Dhammika Dharmapala, Nuno Garoupa & Richard H. McAdams[1]

Hate Crime: A Behavioural Economic Analysis

1. Introduction

In the United States, a »hate crime« is, among other things, a legal category. In recent years, Congress and many state legislatures have taken sentencing discretion away from judges by implementing a highly detailed sentencing code. Within such codes, some jurisdictions increase the sentence if the crime is committed because of the perpetrator's animus or hatred toward a particular group, while many other jurisdictions do so if the victim is selected because of membership in a particular group.[2] For example, the Federal sentencing guidelines provide for a »penalty enhancement« if the defendant »intentionally selected any victim or any property as the object of the offence of conviction because of the actual or perceived race, colour, religion, national origin, ethnicity, gender, disability, or sexual orientation of any person.«[3] Most states have similar laws for state offences, though the list of possible groups varies widely.[4] As a result, in the United States, a hate crime is an act that is already a crime, but which is punished more harshly because of the motive the perpetrator had for committing the offence.[5]

In recent years, the FBI has received reports of hate crimes involving 9,000-12,000 victims per year in the United States.[6] Most of these crimes are based on race. A few are extremely serious, such as the murderous rampages of Benjamin Smith and Buford Furrow, Jr., in 1999, Richard Baumhammers in 2000, and Mark Stroman in 2001.[7] But most hate crimes involve less serious crimes, such as vandalism, intimidation, or

1 Dhammika Dharmapala, Assistant Professor of Economics, University of Connecticut, and Visiting Assistant Professor of Business Economics and Public Policy, University of Michigan (email: dhammika@umich.edu); Nuno Garoupa, Professor of Law and Economics, Universidade Nova de Lisboa and Research Affiliate, FEDEA, Madrid & CEPR, London (email: ngaroupa@fe.unl.pt); Richard H. McAdams, Guy Raymond Jones Professor of Law, University of Illinois (email: rmcadams@law.uiuc.edu).

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

3 28 U.S.C. § 994 (1994).

4 See e.g. R. Grattet, V. Jenness and T. R. Curry »The Homogenization and Differentiation of Hate Crime Law in the United States, 1978 to 1995: Innovation and Diffusion in the Criminalization of Bigotry« 63 *American Sociological Review* 286 (1998).

5 Similarly, in the UK, although hate crime has no specific legal meaning, hate motivation is an aggravating factor in sentencing under 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).

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

7 See D. Dharmapala and R. McAdams »Words that Kill: An Economic Model of the Influence of Speech on Behavior (with Particular Reference to Hate Speech)« 34 *Journal of Legal Studies* 93 (2005) for descriptions of these events.

simple assault.[8] Because many crimes go unreported, and there are always definitional questions about hate crimes, estimates of the number of hate crimes vary widely. Based on the National Crime Victimization Survey, the Bureau of Justice Statistics reported an annual average of 210,000 hate crime victims in the US from 2000 to 2003.[9]

In the United States, there is a large legal and philosophical literature on hate crimes.[10] The central issue in this analysis is whether and how to justify the penalty enhancements for hate crimes. Some critics argue that the penalty enhancement is not justified, as it violates the principle that punishment be based on bad conduct rather than bad thoughts. Those theorists who defend the enhancements argue that the racial or other group-based animus makes the criminal act more harmful or wrongful. Economic reasoning has until recently contributed little to understanding hate crime, but behavioural economics has great potential for illuminating the issues.[11] Specifically, we use behavioural economics to identify a mechanism through which hate crime may cause more social harm than the same crime committed without hate. In this brief essay, we describe our work on this topic,[12] and place it within the context of the wider scholarly debate surrounding hate crimes.

