possible that this could reduce (or even reverse) the revision of \hat{c} described above. In addition, it may be the case that (while underestimating the extent of hate motivation among X^H 's, as assumed above) period 2 X^N 's attribute to X^H 's some private information that Y 's have negative characteristics. This would potentially "explain" the thickness in the tail of the distribution of X^H 's without necessarily revising \hat{c} downwards, and hence may reduce or reverse the effect in Proposition 1. However, all of these scenarios involve some misperception or cognitive bias in addition to the just world bias. The

²⁸If the model were extended to multiple periods, the just world bias would potentially lead to extra crime in each period, and perhaps eventually to a situation where all X^N 's commit crime against Y 's. This is possible for some distributions of preferences. However, for more realistic distributions, a large fraction of X^N 's would be expected to have values of b sufficiently small that they would never commit the crime, even for the maximum possible \hat{c} . Then, there will be some limit to the escalation of crime.

aim of the analysis above is to derive the effects attributable solely to the just world bias *per se* by abstracting from all other potential information asymmetries and cognitive biases.

Proposition 1 shows that marginal increases in s_1^{YH} reduce \hat{c} . On the other hand, increases in s_1^{YN} (the sanction for non-hate-motivated crimes against the disfavored group) do not have an unambiguous effect on the extent of the just world bias. Intuitively, an increase in s_1^{YN} will not generate a surprise for the observer, who has an accurate expectation of the consequences (a fall of $f_N(ps_1^{YN})$ in crimes against Y 's). Instead, the effect (if any) of changes in s_1^{YN} depends on the slope of the pdf f_N , and is not readily susceptible to an intuitive interpretation. Specifically:

Proposition 3 For marginal changes in s_1^{YN} , $\frac{d\hat{c}}{ds_1^{YN}} \begin{matrix} \geq \\ \leq \end{matrix} 0$ if $f_N(ps_1^{YN}) \begin{matrix} \geq \\ \leq \end{matrix} f_N(ps_1^{YN} - \hat{c})$.

Proof: Using the implicit-function rule (as in the proof of Proposition 1):

$$\frac{d\hat{c}}{ds_1^{YN}} = -\frac{p(f_N(ps_1^{YN}) - f_N(ps_1^{YN} - \hat{c}))}{f_N(ps_1^{YN} - \hat{c})} \quad (17)$$

from which the result follows straightforwardly.

This suggests that it is the penalty enhancement for hate-motivated crimes, rather than the sanction for crimes against Y 's *per se*, that has a clearer impact on the perception that the outgroup has negative characteristics. Proposition 3 can thus explain an important and distinctive feature of hate crime penalty enhancements, namely, that they are typically characterized by a concern for offenders' motivation (either in terms of animus against Y 's, or the discriminatory selection of Y 's as victims). This contrasts with "vulnerability" penalty enhancements that are aimed at protecting certain particularly vulnerable groups, such as the elderly (e.g. Moskowitz and De-Boer, 1999: 41-42). The latter kinds of enhancements are imposed without inquiring into offenders' motivation. They are also not reciprocal (in the sense that they do not apply if the victim is from any age group other than the elderly).

Vulnerability enhancements can be straightforwardly understood within the standard economic model of crime: attacks against particularly vulnerable victims are less costly to perpetrators (e.g. in terms of possible resistance or retaliation by the victims) and may cause victims greater direct harm (e.g.

if the elderly are more likely to be injured by an attack of given severity). On either of these grounds, enhanced penalties can be justified on optimal deterrence grounds.

On the other hand, hate crime penalty enhancements are typically characterized by a concern for offenders' motivation, and by reciprocity: for any given statutory criteria, the enhancement attaches to a crime committed against *anyone* on the basis of that criteria. When race is the criterion, for example, anyone in the population, not just members of a particular race, can be the victim of a hate crime justifying the enhancement. These features cannot be readily explained within the standard economic model; however they are consistent with the framework developed here. Proposition 3 shows that this framework gives rise to a concern with offenders' motivation (rather than simply with the group identity of the victim). Reciprocity arises in our model because it is possible for multiple groups to simultaneously be disproportionately victimized. In particular, the model can be readily extended to the case where there is hate motivation against the other group among subsets of both X 's and Y 's, and where the just world bias operates to increase future crime against both X 's and Y 's. In such a setting, both X 's and Y 's suffer more crime than they would in the absence of hate motivation among offenders of the other group, and so each is subject to disparate victimization. Consequently, penalty enhancements can reduce the social harm from increased future crime against both groups. This argument also extends straightforwardly to cases where there are more than two groups.

Finally, it should be noted that in a general welfare analysis, fully equalizing victimization rates (as advocated by Harel and Parchomovsky (1999)) may not be optimal because of the costs of imposing sanctions. The results from a comprehensive welfare analysis are not qualitatively different from those in Dharmapala and Garoupa (2004), where the optimality of penalty enhancements depends crucially on enforcement costs and the probability density function. With no enforcement costs, maximal penalties are optimal, and the question of penalty enhancement is moot. Hence, enforcement costs determine the structure of efficient sanctions. The probability distribution function determines the exact direction and magnitude of penalty enhancements, where a higher marginal benefit in terms of deterrence caused by a penalty enhancement must be traded off against the higher marginal cost of imposing it.

