CRIMINAL JUSTICE POLICY PREFERENCES: BLACKSTONE RATIOS AND THE VEIL OF IGNORANCE

Nicholas Scurich*

INTRODUCTION

The criminal justice system—like any system that involves human judgment and decision making—is ineluctably fallible. Two different types of errors can occur during the administration of criminal justice: a false positive (i.e., convicting a factually innocent person) and a false negative (i.e., acquitting a factually guilty person). These errors are in tension, as reducing one typically results in increasing the prevalence of the other.¹ Thus, a sound criminal justice policy fundamentally entails striking an acceptable tradeoff between the two types of errors based on their respective costs.

This Article considers this debate's historical context, then reports the results of an original empirical study that elicited criminal justice error preferences from an online sample of over 500 adult United States citizens. Consistent with previous research, participants were asked, as a matter of policy, which type of error is worse and to what extent that type of error is worse than the other. Participants’ error preferences were then elicited beneath a Rawlsian veil of ignorance.² The veil of ignorance “universalizes” judgments by forcing individuals to evaluate a policy without knowledge of how it would affect him or her personally. Thus, participants were also asked whether they would personally prefer to endure the consequences of an erroneous conviction vis-à-vis an erroneous acquittal. When framed this way, a non-trivial number of participants switched their previously-stated criminal justice policy preferences.

* Assistant Professor, Department of Criminology, Law & Society, Department of Psychology & Social Behavior, School of Law. Ph.D. Psychology, University of Southern California, 2012.


2. See infra note 17 and accompanying text.
In the debate on the merits between the two types of errors, there is no shortage of opinion on what is an acceptable tradeoff. In 1476, Fortescue seminally posited that twenty false negatives are equal in cost to one false positive. Almost 200 years later, Hale opined that the number is five, and shortly thereafter, the eighteenth-century jurist William Blackstone famously propounded that “[B]etter that ten guilty persons escape, than that one innocent suffer.” Even though jurists of the nineteenth century raised the number considerably, with Starkie notably holding that the proper ratio is ninety-nine to one, Blackstone’s ratio of ten-to-one is the most popular and persistent. Justice Blackmun even characterized it as “perhaps not an unreasonable assumption.”

A false positive has been historically considered more costly because it violates the social contract between the state and the individual, erodes the legitimacy of the justice system, undermines the deterrent effect of punishment, and gratuitously imposes the monetary cost of punishment. However, Blackstone’s ratio must be understood within the context of the eighteenth-century criminal justice system in which a conviction typically resulted in death. It is not clear that such policies would apply to noncapital cases. Moreover, others have argued that false negatives are highly undesirable because releasing a factually-guilty criminal renders him free to reoffend and further victimize the citizenry. Indeed, Jeremy Bentham thought that the “sentimental exaggerations [about the costs of false positives relative to false negatives] tend to give crime impunity, under the pretext of insuring the safety of innocence.”

Although erudite jurists have a strong preference for false negatives relative to false positives, it is not clear that their sentiment is shared by the lay

4. JOHN FORTESCUE, DE LAUDIBUS LEGUM ANGLIE 63 (1616).
6. 4 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 352 (1769).
13. See generally Erik Lillquist, Absolute Certainty and the Death Penalty, 42 AM. CRIM. L. REV. 45, 46 (2005) (exploring the assumptions underlying arguments that a higher level of certainty is required of jurors in capital cases).
15. Id. at 65.
public, an important stakeholder group in a representative democracy. Empirical studies suggest that other populations—typically college undergraduates—prefer a ratio that is closer to equipoise. That is, college undergraduates do not necessarily consider false positives to be the greater evil, and to the extent they do, false positives are not considered much worse than false negatives. Whether these results generalize to the population at large is an unknown empirical question.

