

10-1-2012

## The Plasticity of Harm in the Service of Criminalization Goals

Avani Mehta Sood  
*Berkeley Law*

John M. Darley  
*Princeton University*

Follow this and additional works at: <https://scholarship.law.berkeley.edu/californialawreview>

 Part of the [Law Commons](#)

---

### Recommended Citation

Avani Mehta Sood and John M. Darley, *The Plasticity of Harm in the Service of Criminalization Goals*, 100 CALIF. L. REV. 1313 (2012).

### Link to publisher version (DOI)

<https://doi.org/10.15779/Z38BV51>

This Article is brought to you for free and open access by the California Law Review at Berkeley Law Scholarship Repository. It has been accepted for inclusion in California Law Review by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact [jcera@law.berkeley.edu](mailto:jcera@law.berkeley.edu).

# The Plasticity of Harm in the Service of Criminalization Goals

Avani Mehta Sood\* & John M. Darley\*\*

*The “harm principle,” which suggests that the State should criminalize conduct only if doing so is necessary to prevent harm to others, has long been at the center of debates about the legal enforcement of morality through the law. This Essay presents a series of original experiments that investigate the psychological “plasticity” of harm in the context of criminalization—i.e., the extent to which people may recruit harm to achieve desired outcomes within the terms of a given legal constraint. In the first study, we identified various scenarios that induced defiance of the harm principle, such that the participants wanted to criminalize the conduct even if they did not think that it caused harm to others. We then used these scenarios in a second study to investigate how people would respond when they were told that the law requires a finding of harm in order to impose a criminal penalty. As predicted, the respondents who were presented with the necessity-of-harm constraint continued to criminalize at the same rate, but they imputed harm to the conduct where harm was not previously reported. The third study demonstrated the directionality and universality of the harm plasticity effect by showing that a legally extrinsic ideological factor exacerbated the recruiting of harm among people on both sides of a*

---

Copyright © 2012 Avani Mehta Sood.

\* Doctoral candidate in the Department of Psychology at Princeton University (M.A. Princeton University, 2011; J.D. Yale Law School, 2003; B.A. Princeton University, 1999).

\*\* Dorman T. Warren Professor of Psychology and Public Affairs at Princeton University (Ph.D. Harvard University, 1965; M.A. Harvard University, 1962; B.A. Swarthmore College, 1960).

For comments on earlier drafts, we are grateful to Dan Kahan, Larry Solan, Kim Lane Scheppele, Joel Cooper, Donald Braman, Arthur Leonard, Philip Johnson-Laird, and members of the Darley lab at Princeton University. In addition, Ms. Sood received valuable feedback from participants at the Law and Society Conference in San Francisco (2011); a Law and Public Affairs Program seminar at Princeton University (2011); a Law and Cognition seminar at Harvard Law School (2011); the Conference on Empirical Legal Studies at Yale Law School (2010); and a Law, History, and Society Colloquium at New York Law School (2010). We thank Alison Newman and Brittny Delgado for their help with data collection, Elizabeth Ingriselli for her research assistance, and the excellent editorial team of the *California Law Review*.

Please address correspondence to Avani Mehta Sood, Department of Psychology, Princeton University, Princeton, NJ 08540; e-mail: [avanim@princeton.edu](mailto:avanim@princeton.edu).

*controversial issue (abortion), without their recognition. The final study provided evidence for the nonintentional nature of this effect and pointed toward a pathway to potential remedies. These experiments reveal motivated, outcome-driven perceptions of harm that are at odds with the objective, perception-driven outcomes toward which the legal system strives. We discuss the implications of our findings for the criminal regulation of morality, constitutional law, and theories of punishment and moral reasoning. Furthermore, we consider the application of our results to legal decision making by lay people and professional adjudicators in the context of two high-profile Ninth Circuit cases. We conclude with a discussion of how to potentially reduce motivated cognition in legal judgments.*