The main novelties in the welfare analysis caused by the explicit consideration of the just world bias are the following. On the one hand, the just world bias increases social welfare, because criminal actions are vindicated by the belief that the victims deserve their victimization. The reduction in the perceived wrongfulness of the criminal act makes those who commit crimes better off. While this positive effect on social welfare may be viewed by some as ethically dubious, it is nonetheless worth noting from a purely utilitarian standpoint. On the other hand, the just world bias dilutes deterrence by raising the net benefits from crime; thus, it leads to a higher level of crime for any given expected sanction. In the solution to the social planner's problem, the "extra crime" effect derived in Propositions 1 and 2 leads to a greater willingness to impose penalty enhancements for hate crimes, other things equal. Therefore, if we take the view that this second effect is primary, the model justifies penalty enhancements for hate crimes for reasons quite similar to Dharmapala and Garoupa (2004), that is, not because hate crimes are exogenously more harmful, but because they generate more harm endogenously by increasing the number of committed crimes.

4 Discussion and Conclusion

Myrdal (1944) identified a "vicious circle" in ethnic relations. Individuals tend to believe the world is more just than is actually the case, either because of indoctrination during childhood or as self-deception that productively offsets imperfect willpower. As a result, people have excessively strong priors that outcomes are deserved. When they observe an individual suffering a bad outcome, this prior belief conflicts with the inference that others unfairly imposed the bad outcome on the individual. In the context of ethnic relations, for example, observers will not sufficiently account for discrimination when assessing the causes of the disfavored ethnic group's poverty or poor education. Because they assess too little of the cause to discrimination, they will attribute too much to the group's negative characteristics; this victim-blaming produces additional discrimination. The result is the vicious circle.

We wish to emphasize the generality of our Myrdalian results. As we have noted throughout, the effects of disproportionate victimization and the BJW

may be manifested in a variety of ways. Our model could equally well be illustrated by other forms of hostility, such as social or economic discrimination. When the initial exogenous shock is a rise in hate crime, those subject to a strong BJW who are not willing to engage in crime may still blame the victim and therefore be more likely to engage in socially harmful acts of non-criminal discrimination against the disfavored group. Similarly, the initial exogenous shock may be a form of discrimination other than hate crimes. Besides the more obvious examples of employment or housing discrimination, consider the effect of racial profiling. Suppose that profiling by law enforcement personnel is due (at least partly) to animus towards the disfavored group. Then, observers would ascribe negative characteristics (such as a high probability of guilt) to the victimized group in order to reconcile these practices with Belief in a Just World. This could lead to an increased level of discrimination (or indeed to increased hate crimes) against the disfavored group. Thus, although our example is that hate crime produces more (non-hate) crime, any form of discrimination against a group can lead to greater discrimination against that group. When discrimination produces bad outcomes for victims, those outcomes change beliefs in ways that lead to more discrimination.

Returning to our hate crime example, we can summarize the argument as follows. The existence of animus by some members of one group against members of a second group leads to some potential offenders from the first group deriving greater benefits from crimes against members of the second group. These benefits produce greater victimization of the second group (under a system of uniform sanctions), which conflicts with notions of justice that are widely held (at least among non-hate-motivated members of the first group). Given their Belief in a Just World, some members of the first group ascribe negative characteristics to the second group, making the latter's greater victimization appear more deserved (or less undeserved). This, in turn, raises their net benefits from committing crimes against the second group, and thus results in additional crimes against the latter. These additional crimes (or other manifestations of hostility) constitute a distinctive social harm associated with disproportionate victimization.

Our model also shows that this social harm can be reduced by imposing penalty enhancements for hate crimes against the second group in order to ameliorate the bias. Our analysis generalizes. In particular, because the

potential for discriminatory preferences is reciprocal, so that members of the second group can also commit hate-motivated crimes against members of the first group, our analysis suggests that the penalty enhancement should also be reciprocal (as they are). The analysis also carries through to the case where there are three or more ethnic or racial groups. Finally, as explained above (in discussing Proposition 3), our model demonstrates that the beneficial effect of penalty enhancements applies only to penalties for hate-motivated crimes, rather than for all crimes against the disfavored group.²⁹ This highlights an important difference between hate crime penalty enhancements and what are termed “vulnerability” penalty enhancements that are aimed at protecting certain particularly vulnerable groups, such as the elderly (e.g. Moskowitz and DeBoer, 1999: 41-42).

The latter point is important because one of the central challenges facing any theory of penalty enhancements (e.g. Blake, 2001) is whether and why they should apply to certain kinds of groups (e.g. racial or religious minorities) but not others (e.g. young males aged 18-25, the homeless or the poor). Posner (2001: 233) argues that “. . . advocates of enhanced punishment for hate crimes mean by the term [only] . . . crimes against members of groups for which they have a particular solicitude, such as blacks, Jews, and homosexuals.” We do not claim to have dispelled Posner’s concern; interest group politics may indeed explain how legislatures define hate crimes in the real world. Note, for example, that the first two examples set out at the beginning of this article do not constitute hate crimes under any existing statute even though they illustrate our model as well as a racially motivated crime. However, our analysis suggests some new, hitherto neglected, factors that are relevant to the issue. Most fundamentally, our model suggests the importance of determining empirically, not only when disproportionate victimization exists (because some offenders are motivated by hatred of the group), but when such victimization is perceived as unjust. If there is no hatred or no perception of injustice, then there is no scope for the BJW to operate.³⁰

²⁹Another important generalization concerns what happens when there are more than two periods. Recall that a period is defined in the model by the entry of a new cohort of individuals. As long as each new cohort is subject to the JWB, it will be “surprised” by the victimization of the minority group, giving rise to negative inferences. In this sense, the model can be generalized to multiple periods.