There is an additional reason to question the results of previous studies which have, more or less, simply asked participants, “which error is worse and how much so?” In particular, it is possible that a perspective bias may have affected the results. John Rawls proposed that policy arguments and conceptions of justice ought to be evaluated beneath a “veil of ignorance.” The veil of ignorance removes personal considerations (or biases) by masking one’s position in society. As Rawls noted, “if a man knew that he was wealthy, he might find it rational to advance the principle that various taxes for welfare measures be counted unjust; if he knew that he were poor, he would most likely propose the contrary principle. To represent the desired restrictions, one imagines a situation in which everyone is deprived of this sort of information.” Thus, the veil “universalizes” the judgment in that it forces an individual to apply to himself the same standards he would apply to others.

The veil of ignorance is important and relevant to deciding criminal justice policy for two reasons. First, it reifies the issue by forcing people to confront the reality that any policy choice has a corollary. As noted by Allen and Laudan, virtually all academic discussion has focused on mitigating the occurrence of false positives, as if this has no effect on false negatives, which are usually omitted completely from policy discussion. But decreasing false positives comes at the expense of increasing false negatives. It is simply incoherent to ignore this nexus in contemplating policy.

Second, and with the first reason in mind, it is possible that knowledge of one’s position in social milieu could influence (bias) judgments. It is one thing to claim that false positives are far less desirable than false negatives; however, it is unclear that one would still hold such a preference if one were personally forced to endure the consequences of a false negative (e.g., being victimized).


18. Id. at 18.


20. Laudan, *supra* note 1, at 285 (demonstrating this logical consequence mathematically). As an alternative way to motivate this logic, consider the following thought experiment: suppose one wanted to ensure that false positives never occur. Perhaps the only practical way to achieve this end is by never convicting a single criminal defendant. This would ensure that no innocent person is ever convicted of a crime, but it also would ensure that no (factually) guilty person is convicted either.
rather than a false positive (i.e., wrongfully convicted). Thus, a central question of the current study is whether criminal justice policy preferences are affected by the veil of ignorance.

Mossman and Hart utilized this approach in eliciting the policy preferences regarding involuntary civil commitment. Participants were asked whether “they would prefer being attacked by a man with a knife, or spending a certain time period as a patient in a state psychiatric hospital.” Over one fourth of participants, who were university undergraduates and medical students, indicated a preference for being attacked relative to being involuntarily hospitalized for three days. The finding that at least some medical students would prefer to be attacked rather than hospitalized is potentially troubling; it indicates that they are personally averse to the treatment they provide to other individuals as a matter of course. In some sense, this finding supports the ability of the veil of ignorance to remove perspective biases from policy judgments, since it masks whether one is ordering or receiving the treatment when assessing the relative costs and benefits of involuntary psychiatric commitment. However, it should be noted that the study did not directly test whether the veil of ignorance did in fact affect the stated policy preferences.

I. AN EMPIRICAL STUDY

The purpose of the current study is twofold. The first is to ask a sample of adults about their policy preferences regarding the criminal justice system. Previous studies are either normative in nature (i.e., what jurists think the tradeoff ought to be) or have utilized potentially unrepresentative samples (e.g., university undergraduates). The second purpose is to examine whether the veil of ignorance affects participants stated policy preferences. No research has utilized the veil of ignorance approach within the criminal justice context, nor has any research generally examined whether the veil of ignorance will lead to within-participant preference reversals. It is hypothesized that the veil of ignorance will indeed lead to policy preference reversals, such that false negatives (i.e., acquitting the factually guilty) will be preferred to false positives (i.e., convicting the factually innocent) as a policy matter, yet, as a personal matter, participants will prefer to endure the consequences of a false positive (e.g., imprisonment) relative to a false negative (e.g., victimization).

23. Id. at 186.
A. Participants

Five hundred sixty-eight participants were recruited through Amazon’s Mechanical Turk (AMT). In short, AMT provides a platform to post advertisements for tasks that involve human judgment. Common tasks include surveys, questionnaires and other types of market-research items. Workers view the available tasks and associated compensation and decide whether or not they want to participate. If they satisfactorily complete the task, they are compensated for the participation. It is not uncommon to use AMT to conduct psychological and economic research.