Introduction.....	1315
I. The Harm Principle in the Law and Courts.....	1317
A. Historical and Theoretical Underpinnings.....	1317
B. The Evolving Faces of Harm.....	1319
II. Psychological Mechanism.....	1320
A. The Theory of Motivated Cognition.....	1320
B. The Social Intuitionist Model.....	1322
C. Motivated Cognition in Legal Judgments.....	1323
III. Experimental Studies.....	1324
A. Study 1: Mapping Harm.....	1325
1. Methodology.....	1325
2. Participants.....	1325
3. Results.....	1326
a. Compliance with the Harm Principle.....	1326
b. Defiance of the Harm Principle.....	1326
B. Study 2: The Plasticity of Harm.....	1328
1. Methodology.....	1328
2. Participants.....	1330
3. Predictions.....	1330
4. Results.....	1331
a. The Rigidity of Criminalization.....	1331
b. The Plasticity of Harm.....	1332
C. Study 3: Motivating Role of Ideological Incongruency.....	1336
1. Methodology.....	1337
2. Participants.....	1338
3. Predictions.....	1338
4. Results.....	1339
a. Enhanced Punishment.....	1339
b. Imputed Harm.....	1339
c. Individual Difference Findings.....	1340
D. Study 4: Confronting the Illusion of Objectivity.....	1342

1. Methodology.....	1342
2. Participants.....	1343
3. Predictions .....	1343
4. Results.....	1343
IV. Discussion.....	1345
A. Review of Findings.....	1345
B. Implications for Harm and Beyond.....	1346
C. Constitutional Consequences.....	1348
D. Contribution to Theories of Punishment and Moral Reasoning ...	1349
E. Legal Applications .....	1350
1. Jury Decision Making.....	1350
2. Voter-Based Initiatives .....	1352
3. Judicial Safeguards?.....	1355
F. Potential Remedies.....	1357
Conclusion .....	1357

#### INTRODUCTION

“[T]he only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others.”

—John Stuart Mill, *On Liberty* (1859)

On Election Day 2008, citizens of California who entered the polling booths were asked to cast their votes on Proposition 8—a ballot initiative that sought to amend the California Constitution by eliminating the right of same-sex couples to marry.<sup>1</sup> In the months leading up to the vote, proponents of Proposition 8 had executed an intensive “Protect Marriage” advertising campaign that emphasized “the concern that people of faith and religious groups *would somehow be harmed* by the recognition of gay marriage” and that “children *need to be protected* from exposure to gay people and their relationships.”<sup>2</sup> Proposition 8 passed by a narrow margin of 52 percent,<sup>3</sup> and the California Supreme Court upheld the voter-enacted amendment.<sup>4</sup>

---

1. Specifically, Proposition 8 sought to add a new provision to the California Constitution that would recognize “[o]nly marriage between a man and a woman” as valid. CAL. CONST. art. I, § 7.5.

2. *Perry v. Schwarzenegger*, 704 F. Supp. 2d 921, 990 (N.D. Cal. 2010), *aff’d sub nom.* *Perry v. Brown*, 671 F.3d 1052 (9th Cir. 2012) (emphasis added); see *infra* pp. 1352–53 (describing advertising campaign in favor of Proposition 8). Proponents and opponents of the Proposition spent approximately \$40 million each on their respective campaigns—funds that were raised from contributors across all fifty states and over twenty foreign countries—indicating that this was a contentious issue with implications that would extend well beyond the borders of California. SUSAN B. HANSEN, *RELIGION AND REACTION: THE SECULAR POLITICAL CHALLENGE TO THE RELIGIOUS RIGHT 120* (2011).

3. SECRETARY OF STATE DEBRA BOWEN, STATEMENT OF VOTE: NOVEMBER 4, 2008, GENERAL ELECTION 13 (2008), available at [http://www.sos.ca.gov/elections/sov/2008\\_general/sov\\_complete.pdf](http://www.sos.ca.gov/elections/sov/2008_general/sov_complete.pdf).