³⁰These issues are also important for the question of how hate crime statutes are framed. There are two types of these statutes (e.g. Wang, 1999). One is based on an “animus”

The existing psychological literature does not directly address many of these questions, so it is difficult to reach any firm conclusions. Nonetheless, we hope that our analysis clarifies the issues involved, and suggests significant areas for future inquiry.

The model has a number of additional implications. Note for example the central role of uncertainty; the BJW influences inferences because there is uncertainty about the causes of crime, as where the crime is unsolved and no one directly observes the perpetrator's actions or motives. Under this assumption, any reduction in the level of uncertainty, such as an increase in the fraction of crimes that are solved, will reduce the need for inferring the causes of crime and therefore reduce the scope of the BJW. When crimes are known to have been committed because of hate motivation, they cannot be attributed to nonhating offenders (and hence "explained" by the supposed negative characteristics of the victimized group). This highlights the potential importance of laws that force the revelation of hate motives (through e.g. inquiries related to whether penalty enhancements should be applied). The provision of credible information about the role of hate motivation in the victimization of minority groups can ameliorate the vicious circle by countering the inference that the crimes result from the perception of the group's negative characteristics. Consistent with this notion is the observation that human rights organizations opposed to hate crimes often reveal and disseminate information about the role of hate motivation in crimes against the victimized group. This information is intended to attribute victimization to hate motivation as opposed to the negative characteristics of the victimized group. Thus, such publicity tends to counteract the inference underlying the vicious circle.

Third, although we have used our model to reveal the benefit of hate crime penalty enhancements, one could derive alternative policy implications from our model. Most obviously, given the point just made, is the advantage of detecting more crimes. One could seek to offset the effect of the just world bias on crime not only by greater penalties but also by an enhanced probability of detection. There is some evidence that hate crime statutes

model, focusing on the perpetrator's hostility to the victim's group. The other is the "discriminatory selection" model (upheld by the US Supreme Court in *Wisconsin v. Mitchell*, supra note 6), which punishes the selection of victims on the basis of group membership, independently of motivation.

may actually work this way in jurisdictions where the police department creates a special detective unit for investigating hate crimes that would not be investigated as seriously or at all were there no hate motivation (Bell, 2002). Moreover, the discovery and provision of information about the motivations of offenders may limit the influence of the BJW. The detection of crimes that occur in period one, and the determination by the courts of whether the perpetrators were hate-motivated or not, will tend to publicize information about the extent of hate motivation among offenders. Greater accuracy in courts determination of offenders motivations will also operate in the same direction. This suggests a novel expressive benefit of law enforcement, which is independent of whether or not penalty enhancements are imposed for hate-motivated crimes.³¹

Less obviously, one could seek to offset the effect of the just world bias by policies suppressing the dissemination of information about crime. If members of the majority do not learn of crime, then the bias cannot cause them to believe such victimization is deserved, and thereby increase crime. There are of course significant costs to suppressing true information about crimes (e.g. Dharmapala and McAdams, 2005), for instance, the reduced ability of potential victims to take precautions against crime. In addition, if it were known that information about crime is being suppressed, observers would infer some expected level of crime, based on their prior beliefs. This would reintroduce the just world bias. Indeed, it may even exacerbate it, as there would be no directly observable information about offenders motivation to constrain the inferential bias. In view of these difficulties, we observe that human rights and advocacy groups seeking to combat hate crimes typically do not pursue this approach. Rather, they seek to publicize the role of hate motivation to reduce the propensity of observers to attribute the crimes to nonhaters (and thus to revise their beliefs about the victimized group's characteristics).

There are also a number of other possible informational effects of hate crime statutes that may be relevant to our argument. For example, individuals who are unaware that hate crimes against group X occur in their community may infer from the passage of a hate crime statute that such

³¹However, courts may have more incentive to accurately determine whether crimes are hate-motivated if the sanctions they administer include penalty enhancements for hate crimes.

crimes in fact occur, and that members of group X suffer disproportionate victimization. This, in turn, may lead to negative beliefs about group X through the just world bias, an effect that may partially counteract the deterrent effect from the hate crime statute.³² This would not be relevant, of course, in situations where there has already been considerable publicity about hate crimes. However, where there has been no such publicity (in particular, because there have been no hate crimes), purely symbolic legislation might be counterproductive.³³

5 References

Akerlof, G. and W. T. Dickens (1982) “The Economic Consequences of Cognitive Dissonance” *American Economic Review*, 72, 307-319.

Arrow, K. J. (1973) “The Theory of Discrimination” In O. Ashenfelter and A. Rees (eds.) *Discrimination in Labor Markets* Princeton, N.J.: Princeton University Press, 3-33.

Becker, G. S. (1957) *The Economics of Discrimination* Chicago: University of Chicago Press.