Participants were eligible if they were over the age 18 years old and a United States citizen. The median age of the sample is 28 (IQR = 13). It is comprised of 41% (n = 231) females. 38% (n = 213) of participants identify as a Democrat, 13% (n = 74) identify as a Republican, 35% (n = 199) identify as independent, and the rest indicated some “other” type of political affiliation. Most participants (69%, n = 390) indicated that they do not consider themself a religious person. Using a 9-point scale from extremely liberal (1) to extremely conservative (9), 15% (n = 85) consider themselves extremely liberal whereas 3% (n = 19) consider themselves extremely conservative, with a median score of 3 (IQR = 3).

B. Procedure

Participants first read an informed consent sheet which described the nature of the study. If they agreed, participants then clicked through a series of webpages. Participants were first asked about their preferences as a matter of policy. This condition is referred to as the policy frame. Specifically, participants were asked, “With respect to the criminal justice system, which error do you consider worse: a.) an erroneous conviction (i.e., convicting an innocent person) or b.) an erroneous acquittal (i.e., releasing a guilty person)?” Depending on their responses, participants were then directed to another page in which they were asked to indicate how many of the other errors they would tolerate in order to avoid just a single of the more grievous error. For example, if a participant indicated that an erroneous conviction, or false positive, is worse, then she was asked how many erroneous acquittals, or false negatives, she would tolerate in order to avoid just a single false positive. This rating was made on a 21-point ordinal scale ranging from “0” to “1000 or more.”

25. Id.
26. See generally Peter Deneef & Daniel L. Kent, Using Treatment-Tradeoff Preferences to Select Diagnostic Strategies Linking the ROC Curve to Threshold Analysis, 13 MED. DECISION MAKING 126, 126-132 (1993); Douglas Mossman & Eugene Somoza, Balancing Risks and Benefits: Another Approach to Optimizing Diagnostic Tests, 4 J.
The methodological approach used by Mossman and Hart was adapted to elicit policy preferences beneath the veil of ignorance. This condition is referred to as the personal frame. In this condition, participants were asked, “Would you personally rather be the victim of a violent assault or wrongfully convicted of violent assault?” If they indicated the former, they were taken to another webpage and asked, “How many days would you be willing to spend in prison in order to avoid being the victim of a violent assault?” whereas if they indicated the latter they were asked, “How many times would you be willing to be violently assaulted in order to avoid spending just a single day in prison?” Again, ratings were made on the 21-point ordinal scale. Participants then provided demographic information and were thanked for their participation.

C. Blackstone Ratios as a Matter of Policy (Policy Frame)

With respect to the policy frame, a prominent majority (85%, n = 485) of participants indicated that false positives are worse than false negatives. A binary logistic regression was conducted using demographic variables (i.e., gender; age; political affiliation; political beliefs) to predict the preference for false positives or false negatives. It indicated that only gender and political affiliation were related to policy preferences ($\chi^2 = 30.07$, d.f. = 9, p < .001). Females were twice as likely to prefer false positives to false negatives (exp(b) = 2.09, 95%CI[1.27, 3.46], wald = 8.33, d.f. = 1, p < .01), and higher conservatism was related to a preference for false positives to false negatives (exp(b) = 1.28, 95%CI[1.06, 1.55], wald = 6.30, d.f. = 1, p < .05).

Figure 1 is a histogram of responses to the question of how much worse participants consider a type of error, conditional on their error preference. For example, if a participant indicated that a false positive is worse than a false negative, she then indicated how much worse which would be depicted in a gray bar. Note that the scale on the Y-axis is different between the two plots.


28. Note that it is not being claimed that a false negative automatically results in a violent assault. In other words, acquitting a factually-guilty person does not imply that he will commit another offense. He might commit a less heinous offense, a more heinous offense or no additional offense whatsoever. This cannot be known without the passage of time. Nevertheless, for the sake of concreteness in comparison between the two types of errors, the crime is held constant (i.e., violent assault).
Figure 1. Histograms of Blackstone Ratios Elicited Using the Policy Frame.

For participants who consider false positives worse, the median response is that 20 false negatives are equal to one false positive. For participants who consider false negatives worse, the median response is that 5 false positives are equal to one false negative. A Mann-Whitney U Test indicated that the two distributions (at each category) are significantly different (p < .001).