4. *Strauss v. Horton*, 207 P.3d 48, 122 (Cal. 2009).

Two same-sex couples then filed *Perry v. Schwarzenegger*, a legal action challenging Proposition 8 in federal court, on the grounds that it violated the Due Process and Equal Protection Clauses of the U.S. Constitution.<sup>5</sup> The district court agreed and struck down Proposition 8 as unconstitutional after finding that the harmful consequences of same-sex marriage referenced by the proponents' legal brief and advertising campaign were never legally substantiated at trial.<sup>6</sup> Earlier this year, the Court of Appeals for the Ninth Circuit affirmed this decision, stating: "Proposition 8 operates with no apparent purpose but to impose on gays and lesbians, through the public law, a majority's private disapproval of them and their relationships."<sup>7</sup> The appellants' request for an expanded panel of judges to review this holding was rejected, leading commentators to speculate that the U.S. Supreme Court may be next to address the issue.<sup>8</sup>

The *Perry* case raises important questions about the enforcement of majoritarian morality through the law, and the critical role that "harm" plays in this process. To what extent is the perception of harm psychologically malleable, such that people may less than consciously impute harm to disliked conduct in order to achieve desired outcomes within the terms of given legal constraints? And what cognitive process drives this type of outcome-driven reasoning in legal judgments?

The present research applies experimental methodologies from the field of social psychology to investigate motivated perceptions of harm in the context of regulating morally offensive conduct through punishment. We identified acts that people want to criminalize even without finding harm, to test the following: When people want to penalize such conduct but are told that the law requires a finding of harm in order to impose a criminal penalty, will they recruit harm in order to support their criminalization goals? Our empirical evidence in the affirmative sheds light upon a critical psychological dimension of legal decision making, with significant consequences. Motivated cognition can inadvertently lead to subjective and inconsistent applications of legal rules in ways that undermine fundamental constitutional principles and the rule of law.<sup>9</sup>

This Essay proceeds as follows: Part I traces the evolving role of harm in debates about the legal regulation of morality. Part II discusses the psychological theory of motivated cognition, or motivated reasoning—the cognitive mechanism driving the predicted effect. Part III reports the

---

5. *Perry*, 704 F. Supp. 2d at 921; see U.S. CONST. amend. XIV, § 1.

6. *Perry*, 704 F. Supp. 2d at 931.

7. *Perry v. Brown*, 671 F.3d 1052, 1095 (9th Cir. 2012).

8. See Ethan Bronner, *Same-Sex Marriage Issue Moves Closer to Justices*, N.Y. TIMES, June 6, 2012, at A17.

9. See U.S. CONST. amends. I, XIV; *infra* Part IV.C (discussing constitutional implications of present research).

methodologies and results of four original experiments that demonstrate the “plasticity” of harm in the service of people’s criminalization goals—i.e., the extent to which perceptions of harm may be cognitively molded, without full awareness, to achieve desired outcomes within the constraints of a given law. In Part IV, we turn to the legal implications of the research. We discuss the significance of our findings for criminal regulation, constitutional law, and theories of punishment and moral reasoning. We also consider the application of our results to real-life legal decision making by voters, jurors, and judges in the context of two high-profile cases: *Newton v. National Broadcasting Co.*<sup>10</sup> and *Perry v. Schwarzenegger*.<sup>11</sup> We conclude with a brief discussion of potential remedies that may be used to reduce motivated cognition in legal judgments.

## I.

### THE HARM PRINCIPLE IN THE LAW AND COURTS

Famously articulated by John Stuart Mill one and a half centuries ago, the harm principle suggests that the State should use its powers to regulate individual conduct only if doing so is necessary to prevent harm to others.<sup>12</sup> However, the evolving applications of harm arguments in debates about criminal regulation give rise to questions about the psychological and legal stability of the concept of harm, which the present studies seek to explore.