Bell, J. (2002) *Policing Hatred: Law Enforcement, Civil Rights, and Hate Crime* New York: New York University Press.

Bénabou, R. and J. Tirole (2002) “Self-Confidence and Personal Motiva-

³²Another possibility is that the penalty enhancement causes hate-motivated criminals to conceal their motivation, whereas without the enhancement graffiti artists sometimes reveal their motivation by the racist words or symbols they write on a wall or gravestone. We note, however, that most of the many thousands of annual hate crimes carry no such public indicia of motivation. Moreover, those who do write such hateful words or symbols may have strong enough motives to continue to do so despite risking penalty enhancements: to achieve status among fellow haters or gain the pleasure of hurting those they hate, they may need to inflict more mental suffering on their victims than ordinary graffiti produces.

³³The passage of legislation can potentially convey a variety of types of information, beyond the hate crime context. New statutes may, for instance, reduce uncertainty about the content of the law, or reduce uncertainty concerning the preferences of others (in particular, concerning the extent to which people disapprove of criminal behavior, and the extent to which criminals benefit from crime). The overall effects are quite ambiguous. For a discussion of the informational consequences of legislative enactments, see Dharmapala and McAdams (2003).

tion” *Quarterly Journal of Economics*, 117, 871-915.

Bénabou, R. and J. Tirole (2006) “Belief in a Just World and Redistributive Politics,” *Quarterly Journal of Economics*, 121, 699-746.

Blake, M. (2001) “Geeks and Monsters: Bias Crime and Social Identity” *Law and Philosophy*, 20, 121-139.

Bonanno, G., C. Wortman, D. Lehman, R. Tweed, M. Haring, J. Sonnega, D. Carr, and R. Nesse (2002) “Resilience to Loss and Chronic Grief: A Prospective Study from Preloss to 18-Months Postloss” *Journal of Personality and Social Psychology*, 83, 1150-1164.

Brewer, W. F. (1996) “Good and bad story endings and story completeness” In R. J. Kreuz and M. S. MacNealy (eds.) *Empirical Approaches to Literature and Aesthetics* Norwood, NJ: Ablex, 261-271.

Dharmapala, D. and Garoupa, N. (2004) “Penalty Enhancement for Hate Crimes: An Economic Analysis” *American Law and Economics Review*, 6, 185-207.

Dharmapala, D., N. Garoupa and R. McAdams (2006) “The Just World Bias and Hate Crime Statutes” University of Illinois Law and Economics Research Working Paper 06-11.

Dharmapala, D. and R. McAdams (2003) “The Condorcet Jury Theorem and the Expressive Function of Law: A Theory of Informative Law” *American Law and Economics Review*, 5, 2003, 1-31.

Dharmapala, D. and R. McAdams (2005) “Words that Kill? An Economic Model of the Influence of Speech on Behavior (with Particular Reference to Hate Speech)” *Journal of Legal Studies*, 34, 93-136.

Federal Bureau of Investigation (2005) *Hate Crime Statistics 2004*, Washington, DC: US Department of Justice.

Frederick, S., G. Loewenstein and T. O’Donoghue (2002) “Time Discounting and Time Preference: A Critical Review” *Journal of Economic Literature*, 40, 351-401.

Furnham, A. (1993) “The Just World Belief in Twelve Cultures” *Journal of Social Psychology* 133, 317-29.

Gan, L., R. C. Williams, III, and T. Wiseman (2004) “A Simple Model of

Optimal Hate Crime Legislation” NBER Working Paper #10463 (forthcoming in *Economic Inquiry*).

Garoupa, N. (1997) “The Theory of Optimal Law Enforcement” *Journal of Economic Surveys*, 11, 267-295.

Grattet, R., V. Jenness, and T. R. Curry (1998) “The Homogenization and Differentiation of Hate Crime Law in the United States, 1978 to 1995: Innovation and Diffusion in the Criminalization of Bigotry” *American Sociological Review*, 63, 286-307.

Hafer C. (2000a) “Do Innocent Victims Threaten the Belief in a Just World? Evidence From a Modified Stroop Task” *Journal of Personality and Social Psychology*, 79, 165-173.

Hafer C. (2000b) “Investment in Long-Term Goals and Commitment to Just Means Drive the Need to Believe in a Just World” *Personality and Social Psychology Bulletin*, 26, 1059-1073.

Hafer C. (2002) “Why We Reject Innocent Victims” in M. Ross and D. Miller (eds.) *The Justice Motive in Everyday Life*, New York: Cambridge University Press, pp.109-126.

Harel, A. and G. Parchomovsky (1999) “On Hate and Equality” *Yale Law Journal*, 109, 507-539.

Harlow, C.W. (2005) “Hate Crime Reported by Victims and Police” *Bureau of Justice Statistics Special Report*, Washington, DC: US Department of Justice.

Hurd, H. and M. Moore (2004) “Punishing Hate and Prejudice” *Stanford Law Review*, 56, 1081-1146.

Jolls, C., C. R. Sunstein and R. H. Thaler (1998) “A Behavioral Approach to Law and Economics” *Stanford Law Review*, 50, 1471-1550.

Laibson, D. (1997) “Golden Eggs and Hyperbolic Discounting” *Quarterly Journal of Economics*, 112, 443-77.