2. The »Extra Harm« of Hate Crimes

What is it about hate crimes that might justify enhanced punishment? There are many possible answers, which we review briefly. The literature in this area has predominantly been based on a deontological and retributivist framework. A deontological claim that has been made in this literature is that hate crimes constitute, for some reason, a greater wrong than the same crime committed without a hate motivation. However, this claim is controversial among scholars, and has been challenged from within the deontological paradigm.[13] Alternatively, distributive justice theorists have argued that hate crimes cause disproportionate victimization of minority groups, which is intrinsically unfair.[14] However, whether fairness can be determined from outcomes in this way remains a contested notion among theorists of justice.

---

8 See e.g. Federal Bureau of Investigation, *supra* note 6.

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

10 See e.g. F. Lawrence *Punishing Hate: Bias Crime under American Law* Cambridge, MA: Harvard University Press (1999), and H. Hurd and M. Moore »Punishing Hate and Prejudice« 56 *Stanford Law Review*, 1081 (2004).

11 For a critical review of behavioural economics as applied to crime, see N. Garoupa »Behavioral Economic Analysis of Crime: A Critical Review« 15 *European Journal of Law and Economics*, 5 (2003).

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

13 See e.g. A. Dillof »Punishing Bias: An Examination of the Theoretical Foundations of Bias Crime Statutes« 91 *Northwestern Law Review* 1015 (1997).

14 See A. Harel and G. Parchomovsky »On Hate and Equality« 109 *Yale Law Journal* 507 (1999).

The most straightforward consequentialist justification for penalty enhancements would be that hate crimes cause more harm to the victim than would the same crime in the absence of a hate motivation.[15] However, it is difficult to quantify or compare these harms, especially when they are of a psychological nature. In some cases, hate crimes create a greater risk of future harm by inviting retaliation from members of the victims' group, possibly leading to an escalating cycle of violence. However, such cases are likely to be relatively rare, and in any event penalty enhancements in US law are not restricted to these kinds of circumstances.

The economic theory of crime, constituting a subset of the wider consequentialist paradigm, has until recently played little role in the debate about hate crime penalty enhancements. It is, however, possible to extend the standard economic model of crime to analyze hate crimes,[16] by assuming that the population consists of different groups, with some potential offenders being motivated by animus or hatred towards certain groups (in the sense that they derive greater benefits from committing an otherwise identical crime against members of a minority group than from committing the same crime against a member of their own group). A central conclusion of this analysis is that, within the economic perspective, there is no distinctive harm that arises from the disproportionate victimization of minority groups. Rather, whether or not penalty enhancements are justified from the standpoint of optimal deterrence theory depends on a variety of contingent empirical factors, such as the costs of imposing sanctions.[17]

Our approach in the research described here differs from all of the theories sketched above. Specifically, we advance a consequentialist claim that hate crimes cause more harm to *society* by causing more crime. The extra harm that our approach focuses on is not that individual hate crime victims suffer more than other individual victims, but that disproportionate victimization of minority groups may give rise to inferences about group characteristics that in turn lead to more crime in the future. That is, hate crimes ensure that there will be more crime victims (though not necessarily more *hate* crime victims). This approach and its implications are described below.

3. The Basic Framework and Argument

In the conventional economic model of crime, an individual offends whenever her expected benefits from crime exceed her expected costs (including the expected sanction). This framework can be extended to study the issue of hate crime by assuming that victims can differ in their group identity. Though we could make our point by considering many kinds of groups defined along different dimensions (e.g., race, religion,

15 The argument of L. Wang »The Transforming Power of 'Hate': Social Cognition Theory and the Harms of Bias-Related Crime« 71 *Southern California Law Review* 47 (1997) can be understood in this light, essentially claiming that the psychological harms suffered by victims of hate crimes are greater than those suffered by victims of parallel nonhate crimes.

16 See D. Dharmapala and N. Garoupa »Penalty Enhancement for Hate Crimes: An Economic Analysis« 6 *American Law and Economics Review*, 185 (2004).