#### *A. Historical and Theoretical Underpinnings*

The harm principle has long been at the center of discourse about the legal enforcement of morality, often referred to as the Hart–Devlin debate, with the central question:

Should the criminal law be limited, as Mill and (with qualification) H.L.A. Hart had argued, by the *harm principle* . . . or could the criminal law in principle be used to enforce *any* of the important moral convictions of the community—even if the targeted conduct . . . was not in any obvious way rights-violative or harmful to others?<sup>13</sup>

In the course of this debate, the harm principle has historically been converted into a “trustworthy weapon in the arsenal of liberalism,”<sup>14</sup> invoked to support the conclusion that behavior the majority considers to be immoral should not be criminalized unless it causes harm to others. Legal moralists, on the other hand,

10. *Newton v. Nat’l Broad. Co.*, 677 F. Supp. 1066 (D. Nev. 1987), *rev’d*, 930 F.2d 662 (9th Cir. 1990).

11. *Perry*, 704 F. Supp. 2d 921.

12. JOHN STUART MILL, ON LIBERTY (1859).

13. Jeffrie G. Murphy, *Legal Moralism and Liberalism*, 37 ARIZ. L. REV. 73, 74 (1995). Compare H.L.A. HART, LAW, LIBERTY AND MORALITY (1963), with PATRICK DEVLIN, THE ENFORCEMENT OF MORALS (1965).

14. Steven D. Smith, *Is the Harm Principle Illiberal?*, 51 AM. J. JURIS. 1, 1 (2006).

have favored greater state regulation of individual conduct based on Lord Devlin's argument that "society may use the law to preserve morality in the same way as it uses it to safeguard anything else that is essential to its existence."<sup>15</sup>

The Model Penal Code, promulgated in 1962 to help standardize criminal codes across the states, reflects the harm principle in its definition of crimes. The Code strives "to forbid and prevent conduct that unjustifiably and inexcusably inflicts or threatens substantial harm to individual or public interests," as well as "to safeguard conduct that is without fault from condemnation as criminal."<sup>16</sup> Following this lead, numerous states have incorporated harm-based definitions of crime into their penal codes, and criminal law casebooks have accorded the harm principle a prominent role as well.<sup>17</sup> Gradually, harm has become "*the* critical principle used to police the line between law and morality within the Anglo-American philosophy of law."<sup>18</sup>

Yet, the Hart–Devlin debate lives on—reflected, for example, in the U.S. Supreme Court's jurisprudence on same-sex sexual activity, when the Court confronted the question of "whether the majority may use the power of the State to enforce [its] views on the whole society through operation of the criminal law."<sup>19</sup> In its 1986 *Bowers v. Hardwick* decision, the Court upheld as constitutional a Georgia statute that criminalized sodomy, stating that laws are "constantly based on notions of morality, and if all laws representing essentially moral choices are to be invalidated under the Due Process Clause, the courts will be very busy indeed."<sup>20</sup> This opinion represented a move in favor of legal moralism over the harm principle; a later Seventh Circuit decision (not a sodomy case) invoked *Bowers* for the premise that "[l]egislatures are permitted to legislate with regard to morality . . . rather than confined to preventing demonstrable harms."<sup>21</sup> As one legal commentator noted, "Had the [Supreme] Court recognized a harm-based account of offense definition as a legitimate way of protecting individual liberty it probably would have found the use of criminal sanction in *Bowers* unjustified."<sup>22</sup>

Seventeen years later, in *Lawrence v. Texas*, the Court overturned its *Bowers* precedent by striking down a Texas statute criminalizing sexual

---

15. DEVLIN, *supra* note 13, at 11.

16. MODEL PENAL CODE § 1.02(a), (c) (1962).

17. *See, e.g.*, PAUL H. ROBINSON, CRIMINAL LAW 131–37 (1997).

18. Bernard E. Harcourt, *The Collapse of the Harm Principle*, 90 J. CRIM. L. & CRIMINOLOGY 109, 131 (1999).

19. *Lawrence v. Texas*, 539 U.S. 558, 571 (2003).