Lawrence, F. (1999) *Punishing Hate: Bias Crime under American Law* Cambridge, MA: Harvard University Press.

Lerner, M. (1965) “Evaluation of Performance as a Function of Performer’s Reward and Attractiveness” *Journal of Personality and Social Psychology*,

1, 355-360.

Lerner, M. (1980) *The Belief in a Just World: A Fundamental Delusion* New York, NY: Plenum Press.

Lerner, M. (1998) "The Two Forms of Belief in a Just World: Some Thoughts on Why and How People Care about Justice" in L. Montada and M. Lerner (eds.) *Responses to Victimization and Belief in a Just World* New York, NY: Plenum Press, pp. 247-269.

Lerner, M. and E. Agar (1972) "The Consequences of Perceived Similarity: Attraction and Rejection, Approach and Avoidance" *Journal of Experimental Research in Personality*, 6, 69-75.

Lerner, M. and D. Miller (1978) "Just World Research and the Attribution Process: Looking Back and Ahead" *Psychological Bulletin*, 85, 1030-1051.

Lerner, M. and C. H. Simmons (1966) "The Observer's Reaction to the "Innocent Victim": Compassion or Rejection?" *Journal of Personality and Social Psychology*, 4, 203-210.

Lincoln, A. and G. Levinger (1972) "Observers' Evaluations of the Victim and the Attacker in an Aggressive Incident" *Journal of Personality and Social Psychology*, 22, 202-210.

Lipkus, I. M. and V. L. Bissonnette (1996) "Relationships among Belief in a Just World, Willingness to Accommodate, and Marital Well-Being" *Personality and Social Psychology Bulletin*, 22, 1043-56.

Long, G.T. and M.J. Lerner (1974) "Deserving, the 'Personal Contract,' and Altruistic Behavior by Children" *Journal of Personality and Social Psychology*, 29, 551-56.

Maes, J. (1994) "Blaming the Victim: Belief in Control or Belief in Justice?" *Social Justice Research*, 7, 69-90.

Maes, J. (1998) "Eight Stages in the Development of Research on the Construct of a Belief in a Just World" in L. Montada and M. Lerner (eds.) *Responses to Victimization and Belief in a Just World* New York, NY: Plenum Press, pp.163-185.

McAdams, R. (1995) "Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination" *Harvard Law Review*, 108, 1003-1084.

Mischel, W. (1974) "Processes in Delay of Gratification" in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, New York: Academic Press.

Moskowitz, S. and M. J. DeBoer (1999) "When Silence Resounds: Clergy and the Requirement to Report Elder Abuse and Neglect" *DePaul Law Review*, 49, 1-83.

Myrdal, G. (1944) *An American Dilemma: The Negro Problem and Modern Democracy* New York, NY: Harper and Row.

O'Donoghue, T. and M. Rabin (1999) "Doing It Now or Later" *American Economic Review*, 89, 103-24.

Piaget, J. (1965) *The Moral Judgment of the Child*, New York: Free Press.

Polinsky, A. M. and S. Shavell (2000) "The Economic Theory of Public Enforcement of Law" *Journal of Economic Literature*, 38, 45-76.

Posner, R. (2001) *Frontiers of Legal Theory* Cambridge, MA: Harvard University Press.

Rayburn, N. R., M. Mendoza and G. C. Davison (2003) "Bystanders' Perceptions of Perpetrators and Victims of Hate Crime" *Journal of Interpersonal Violence*, 18, 1055-1074.

Reichle, B. and M. Schmitt (2002) "Helping and Rationalization as Alternative Strategies for Restoring the Belief in a Just World: Evidence from Longitudinal Change Analyses" in M. Ross and D. Miller (eds.) *The Justice Motive in Everyday Life*, New York: Cambridge University Press, pp.127-148.

Ross, M. and D. Miller (eds.) (2002) *The Justice Motive in Everyday Life*, New York: Cambridge University Press.

Rubin, Z. and L. A. Peplau (1973) "Belief in a Just World and Reactions to Another's Lot: A Study of Participants in the National Draft Lottery" *Journal of Social Issues*, 29, 73-93.

Rubin, Z. and L. A. Peplau (1975) "Who Believes in a Just World?" *Journal of Social Issues*, 31, 65-89.

Selznick, G.J. and S. Steinberg (1969) *The Tenacity of Prejudice*, New York: Harper and Row.

Wang, L. (1997) "The Transforming Power of "Hate": Social Cognition Theory and the Harms of Bias-Related Crime" *Southern California Law Review*, 71, 47-135.

Wang, L. (1999) "The Complexities of 'Hate'" *Ohio State Law Journal*, 60, 799-900.

Wyer, R., G. Bodenhausen, and T. Gorman (1985) "Cognitive Mediators of Reactions to Rape" *Journal of Personality and Social Psychology*, 48, 324-338.