17 Dharmapala and Garoupa, *supra* note 16.

sexual preference, etc.), for simplicity, we assume that the population consists of just two groups, one dominant and one disfavoured, with the groups being of equal size. Suppose that a subset of potential offenders from the dominant group has discriminatory or hateful preferences, so they derive greater benefits from committing a crime against members of the disfavoured group than from committing the otherwise identical crime against a member of their own group. If expected sanctions are the same regardless of the victim's group, the hate motivation ensures that members of the disfavoured group will face a disproportionately high probability of victimization. Note that victimization is »disproportionate« here in the sense that it occurs at a higher rate than in the absence of hate motivation (as opposed to *statistically* disproportionate victimization where their members constitute a greater percentage of crime victims than is warranted by their percentage in the total population). Thus, this notion of disproportionate victimization can be readily extended to the case of hatred against multiple groups (so that, if a subset of potential offenders from the disfavoured group have discriminatory preferences toward the dominant group, then members of the dominant group would, along with the disfavoured group, suffer a disproportionately high probability of victimization). However, in the interests of simplicity, we focus only on the hatred by some offenders from the dominant group.

We extend this economic framework by introducing a new variable distinct from biased tastes: the beliefs that a potential offender holds about the moral characteristics of a potential victim (»moralistic beliefs«). Here, we assume that all individuals have a preference for avoiding the shame of social disapproval, which they expect to incur if they are discovered to have committed a crime. Alternatively, we may assume that offenders anticipate incurring psychological aversion or guilt from committing a crime. In either case, we further assume that the shame or guilt varies with the perceived intrinsic value or »moral worth« of the individual one victimizes. That is, the offender expects to incur less guilt or shame – i.e. less cost – from committing an offence 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. These moralistic beliefs, like discriminatory preferences, partly determine the net costs of crime. For example, members of the dominant group may believe that members of the disfavoured group are more likely to be untrustworthy, to be loyal to a foreign nation, or to use welfare programs. The more negative these characteristics are believed to be, the lower are the expected costs from committing a crime against a member of the disfavoured group.

Our central claim concerns the interaction of moralistic beliefs and discriminatory preferences: in particular, disparate victimization caused by discriminatory preferences can in turn influence individuals' moralistic beliefs in a way that produces more crime. Specifically, we identify the following causal chain:

(1) discriminatory preferences cause disproportionate victimization of the disfavoured group;

(2) disproportionate victimization causes members of the dominant group to infer that others hold more negative moralistic beliefs about the disfavoured group than was previously thought;

(3) because the beliefs of others provide some (partial) evidence about the true state of affairs, individuals will now revise downward their own moralistic beliefs about the disfavoured group;

(4) more negative moralistic beliefs lower the expected cost of crime and therefore increases the amount of crime against the disfavoured group.

Next, we explain the causal chain above in more detail. Step (1) is straightforward and requires no behavioural insights. The benefit individuals receive from satisfying their discriminatory preferences for crime make it likely that some individuals will commit some crimes that would have, in the absence of the preference, been deterred. The preferences thus generate disproportionate victimization. Step (4) is similarly straightforward: believing the group to have more negative characteristics lowers the expected cost of crime, and so will lead to a higher level of crime (holding fixed the expected sanction).

Steps (2) and (3) concern the inferences made by those who observe disproportionate victimization. Individuals will, in general, have some prior beliefs about the parameters that determine the crime level before they observe the actual crime rate. If it differs from what was expected, then rational individuals will update their beliefs about the relevant parameters. Our claim – based on psychological evidence outlined below – is that individuals will systematically underestimate the contribution of discriminatory preferences and so expect less crime against the disfavoured group than actually occurs. When they observe the higher-than-expected crime rate, they can attribute the greater victimization of the disfavoured group either (a) to offenders' discriminatory preferences or (b) to the victims' negative characteristics (or equivalently, to offenders' moralistic beliefs). If individuals underestimate the force of discriminatory preferences, they will therefore over-attribute the extra crime to moralistic beliefs (and hence to the disfavoured group's negative characteristics). Strictly speaking, there are two steps here. First, the observer infers that the extra crimes she observes are caused by the fact that others hold more negative moralistic beliefs than the observer had previously believed. Second, as the beliefs of others provide some (partial) evidence about the true state of affairs, the observer infers from the more negative moralistic beliefs others hold about the disfavoured group that the disfavoured group in fact has more negative characteristics than the observer previously believed.