20. *Bowers v. Hardwick*, 478 U.S. 186, 196 (1986).

21. *Milner v. Apfel*, 148 F.3d 812, 814 (7th Cir. 1998).

22. Claire Finkelstein, *Positivism and the Notion of an Offense*, 88 CALIF. L. REV. 335, 380 (2000).

intercourse between individuals of the same sex.<sup>23</sup> “The fact that the governing majority in a State has traditionally viewed a particular practice as immoral is not a sufficient reason for upholding a law prohibiting the practice,” stated the majority, holding that the Due Process Clause gives same-sex couples “the full right to engage in their conduct without intervention from the government.”<sup>24</sup> In reaching this conclusion, the Court cited the Model Penal Code’s recommendation against imposing criminal penalties for private, consensual sexual acts—which was justified, *inter alia*, on the ground that such conduct is not harmful to others.<sup>25</sup> The *Lawrence* decision has been described to have “effectively constitutionalized the harm principle,”<sup>26</sup> although the issue periodically gets reignited in cases pertaining to morally controversial conduct.<sup>27</sup>

### B. *The Evolving Faces of Harm*

Notwithstanding attempts by lawmakers and courts to establish “a general secular harm principle,”<sup>28</sup> the concept of harm has been difficult to pin down. Even before the *Lawrence* decision, implicit expectations of harm-based arguments may have motivated those in favor of broader criminalization to deploy the rhetoric of harm to counter the very anticriminalization stance that the harm principle once supported.<sup>29</sup> For example, the “broken windows theory,” which has been invoked to broaden policing powers, transformed the

23. 539 U.S. 558.

24. *Id.* at 577–78.

25. *Id.* at 572 (citing MODEL PENAL CODE, Commentary 277-80 (Tent. Draft No. 4) (1955)). Dissenting from the majority opinion, Justice Scalia continued to express support for “the ancient proposition that a governing majority’s belief that certain sexual behavior is ‘immoral and unacceptable’ constitutes a rational basis for regulation.” *Lawrence*, 539 U.S. at 589 (Scalia, J., dissenting).

26. Smith, *supra* note 14, at 13; see Dan M. Kahan, *The Supreme Court 2010 Term—Foreword: Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law*, 125 HARV. L. REV. 1, 50 (2011) [hereinafter Kahan, *Foreword*] (“Read for all its worth, this language . . . can be seen as establishing a general *secular harm* principle, one reinforced by the holding of *Romer v. Evans* that the expression of ‘animosity’ toward a group cannot be deemed a ‘legitimate governmental purpose’ under the Equal Protection Clause.”).

27. The issue has been reignited, for example, in cases pertaining to state laws against the sale of sex toys, creating a circuit split among federal courts. Editorial, *The Courts and Privacy*, N.Y. TIMES, Oct. 2, 2009, at A30. Compare *Reliable Consultants, Inc. v. Earle*, 517 F.3d 738, 745–46 (5th Cir. 2008) (holding that “interests in ‘public morality’” alone cannot constitutionally sustain a Texas statute banning the promotion of personal sex devices), with *Williams v. Morgan*, 478 F.3d 1316, 1323 (11th Cir. 2007) (holding that, even after *Lawrence*, “the State’s interest in the preservation of public morality remains a rational basis” for Alabama’s statutory ban on the sale of personal sex devices), and *1568 Montgomery Highway, Inc. v. City of Hoover*, 45 So. 3d 319, 340–41 (Ala. 2009) (upholding an Alabama statute restricting the location of “adult-only enterprises” as rationally related to the State’s interest in preserving public morality).