Readers with comments should address them to:

Professor Richard H. McAdams
University of Chicago Law School
1111 East 60th Street
Chicago, IL 60637
rmcadams@uchicago.edu

Chicago Working Papers in Law and Economics
(Second Series)

For a listing of papers 1–299 please go to Working Papers at
<http://www.law.uchicago.edu/Lawecon/index.html>

300. Adam B. Cox, The Temporal Dimension of Voting Rights (July 2006)
301. Adam B. Cox, Designing Redistricting Institutions (July 2006)
302. Cass R. Sunstein, Montreal vs. Kyoto: A Tale of Two Protocols (August 2006)
303. Kenneth W. Dam, Legal Institutions, Legal Origins, and Governance (August 2006)
304. Anup Malani and Eric A. Posner, The Case for For-Profit Charities (September 2006)
305. Douglas Lichtman, Irreparable Benefits (September 2006)
306. M. Todd Henderson, Paying CEOs in Bankruptcy: Executive Compensation when Agency Costs Are Low (September 2006)
307. Michael Abramowicz and M. Todd Henderson, Prediction Markets for Corporate Governance (September 2006)
308. Randal C. Picker, Who Should Regulate Entry into IPTV and Municipal Wireless? (September 2006)
309. Eric A. Posner and Adrian Vermeule, The Credible Executive (September 2006)
310. David Gilo and Ariel Porat, The Unconventional Uses of Transaction Costs (October 2006)
311. Randal C. Picker, Review of Hovenkamp, The Antitrust Enterprise: Principle and Execution (October 2006)
312. Dennis W. Carlton and Randal C. Picker, Antitrust and Regulation (October 2006)
313. Robert Cooter and Ariel Porat, Liability Externalities and Mandatory Choices: Should Doctors Pay Less? (November 2006)
314. Adam B. Cox and Eric A. Posner, The Second-Order Structure of Immigration Law (November 2006)
315. Lior J. Strahilevitz, Wealth without Markets? (November 2006)
316. Ariel Porat, Offsetting Risks (November 2006)
317. Bernard E. Harcourt and Jens Ludwig, Reefer Madness: Broken Windows Policing and Misdemeanor Marijuana Arrests in New York City, 1989–2000 (December 2006)
318. Bernard E. Harcourt, Embracing Chance: Post-Modern Meditations on Punishment (December 2006)
319. Cass R. Sunstein, Second-Order Perfectionism (December 2006)
320. William M. Landes and Richard A. Posner, The Economics of Presidential Pardons and Commutations (January 2007)
321. Cass R. Sunstein, Deliberating Groups versus Prediction Markets (or Hayek’s Challenge to Habermas) (January 2007)
322. Cass R. Sunstein, Completely Theorized Agreements in Constitutional Law (January 2007)
323. Albert H. Choi and Eric A. Posner, A Critique of the Odious Debt Doctrine (January 2007)
324. Wayne Hsiung and Cass R. Sunstein, Climate Change and Animals (January 2007)
325. Cass R. Sunstein, Cost-Benefit Analysis without Analyzing Costs or Benefits: Reasonable Accommodation, Balancing and Stigmatic Harms (January 2007)
326. Cass R. Sunstein, Willingness to Pay versus Welfare (January 2007)
327. David A. Weisbach, The Irreducible Complexity of Firm-Level Income Taxes: Theory and Doctrine in the Corporate Tax (January 2007)
328. Randal C. Picker, Of Pirates and Puffy Shirts: A Comments on “The Piracy Paradox: Innovation and Intellectual Property in Fashion Design” (January 2007)
329. Eric A. Posner, Climate Change and International Human Rights Litigation: A Critical Appraisal (January 2007)
330. Randal C. Picker, Pulling a Rabbi Out of His Hat: The Bankruptcy Magic of Dick Posner (February 2007)
331. Bernard E. Harcourt, Judge Richard Posner on Civil Liberties: Pragmatic (Libertarian) Authoritarian (February 2007)