As stated, however, steps (2) and (3) assume a certain cognitive bias. To clarify this point, let us assume that individuals make perfectly rational inferences. Note that even rational individuals have imperfect information, so their estimates about the relevant parameters will be subject to error. Thus, some individuals may be surprised by the crime levels they observe. For Bayesians, however, there is no reason to believe that their priors will be wrong *on average*. Some individuals will underestimate the strength or pervasiveness of discriminatory preferences and therefore have expected less crime against the disfavoured group than actually occurs. Others will overestimate the discriminatory preferences and therefore have expected more crime against the dis-

favoured group. Both groups will update their beliefs in light of the actual amount of crime that is observed; there will be no systematic effect on the inference about the disfavoured group's negative characteristics. Thus, if all individuals are Bayesians, hate crime will not increase total crime.

4. Belief in a Just World and the »Just World Bias«

We claim, however, that people are not generally Bayesians. In some ways, this point seems obvious to almost anyone who is not an economist; statistics teachers have to struggle mightily to teach the perfect logic of Bayesian inference. Note, however, that it is not sufficient for our purposes that people deviate in any manner from Bayesian inference. Our claim is that people deviate in a systematic way: to a greater degree than is rational, individuals resist making inferences that would cause them to view the world as being less just than they previously believed. Instead, they strive to interpret the world in a way that preserves their belief that the world tends to give people what they deserve.

In support of this assumption, we draw on a social psychological literature that studies the existence and effect of a belief in a just world. In one of the pioneering experiments in this literature,[18] subjects viewed on a television what appeared to be a contemporaneous experiment on learning, in which a subject (actually a confederate of the experimenters) was receiving extremely 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 adjusted their views of the victim downward to make her seem more deserving of her bad outcome.[19]

---

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

19 One 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.

The experimental setting described above may seem artificial. However, psychologists have found similar effects with respect to a variety of more typical victims. Most relevant for our purposes, this effect has been found for victims of crime.[20] Moreover, these findings have been replicated using a variety of alternative methodologies[21] and across a range of cultures. The essential point that these studies find is not that people believe that the world is perfectly just, but that they strive to interpret it as being as just as possible. Causal attributions for bad outcomes are complex and difficult. We expect even rational Bayesians to make errors, overestimating or underestimating the degree to which an individual is responsible for bad things that happen to her. However, the psychological research summarized above implies that instead of these errors being randomly distributed around a mean that represents the correct causal attribution, they are skewed towards over-attributing bad outcomes to the negative characteristics of the individuals who suffer them.

This process does not require that individuals consciously denigrate the victim: Lerner argues that:[22] »When these reactions appear, they are naturally framed in ways that do not directly violate conventional rules of logic or morality, e.g., the person who derogates a victim will generate a culturally plausible basis for that condemnation.« This notion of cultural plausibility also suggests that (even though the classic experiments in this area typically focus on inferences about individuals), it is reasonable to extend the notion to inferences about groups. The perceived negative characteristics of a disfavoured group to which the victim belongs naturally provides such a culturally plausible basis for negative inferences.

There are various explanations for the apparent persistence of the belief in a just world – for instance, as a rule of thumb for making complex attributions or as a way of coping with anxiety. However, regardless of the reasons for its persistence, it is possible to characterise the effects of this belief as a cognitive bias. That is, the desire to maintain (as far as possible) the belief in a just world will cause individuals to depart from rational Bayesian inference procedures, and to make inferences about victims' characteristics that are more negative than would be the inferences of a pure Bayesian. We term this cognitive bias the »just world bias« (JWB).