28. Kahan, *Foreword*, *supra* note 26, at 50.

29. See Harcourt, *supra* note 18, at 109–10 (“Instead of arguing about morals, the proponents of enforcement are talking about individual and social harms in contexts where, thirty years ago, the harm principle would have precluded regulation or prohibition.”).



perception of “untended behaviors,” like loitering and panhandling, from mere nuisances into acts that can cause serious harm by making an area “vulnerable to criminal invasion.”<sup>30</sup>

As observed by Bernard Harcourt:

In contrast to the earlier pairing of harm and legal moralist arguments, or even to the later dominance of the harm argument over legal moralism, today the debate is no longer structured. It is, instead, a harm free-for-all: a cacophony of competing harm arguments without any way to resolve them.<sup>31</sup>

This calls into question the cognitive reliability and legal objectivity of the concept of harm. We suggest that, in the context of punishment, outcome-driven psychological processes may inadvertently drive people’s perceptions of harm.

## II.

### PSYCHOLOGICAL MECHANISM

The psychological theory of motivated cognition, also known as motivated reasoning, can account for how and why judgments about harm might be motivated, without the decision maker’s awareness, in morally loaded contexts. The related social intuitionist model could also provide some useful insights, but it is narrower in scope and differs from motivated cognition in ways that make it less applicable to legal judgments.<sup>32</sup> Understanding the cognitive process underlying the plasticity of harm will clarify the implications of our findings for the realms of both psychology and law.

#### *A. The Theory of Motivated Cognition*

The theory of directional motivated cognition suggests that being motivated to reach a particular conclusion can lead to inadvertently biased

---

30. George L. Kelling & James Q. Wilson, *Broken Windows*, ATLANTIC MONTHLY, Mar. 1983, at 4; see Harcourt, *supra* note 18, at 149–51. According to the broken windows theory, “The unchecked panhandler is, in effect, the first broken window. Muggers and robbers, whether opportunistic or professional, believe they reduce their chances of being caught or even identified if they operate on streets where potential victims are already intimidated by prevailing conditions.” Kelling & Wilson, *supra*, at 6. This theory was operationalized in New York City’s “quality-of-life” initiatives in the 1990s, which entailed aggressive policing of minor misdemeanors in order to reduce more serious crime. See Bernard E. Harcourt & Jens Ludwig, *Broken Windows: New Evidence from New York City and a Five-City Social Experiment*, 73 U. CHI. L. REV. 271, 272 (2006) (noting that New York City Mayor Rudolph Giuliani and his police commissioner William Bratton cited the broken windows theory as the impetus for their quality-of-life initiatives); GEORGE L. KELLING & CATHERINE M. COLES, *FIXING BROKEN WINDOWS: RESTORING ORDER AND REDUCING CRIME IN OUR COMMUNITIES* 137, 149 (1996) (noting that the quality-of-life initiatives were based on the theory that restoring order reduces crime).

31. Harcourt, *supra* note 18, at 119.

32. See Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOL. REV. 814 (2001).

reasoning.<sup>33</sup> This psychological phenomenon has been experimentally demonstrated in a variety of domains, such as in the formation of beliefs about the self, others, and events, as well as in political, business, and legal judgments.<sup>34</sup> Explaining the underlying cognitive process, psychologist Ziva Kunda suggested that people conduct a biased search through their memory for beliefs and rules that help them reach directional goals and “may also creatively combine accessed knowledge to construct new beliefs that could logically support the desired conclusion.”<sup>35</sup>

The motivated cognition process is not, however, without limits. For one, it does not enable decision makers to reach a completely baseless outcome: “people feel a need to be able to justify their judgment to themselves, and possibly to others.”<sup>36</sup> Therefore, as noted by Kunda:

[P]eople motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer. They draw the desired conclusion only if they can muster up the evidence necessary to support it.<sup>37</sup>

In addition, decision makers engage in motivated cognition only to the extent that it is necessary to reach their desired outcome.<sup>38</sup>

Finally, this psychological process operates under an illusion of objectivity, whereby people are unaware of the biases driving their decision making. This distinguishes motivated cognition from the more deliberate analytical malleability that characterizes jury nullification or intentionally biased legal judgments. People do not fully realize that “in the presence of different directional goals . . . they might even be capable of justifying opposite conclusions.”<sup>39</sup>

33. Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCHOL. BULL. 480 (1990).