332. Cass R. Sunstein, If People Would Be Outraged by Their Rulings, Should Judges Care? (February 2007)
333. Eugene Kontorovich, What Standing Is Good For (March 2007)
334. Eugene Kontorovich, Inefficient Customs in International Law (March 2007)
335. Bernard E. Harcourt, From the Asylum to the Prison: Rethinking the Incarceration Revolution. Part II: State Level Analysis (March 2007)
336. Cass R. Sunstein, Due Process Traditionalism (March 2007)
337. Adam B. Cox and Thomas J. Miles, Judging the Voting Rights Act (March 2007)
338. M. Todd Henderson, Deconstructing Duff & Phelps (March 2007)
339. Douglas G. Baird and Robert K. Rasmussen, The Prime Directive (April 2007)
340. Cass R. Sunstein, Illusory Losses (May 2007)
341. Anup Malani, Valuing Laws as Local Amenities (June 2007)
342. David A. Weisbach, What Does Happiness Research Tell Us about Taxation? (June 2007)
343. David S. Abrams and Chris Rohlf, Optimal Bail and the Value of Freedom: Evidence from the Philadelphia Bail Experiment (June 2007)
344. Christopher R. Berry and Jacob E. Gersen, The Fiscal Consequences of Electoral Institutions (June 2007)
345. Matthew Adler and Eric A. Posner, Happiness Research and Cost-Benefit Analysis (July 2007)
346. Daniel Kahneman and Cass R. Sunstein, Indignation: Psychology, Politics, Law (July 2007)
347. Jacob E. Gersen and Eric A. Posner, Timing Rules and Legal Institutions (July 2007)
348. Eric A. Posner and Adrian Vermeule, Constitutional Showdowns (July 2007)
349. Lior Jacob Strahilevitz, Privacy versus Antidiscrimination (July 2007)
350. Bernard E. Harcourt, A Reader's Companion to *Against Prediction*: A Reply to Ariela Gross, Yoram Margalioth, and Yoav Sapir on Economic Modeling, Selective Incapacitation, Governmentality, and Race (July 2007)
351. Lior Jacob Strahilevitz, "Don't Try This at Home": Posner as Political Economist (July 2007)
352. Cass R. Sunstein, The Complex Climate Change Incentives of China and the United States (August 2007)
353. David S. Abrams and Marianne Bertrand, Do Judges Vary in Their Treatment of Race? (August 2007)
354. Eric A. Posner and Cass R. Sunstein, Climate Change Justice (August 2007)
355. David A. Weisbach, A Welfarist Approach to Disabilities (August 2007)
356. David S. Abrams, More Time, Less Crime? Estimating the Deterrent Effect of Incarceration using Sentencing Enhancements (August 2007)
357. Stephen J. Choi, G. Mitu Gulati and Eric A. Posner, Professionals or Politicians: The Uncertain Empirical Case for an Elected Rather than Appointed Judiciary (August 2007)
358. Joseph Bankman and David A. Weisbach, Consumption Taxation Is Still Superior to Income Taxation (September 2007)
359. Douglas G. Baird and M. Todd Henderson, Other People's Money (September 2007)
360. William Meadow and Cass R. Sunstein, Causation in Tort: General Populations vs. Individual Cases (September 2007)
361. removed by author
362. Richard McAdams, Reforming Entrapment Doctrine in *United States v. Hollingsworth* (September 2007)
363. M. Todd Henderson, From *Seriatim* to Consensus and Back Again: A Theory of Dissent (October 2007)
364. Timur Kuran and Cass R. Sunstein, Availability Cascades and Risk Regulation (October 2007)
365. David A. Weisbach, The Taxation of Carried Interests in Private Equity (October 2007)
366. Lee Anne Fennell, Homeownership 2.0 (October 2007)
367. Jonathan R. Nash and Rafael I. Pardo, An Empirical Investigation into Appellate Structure and the Perceived Quality of Appellate Review (October 2007)
368. Thomas J. Miles and Cass R. Sunstein, The Real World of Arbitrariness Review (November 2007)
369. Anup Malani, Maciej F. Boni, Abraham Wickelgren, and Ramanan Laxminarayan, Controlling Avian Influenza in Chickens (November 2007)
370. Richard H. McAdams, The Economic Costs of Inequality (November 2007)

371. Lior Jacob Strahilevitz, Reputation Nation: Law in an Era of Ubiquitous Personal Information (November 2007)
372. Thomas J. Miles and Cass R. Sunstein, The New Legal Realism (December 2007)
373. M. Todd Henderson, Everything Old Is New Again: Lessons from *Dodge V. Ford Motor Company* (December 2007)
374. Jonathan Remy Nash, Economic Efficiency versus Public Choice: The Case of Property Rights in Road Traffic (December 2007)
375. Edward L. Glaeser and Cass R. Sunstein, Extremism and Social Learning (December 2007)
376. Stephen J. Choi, G. Mitu Gulati, and Eric A. Posner, Are Judges Overpaid?: A Skeptical Response to the Judicial Salary Debate (December 2007)
377. Eric A. Posner, Does Political Bias in the Judiciary Matter?: Implications of Judicial Bias Studies for Legal and Constitutional Reform (January 2008)
378. Susan Bandes, The Heart Has Its Reasons: Examining the Strange Persistence of the American Death Penalty (January 2008)
379. Susan Bandes, After Innocence: Framing Wrongful Convictions (January 2008)
380. Jacob E. Gersen and Ann Joseph O'Connell, Deadlines in Administrative Law (January 2008)
381. Richard A. Epstein, The Property Rights Movement and Intellectual Property (January 2008)
382. Richard A. Epstein, Some Reflections on Custom in the IP Universe (January 2008)
383. Richard A. Epstein, Decentralized Responses to Good Fortune and Bad Luck (January 2008)
384. Richard A. Epstein, How to Create—or Destroy—Wealth in Real Property (January 2008)
385. Richard A. Epstein, The Human and Economic Dimensions of Altruism: The Case of Organ Transplantation (January 2008)
386. Cass R. Sunstein, Adolescent Risk-Taking and Social Meaning: A Commentary (January 2008)
387. Cass R. Sunstein and Adrian Vermeule, Conspiracy Theories (January 2008)
388. Ariel Porat, Expanding Restitution: Liability for Unrequested Benefits (January 2008)
389. Randal C. Picker, *Twombly*, *Leegin* and the Reshaping of Antitrust (February 2008)
390. Alon Harrel and Ariel Porat, Aggregating Probabilities across Offences in Criminal Law (March 2008)
391. John Bronsteen, Christopher Buccafusco, and Jonathan Masur, Hedonic Adaptation and the Settlement of Civil Lawsuits (March 2008)
392. Randal C. Picker, Fair Use v. Fair Access (March 2008)
393. Jonathan S. Masur, Process as Purpose: Administrative Procedures, Costly Screens, and Examination at the Patent Office (March 2008, revised July 2008)
394. Eric A. Posner, Human Welfare, Not Human Rights (March 2008)
395. Lee Anne Fennell, Slices and Lumps, 2008 Coase Lecture (March 2008)
396. Eric A. Posner, Fault in Contract Law (March 2008)
397. Cass R. Sunstein, Is OSHA Unconstitutional? (March 2008)
398. Randal C. Picker, Take Two: Stare Decisis in Antitrust/The Per Se Rule against Horizontal Price-Fixing (March 2008)
399. M. Todd Henderson and Anup Malani, Corporate Philanthropy and the Market for Altruism (April 2008, revised September 2008)
400. Shyam Balganes, Foreseeability and Copyright Incentives (April 2008)
401. Cass R. Sunstein and Reid Hastie, Four Failures of Deliberating Groups (April 2008)
402. M. Todd Henderson, Justin Wolfers and Eric Zitzewitz, Predicting Crime (April 2008)
403. Richard A. Epstein, Bell Atlantic v. Twombly: How Motions to Dismiss Become (Disguised) Summary Judgments (April 2008)
404. William M. Landes and Richard A. Posner, Rational Judicial Behavior: A Statistical Study (April 2008)
405. Stephen J. Choi, Mitu Gulati, and Eric A. Posner, Which States Have the Best (and Worst) High Courts? (May 2008)
406. Richard H. McAdams and Janice Nadler, Coordinating in the Shadow of the Law: Two Contextualized Tests of the Focal Point Theory of Legal Compliance (May 2008)
407. Cass R. Sunstein, Two Conceptions of Irreversible Environmental Harm (May 2008)
408. Richard A. Epstein, Public Use in a Post-*Kelo* World (June 2008)
409. Jonathan R. Nash, The Uneasy Case for Transjurisdictional Adjudication (June 2008)
410. Adam B. Cox and Thomas J. Miles, Documenting Discrimination? (June 2008)