5. Applying the Just World Bias to the Analysis of Hate Crimes

Now, assume that individuals are subject to the JWB, and return to the simple causal chain outlined in Section 3. Suppose that the underlying preferences are such that both haters (i.e. those motivated by discriminatory preferences) and nonhaters (i.e. those

20 See e.g. R. Wyer, G. Bodenhausen, and T. Gorman »Cognitive Mediators of Reactions to Rape« 48 *Journal of Personality and Social Psychology* 324 (1985).

21 See e.g. C. Hafer »Do Innocent Victims Threaten the Belief in a Just World? Evidence From a Modified Stroop Task« 79 *Journal of Personality and Social Psychology*, 165 (2000).

22 M. Lerner »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 247 (1998) at 255.

motivated by moralistic beliefs) within the dominant group commit crimes against the disfavoured group. In step (2), observers observe the rate of victimisation of the disfavoured group, and know that crimes committed by haters are attributable to the offenders' preferences, while crimes committed by nonhaters are attributable to the offenders' moralistic beliefs about the victimized group's negative characteristics. It seems realistic to assume that not all crimes are solved, so there are at least some crimes for which there is uncertainty about the offenders' motivations. These crimes could be attributed either to moralistic beliefs or to discriminatory preferences.

Psychological evidence suggests that most people view crime victimization caused by hate motivation as being more unjust than victimization caused by negative characteristics of the victims. In particular, it has been found 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, they rate the victim of a hate crime as less culpable than the victim of a non-hate crime.[23] These results support the assumption that most people view hate-motivated victimization as particularly unjust.

Attributing the unsolved crimes at step (2) to the discriminatory preferences of haters entails that the victims were targeted through no fault of their own, and thus conflicts with a belief in a just world. If observers are subject to the JWB, they will thus strive to interpret the world in a way that reduces the need to attribute unsolved crimes to discriminatory preferences. In this context, this entails attributing the (unsolved) crimes to nonhaters. However, because nonhaters' crimes are motivated by moralistic beliefs about the disfavoured group's negative characteristics, the only way to reconcile this attribution with the observed rate of victimization is to revise (negatively) moralistic beliefs about the disfavoured group. That is, the observer subject to the JWB attributes to non-hate-motivated offenders a more negative moralistic belief about the disfavoured group than those offenders actually hold. As others' (perceived) beliefs are a source of (imperfect) information about the true state of the world, observers will themselves negatively revise their moralistic beliefs about the disfavoured group. Observers will thus come to the conclusion that the disfavoured group is even less trustworthy, or even more disloyal to the country, or even lazier, than they previously believed. Thus, given the JWB, step (3) of our causal chain will be valid.

By underestimating the role of discriminatory preferences in the face of uncertainty, observers preserve (to some degree) their belief in the basic justness of the world. However, the revised beliefs about the victimized group's negative characteristics raise the net benefits from crimes against that group. As discussed above, this can lead (through step (4)) to additional crimes being committed against that group. The harms from these additional crimes would not occur in the absence of the earlier hate crimes; in this sense, hate crimes generate extra crimes against the disfavoured group. Moreo-

23 N. R. Rayburn, M. Mendoza and G. C. Davison »Bystanders' Perceptions of Perpetrators and Victims of Hate Crime« 18 *Journal of Interpersonal Violence* 1055 (2003), at 1062-3 and 1069.

ver, these additional crimes are clearly costly to society, and these costs are in principle measurable (unlike, for instance, the psychic harms to victims).

6. A Simple Example

A simple example helps to clarify the essence of the argument above. Consider a country where the population is divided into two ethnic groups – A's (the majority group) and B's (the minority group). Suppose that the two groups are identical in terms of factors that affect the probability of crime victimization, such as residential location, wealth, and age. On the basis of these characteristics alone, we assume that the average probability of victimization of any individual in the country is 1% (for example, this can be interpreted as follows: every year, one person is victimized out of hundred residents in the country). However, suppose that some (relatively small) subgroup of A's have hate preferences with respect to crimes against B's (for example, these may be followers of an extremist ideology that advocates the »ethnic cleansing« of B's). These preferences among this subgroup of A's leads to disproportionate victimization of B's – for example, suppose that for a random member of the minority the average probability of victimization is 2%, rather than 1%.