D. Blackstone Ratios as a Personal Matter (Personal Frame)

With respect to the personal frame, 75% (n = 424) of participants indicated that they would personally prefer to endure the consequences of a false negative relative to a false positive. A logistic regression using the above-noted
demographic variables to predict the preference for false positives or false negatives found that only gender was statistically significant ($\chi^2 = 21.46$, d.f. = , p < .01), with females being almost twice as likely to prefer false positives to false negatives ($\exp(b) = 1.94$, 95%CI[1.23, 2.91], wald = 10.41, d.f. = 1, p < .001).

Figure 2 displays the frequency of responses to the question of how much worse participants consider a type of error, conditional on their error preference.

For participants who would prefer the consequences of a false negative, the median response is that 5 false negatives are equal to one false positive. In
other words, they prefer to be violently assaulted 5 times than spend a single day in prison. For participants who would prefer to be a false positive, the median response is that they would prefer to spend 30 days in prison than be violently assaulted. A Mann-Whitney U Test indicated that the two distributions (at each category) are significantly different ($p < .001$).

E. The Stability of Preferences Across the Two Frames

This analysis examines whether the preferences of participants were consistent across the two different frames (i.e., the policy frame and the personal frame). The frequency of responses to both frames is presented in Table 1.

Table 1. Participants’ ($n = 568$) error-type preferences decomposed by the frame of question

<table>
<thead>
<tr>
<th></th>
<th>Personal Frame</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>False Negative</td>
<td>False Positive</td>
</tr>
<tr>
<td>False Positive</td>
<td>380</td>
<td>105</td>
</tr>
<tr>
<td>False Negative</td>
<td>44</td>
<td>39</td>
</tr>
<tr>
<td>(total)</td>
<td>424</td>
<td>144</td>
</tr>
</tbody>
</table>

The shaded cells correspond to participants (74%, $n = 419$) who were consistent in their preferences across the two frames. For example, participants in the upper left cell indicated that false positives are worse under the policy frame and that they personally would rather endure the consequences of a false negative than suffer the wrath of a false positive. The non-shaded cells (26%, $n = 149$) correspond to participants who changed their preference between the two frames. For example, 22% ($n = 105$) of participants who indicated that false positives are worse than false negatives in the policy frame, indicated that they would personally rather experience the consequences of a false positive than a false negative. In other words, they considered an erroneous conviction to be worse than an erroneous acquittal, yet they personally would rather be imprisoned than violently attacked.

A logistic regression was conducted to examine whether the above-noted demographic variables could predict a preference reversal between the two frames. It detected two significant effects ($\chi^2 = 16.83$, d.f. = 5, $p < .01$). Females were one and a half times more likely to switch their preference between the two frames ($\exp(b) = 1.55$, 95%CI[1.04, 2.33], $\text{wald} = 4.70$, d.f. = 1, $p < .05$). Political conservatism was also related to an increase in the likelihood of a preference reversal ($\exp(b) = 1.14$, 95%CI[1.01, 1.28], $\text{wald} = 4.71$, d.f. = 1, $p < .05$).
II. DISCUSSION

The results of the present study indicate that there is not unanimity in policy preferences of the criminal justice system. A majority of participants indicated that false positives are less desirable than false negatives, which is consistent with a bevy of legal scholarship. On the issue of how much worse, there is considerable variability even amongst those who agree that one type of error is worse than the other. Indeed, responses spanned the entire ranges of possibilities. One implication is that even a democratically-elected criminal justice policy is likely to not sit well large segments of the populous, which could present problems for the legitimacy of the law.  

On a more practical level, this tremendous variability is also likely to create inconsistencies in the administration of criminal law. The tradeoff between false positives and false negatives is linked to the beyond a reasonable doubt standard of proof, which is constitutionally required in all criminal trials. As Justice Harlan noted in his oft-cited concurrence, “I view the requirement of reasonable doubt in a criminal case as bottomed on a fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free.” Harlan’s commentary stopped noticeably short of specifying how much worse, presumably relegating that duty to the jury as representatives of the community.