411. M. Todd Henderson, Alan D. Jagolinzer, and Karl A. Muller, III, *Scienter Disclosure* (June 2008)
412. Jonathan R. Nash, *Taxes and the Success of Non-Tax Market-Based Environmental Regulatory Regimes* (July 2008)
413. Thomas J. Miles and Cass R. Sunstein, *Depoliticizing Administrative Law* (June 2008)
414. Randal C. Picker, *Competition and Privacy in Web 2.0 and the Cloud* (June 2008)
415. Omri Ben-Shahar, *The Myth of the "Opportunity to Read" in Contract Law* (July 2008)
416. Omri Ben-Shahar, *A Bargaining Power Theory of Gap-Filling* (July 2008)
417. Omri Ben-Shahar, *How to Repair Unconscionable Contracts* (July 2008)
418. Richard A. Epstein and David A. Hyman, *Controlling the Costs of Medical Care: A Dose of Deregulation* (July 2008)
419. Eric A. Posner, *Erga Omnes Norms, Institutionalization, and Constitutionalism in International Law* (August 2008)
420. Thomas J. Miles and Eric A. Posner, *Which States Enter into Treaties, and Why?* (August 2008)
421. Cass R. Sunstein, *Trimming* (August 2008)
422. Cass R. Sunstein, *Second Amendment Minimalism: Heller as Griswold* (August 2008)
423. Richard A. Epstein, *The Disintegration of Intellectual Property* (August 2008)
424. John Bronsteen, Christopher Buccafusco, and Jonathan Masur, *Happiness and Punishment* (August 2008)
425. Adam B. Cox and Thomas J. Miles, *Judicial Ideology and the Transformation of Voting Rights Jurisprudence* (August 2008)
426. Daniel Abebe and Jonathan S. Masur, *A Nation Divided: Eastern China, Western China, and the Problems of Global Warming* (August 2008)
427. William Birdthistle and M. Todd Henderson, *One Hat Too Many? Investment Desegregation in Private Equity* (August 2008)
428. Irina D. Manta, *Privatizing Trademarks* (abstract only) (September 2008)
429. Paul J. Heald, *Testing the Over- and Under-Exploitation Hypothesis: Bestselling Musical Compositions (1913–32) and Their Use in Cinema (1968–2007)* (September 2008)
430. M. Todd Henderson and Richard A. Epstein, *Introduction to "The Going Private Phenomenon: Causes and Implications"* (September 2008)
431. Paul Heald, *Optimal Remedies for Patent Infringement: A Transactional Model* (September 2008)
432. Cass R. Sunstein, *Beyond Judicial Minimalism* (September 2008)
433. Bernard E. Harcourt, *Neoliberal Penalty: The Birth of Natural Order, the Illusion of Free Markets* (September 2008)
434. Bernard E. Harcourt, *Abolition in the U.S.A. by 2050: On Political Capital and Ordinary Acts of Resistance* (September 2008)
435. Robert Cooter and Ariel Porat, *Liability for Lapses: First or Second Order Negligence?* (October 2008)
436. Ariel Porat, *A Comparative Fault in Defense Contract Law* (October 2008)
437. Richard H. McAdams, *Beyond the Prisoners' Dilemma: Coordination, Game Theory and the Law* (October 2008)
438. Dhammika Dharamapala, Nuno Garoupa, and Richard H. McAdams, *Belief in a Just World, Blaming the Victim, and Hate Crime Statutes* (October 2008)