Now, consider the inference problem faced by a newcomer to the country (who, for the sake of the argument, belongs to group A), or by a member of a new generation of A's. Suppose that this individual is uncertain about the extent of hate preferences among her fellow-A's (for instance, she does not know precisely how many A's subscribe to the extremist ideology that targets B's). She directly observes, however, the rates of crime victimization for each group. If she engages in Bayesian inference, then (on average) she will infer that the disproportionate victimization of B's is caused by the hate preferences of the extremist A's. On average, she will correctly infer the fraction of A's who hold such preferences, based on the observed rates of victimization and knowledge of other factors such as expected sanctions.

However, suppose instead that the observer is subject to the JWB. Then, the inference described above (where B's are disproportionately victimized through no fault of their own) will conflict with the desire to believe in a just world. Moreover, suppose that there is a widespread belief among A's that B's have certain negative characteristics, such as dishonesty in commercial transactions. Then, the »extra« crimes suffered by B's could be attributed either to extremist A's (motivated by hate preferences) or to nonextremist A's (for instance, motivated by rage after being cheated by a B). The model sketched in Section 5 implies that observers subject to the JWB will over-attribute crimes against B's to the latter cause. It follows that the observer must also revise her beliefs about the prevalence of dishonesty among B's – if the extra crimes against B's are believed to be committed by nonextremist A's who are angered by the dishonesty of B's, then there must be more dishonesty among B's than previously believed (or than would be inferred by a Bayesian).

Of course, there are costs involved in making this type of biased inference. If one believes that B's are more dishonest than they truly are, that will discourage interaction with them, leading to foregone profitable trading opportunities. However, if a be-

lief in a just world is psychologically valuable, then there is reason to believe the BJW will persist over time. Also, if the crimes are all solved and the motivations of perpetrators revealed, then it is not possible to make this biased inference. However, in practice, some crimes remain unsolved, and motivation is not always clear, even for solved crimes.

Given a biased inference of the type described above, consider the observer's decision to commit future crimes against B's. Under the assumptions we described in Section 3, the more dishonest B's are believed to be, the lower the psychological costs associated with committing crimes against them. This could be because the offender finds it easier to justify the attack to herself, or because the offender expects less social disapproval from others. In either case, the net benefits to A's of committing crimes against B's will increase, leading to more crimes against B's in the future.

7. Some Implications

The most important implication that follows from the analysis sketched above concerns penalty enhancements for hate crimes. Penalty enhancements raise the sanctions imposed on hate-motivated offenders only (it is assumed that courts can distinguish between hate-motivated and non-hate-motivated offenders, or at least can do so with some probability greater than would result from random chance). These higher sanctions deter at least some hate-motivated crimes (by increasing the expected costs of crime for hate-motivated offenders), and so reduce the extent of disproportionate victimisation of the disfavoured group. Thus, in step (2), the extent of the observer's »surprise« (on observing a rate of victimization that exceeds the expected rate) is smaller. In step (3), observers can maintain their belief in a just world while making a smaller adjustment to their moralistic beliefs. Thus, the eventual revised moralistic beliefs about the disfavoured group are less negative than would be the case in the absence of hate crime penalty enhancements. Consequently, the increase in the expected benefits of crimes (and hence in the number of crimes) against the disfavoured group will be smaller, the larger the penalty enhancement. When observers are subject to the JWB, it follows that penalty enhancements can reduce the social costs associated with the additional crimes generated by hate crimes.

We are not claiming that JWB is a necessary condition for penalty enhancements. The mere existence of discriminatory preferences might justify them, as argued elsewhere.[24] Nevertheless, the JWB provides a dynamic justification for imposing penalty enhancements in the present in order to reduce victimization in the future. In addition, while our thesis provides an efficiency rationale for hate crime penalty enhancements, other legal economists have highlighted concerns relating to interest group politics as an explanation for the enactment of hate crime laws.[25] It may well be that as an empirical matter hate crime legislation is best explained by interest group lobbying. However,

24 Dharmapala and Garoupa, *supra* note 16.

25 See R. Posner *Frontiers of Legal Theory* Cambridge, MA: Harvard University Press (2001) at 233.

our thesis provides grounds to believe that there are important efficiency issues that should not be neglected in a comprehensive economic analysis of hate crime statutes.