The translation of Blackstone-type ratios into standards of proof is complex but not intractable. One approach is to treat the particular Blackstone ratio as an odds ratio which indicates the minimum necessary odds (posterior odds) of guilt required in order to convict. For instance, a Blackstone ratio of 10:1 requires that the posterior odds of guilt exceed 10:1 or about 91% to warrant a conviction.  

Note that this approach does not guarantee that the observed

29. Tyler, supra note 10, at 19.
30. See In re Winship, 397 U.S. 358, 371 (1970) (Harlan, J., concurring) (“The standard of proof influences the relative frequency of these two types of erroneous outcomes. If, for example, the standard of proof for a criminal trial were a preponderance of the evidence rather than proof beyond a reasonable doubt, there would be a smaller risk of factual errors that result in freeing guilty persons, but a far greater risk of factual errors that result in convicting the innocent. Because the standard of proof affects the comparative frequency of these two types of erroneous outcomes, the choice of the standard to be applied in a particular kind of litigation should, in a rational world, reflect an assessment of the comparative social disutility of each.”).
31. Id.
32. Id. at 372.
35. To convert odds to probability take the first number, i.e., 10, and divide it by the sum of the first and second numbers, i.e., 10+1, which equals 10/11 or ~0.91.
frequency of errors will necessarily resemble any particular Blackstone ratio; the actual frequency of errors depends on the base rate of guilty defendants who are brought to trial. Nevertheless, this approach provides a useful heuristic for understanding the relation between Blackstone ratios and legal standards of proof.

The most conspicuous manifestation of the variability observed in the present study is likely to be seen when jurors apply the reasonable doubt standard. Survey research finds a similarly high degree of variability when participants are directly asked about the level of confidence that the reasonable doubt rule requires. For example, Catherine McCauliff asked a sample of 171 federal judges about the level of certainty that the reasonable doubt rule requires. The responses ranged from 100% all the way to only 50% with a mode of 90%. Although there is reason to question these values because of the elicitation approach, they are consistent with and can be potentially explained by the present findings. Specifically, the present findings indicate that jurors might operationalize reasonable doubt differently because they have different underlying Blackstone ratios regarding the appropriate tradeoff between false positives and false negatives, and different tradeoffs imply different levels of requisite certainty. For instance, a 5:1 Blackstone ratio requires certainty in excess of 83% whereas the 10:1 ratio requires certainty in excess of 91%. Thus, even if jurors agree on the (probative) value of evidence proffered at trial, they could reach different verdicts based solely on differences in their threshold for conviction, deliberation notwithstanding.

Several caveats of the present research must be acknowledged. The findings reported in this Article are based on a sample that is not statistically representative of the U.S. population. The extent to which the preferences generalize to the population at large is an open empirical question and therefore the results should be considered provisional until such an undertaking occurs. Further studies might utilize different response scales, as well as different crimes when eliciting policy preferences. On this latter point, the crime of violent assault was selected because it is both a serious crime and it is tenable for the purposes of the veil of ignorance. Some crimes are simply incompatible with the veil of ignorance. For instance, there is no meaningful sense in which one can prefer to be wrongfully convicted of murder than to be murdered. Finally, empirical research finds that jurors do bring to bear different standards of proof (and by implication underlying policy preferences) depending on the

36. Dekay, supra note 33, at 121.
37. For a review of this research, see generally Hastie, supra note 16, at 102.
39. Id.
40. See Barbara D. Underwood, The Thumb on the Scales of Justice: Burdens of Persuasion in Criminal Cases, 86 YALE L.J. 1299, 1309-10 (1977) (presenting evidence that jurors conceive “reasonable doubt” differently depending on the term that is presented to them).
specifics of the crime and what is involved for both those who might be erroneously convicted or erroneously acquitted. Thus, one would hypothesize that the actual policy preferences would vary as a function of the particular consequences associated with the alleged crime. Further empirical research is needed to directly test this hypothesis.