The mechanism outlined above depends crucially on 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): knowing that haters have discriminatory preferences is in itself sufficient to explain their crimes. Rather, the bias entails attributing a larger proportion of unsolved crimes to nonhaters than would an unbiased Bayesian.

A further implication is that any reduction in the level of uncertainty (such as an increase in the fraction of crimes that are solved) will reduce the scope of the inferential bias. When crimes are known to have been committed because of hate motivation, they cannot be attributed to nonhating offenders motivated by moralistic beliefs (and hence »explained« by the supposed negative characteristics of the disfavoured group). There is some evidence that hate crime statutes may increae the probability of detection 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.[26] This idea also highlights the potential importance of laws that force the revelation of hate motives (for instance, through inquiries related to whether penalty enhancements should be applied).

The provision of information about the role of hate motivation in the victimization of minority groups can also reduce the inferential bias. Consistent with this notion is the observation that human rights organizations and NGOs 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, for example, to the negative characteristics of the victimized group). Thus, such publicity tends to counteract the inferential bias, by providing information that makes observers less likely to attribute observed hate crimes to moralistic beliefs rather than to discriminatory preferences.

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 B occur in their community may infer from the passage of a hate crime statute that such crimes in fact occur, and that members of group B suffer disproportionate victimization. This, in turn, may lead to negative moralistic beliefs about group B through the JWB, an effect that may partially counteract the deterrent effect from the hate crime statute. This would be less relevant in situations where there has already been considerable publicity about hate crimes. However, where there has been no such publicity (and particularly where there have been no hate crimes), purely symbolic legislation may be counterproductive.[27]

An important caveat is that, to date, none of the experiments on belief in a just world uses the distinctive methods of experimental economics. These methods involve tests

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

of behaviour in settings where subjects are provided with monetary incentives. One may, for example, imagine subjects being exposed to an individual's victimisation and then placed 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 are raised by the JWB).

It is also important to note that while our analysis focuses on crime, the basic mechanism could apply more generally to social or economic discrimination. Moreover, our analysis can also be related to one of the pioneering ideas in the sudy of discrimination, Myrdal's »vicious cycle«:[28] when discriminination against a disfavoured group leads to worse outcomes for that group, these outcomes are then viewed by members of the dominant group as evidence of the disfavoured group's intrinsic negative characteristics, leading to more discrimination, even worse outcomes for the disfavored group, and even more negative inferences. This notion is inconsistent with Bayesian rationality, as it assumes that the effects of discrimination on the disfavoured group's outcomes are ignored. However, it is consistent with the existence of a just world bias, where unjust outcomes are attributed to the negative characteristics of the disfavoured group, rather than to the discriminatory preferences of the dominant group.

In conclusion, we have sought to provide a brief introduction to the application of behavioural economics to the study of hate crime. We believe that this is an area in which there is great potential for behavioural-economic analysis to shed light on issues that have proved puzzling from the standpoint of more traditional paradigms of legal scholarship. For instance, related research[29] has explored the links between hate speech and hate crime, analyzing how cognitive biases may influence whether speech that is hostile towards minority groups leads to violent crime against members of those groups. We hope that the applications discussed above demonstrate the potential power of behavioural-economic analysis, and that future research using this approach contributes to illuminating important legal issues.

27 See D. Dharmapala and R. McAdams »The Condorcet Jury Theorem and the Expressive Function of Law: A Theory of Informative Law« 5 *American Law and Economics Review* 1 (2003) for a wider discussion of the informational effects of legislation.

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

29 Dharmapala and McAdams, *supra* note 7.