Although policy preferences (and standards of proof) might vary between crimes de facto, it is not clear that this variability is legally permissible. The United States Supreme Court has never stated that the beyond a reasonable doubt standard of proof applies with lesser force to less serious crimes, but some Justices have certainly considered this possibility. In addition, some commentators have argued that the nebulous instructions provided to jurors on the standard of proof and the Court's continual refusal to elaborate on this topic signals implicit approval of jurors applying their own policy preferences. Indeed, some legal scholars argue that such variability is both legally permissible and desirable. Others question whether question whether jurors should be the arbitrators of social policy, and instead believe that jurors ought to simply heed the guidance contained in jury instructions. The present findings cannot resolve the ongoing debate about variability in the application of legal standards, but they do help to reify discussions about what it means for jurors to apply different standards of proof. Moreover, the effect of the veil of ignorance suggests that some jurors might apply a different standard of proof to others than they would apply to themselves.

41. See Anne W. Martin & David A. Schum, Quantifying Burdens of Proof: A Likelihood-Ratio Approach, 27 JURIMETRICS J. 383, 397-98 (1986) (finding that participants required a higher level of certainty in order to convict as the severity of the associated punishment increased).
42. See In re Winship, 397 U.S. 358, 371 (1970) (Harlan, J., concurring) (“The choice of the standard to be applied in a particular kind of litigation should, in a rational world, reflect an assessment of the comparative social disutility of each.”).
43. See generally Henry A. Diamond, Reasonable Doubt: To Define, or Not to Define, 90 COLUM. L. REV. 1716 (1990) (detailing the ongoing dispute among the courts whether or not to define “reasonable doubt” in the post-Winship era).
44. See Lilquist, supra note 12, at 92 (advocating for a “flexible” theory of reasonable doubt which “views the reasonable doubt standard as attempting to strike a balance between the utilities and disutilities associated with errors”); Elisabeth Stoffelmayr & Shari S. Diamond, The Conflict Between Precision and Flexibility in Explaining Beyond a Reasonable Doubt, 6 PSYCHOL. PUB. POL’Y & L. 769, 778 (2000) (arguing that “the reasonable doubt instruction should leave room for flexible tailoring of the standard to the costs of error”); see also Alan Dershowitz, When Are Doubts Reasonable?, in BEYOND A REASONABLE DOUBT 21 (Larry King ed., 1996) (“For me the reasonableness of the doubt required to acquit should depend on the seriousness of the crime and the severity of the punishment. No doubt is reasonable if the punishment is death. Very little doubt should be deemed reasonable if the punishment is imprisonment. But if the punishment is merely a fine or a suspended sentence, the required degree of doubt might be greater.”).
45. Larry Laudan & Larry D. Saunders, Re-thinking the Criminal Standard of Proof: Seeking Consensus About the Utilities of Trial Outcomes, 7 INT’L COMMENT. EVIDENCE 1, 32 (2009) (advocating for a more stringent “quantified standard of proof”).
A comprehensive exposition of the Rawlsian thesis about the veil of ignorance would require space far beyond that which is permitted here. The gist of Rawls’ contention is that people ought to apply to themselves the standard they apply to others. With this in mind, the present study found that a non-trivial number of participants were not consistent in their preferences. What they consider preferable from a policy perspective switched when the corollary of that policy was applied to them. Whether or not one accepts the Rawlsian thesis as normative, the finding that the veil of ignorance can induce policy preference reversals is relevant generally to any policy discussion, and it is suggested that this methodology be employed when empirically examining policy preferences in other domains.

A. Final Thoughts

Legal scholars Allen and Laudan describe a truism that is often lost in discussions about the policies governing the criminal justice system:

The core tension in the design of a system of crime and punishment is that the two obligations to protect against crime and false convictions pull in contrary directions. Many of the methods that can control crime put innocent persons at the risk of false conviction. Likewise, many of the remedies for reducing the number of false convictions increase the risk of false acquittals, and with it, the risk of rising crime victimization.46

The present study does not provide any sharp resolution to the issue of what United States citizens think the appropriate tradeoff ought to be in the context of criminal justice. Indeed, the responses were highly variable. But the methodology—particularly the veil of ignorance—highlights the reality that a choice of any specific policy has consequences, some of which may be undesirable. Confronting rather than eschewing this reality is necessary for a coherent discussion of criminal justice policy.

46. Allen & Laudan, supra note 14, at 84.