34. *Id.* See also Charles S. Taber et al., *The Motivated Processing of Political Arguments*, 31 POL. BEHAV. 137 (2009); Eileen Braman & Thomas E. Nelson, *Mechanism of Motivated Reasoning? Analogical Perception in Discrimination Disputes*, 51 AM. J. POL. SCI. 940 (2007); Eileen Braman, *Reasoning on the Threshold: Testing the Separability of Preferences in Legal Decision Making*, 68 J. POL. 308 (2006); Charles S. Taber & Milton Lodge, *Motivated Skepticism in the Evaluation of Political Beliefs*, 50 AM. J. POL. SCI. 755 (2006); David P. Redlawsk, *Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making*, 64 J. POL. 1021 (2002); Mark Fischle, *Mass Response to the Lewinsky Scandal: Motivated Reasoning or Bayesian Updating?* 21 POL. PSYCHOL. 135 (2000); Richard E. Redding & N. Dickon Reppucci, *Effects of Lawyers' Socio-Political Attitudes on Their Judgments of Social Science in Legal Decision Making*, 23 LAW & HUM. BEHAV. 31 (1999); Lindsley G. Boiney et al., *Instrumental Bias in Motivated Reasoning: More When More Is Needed*, 72 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 1 (1997).

35. Kunda, *supra* note 33, at 483.

36. Boiney et al., *supra* note 34, at 19–20.

37. Kunda, *supra* note 33, at 482–83.

38. See Boiney et al., *supra* note 34, at 20.

39. Kunda, *supra* note 33, at 483. See Tom Pyszczynski & Jeff Greenberg, *Toward an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis-*

*B. The Social Intuitionist Model*

Another psychological theory that is arguably relevant to debates about harm in the criminal context is Jonathan Haidt's social intuitionist model, which suggests that "moral judgment is caused by quick moral intuitions and is followed (when needed) by slow, ex post facto moral reasoning."<sup>40</sup> One of Haidt's studies tested whether judgments about victimless sexual taboo violations—such as consensual sex between siblings with multiple forms of protection against pregnancy—are driven by informational assumptions about harmfulness, or rather, by affective reactions such as disgust and discomfort that are "later cloaked by harm-based rationalizations."<sup>41</sup> The researchers found that although harm was often cited as a reason for condemnation, it was negative emotional response, not the perception of harmfulness, that operated as a significant predictor of moral judgment.<sup>42</sup> This result was reinforced by a qualitative finding of moral dumbfounding: "[P]articipants often condemned the scenarios instantly, and then seemed to search and stumble through sentences laced with pauses, 'ums,' and 'I don't know,' before producing a statement about harm."<sup>43</sup>

Like the illusion of objectivity in motivated cognition, the social intuitionist model suggests that people fall prey to a "wag-the-dog" illusion: "We believe that our own moral judgment (the dog) is driven by our own moral reasoning (the tail)," when in reality, our intuitive judgments are driving the reasoning process.<sup>44</sup> However, the social intuitionist model differs from motivated cognition in at least two important ways. For one, motivated cognition is a broader phenomenon that spans various domains of reasoning, whereas the social intuitionist model is specific to the domain of moral judgments.<sup>45</sup> Second, the social intuitionist model does not have the same rational limits as motivated cognition. Haidt's model posits that people put forth "argument after argument, never wavering in the conviction . . . even after one's last argument has been shot down."<sup>46</sup> Thus, the social intuitionist model may be less applicable in legal contexts, which do not always involve morally loaded judgments and where decisions are bound by the institutional checks and constraints of the law.

---

*Testing Model*, in 20 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 297, 310–11 (L. Berkowitz ed., 1987).

40. Haidt, *supra* note 32, at 817.

41. Jonathan Haidt & Matthew A. Hersh, *Sexual Morality: The Cultures and Emotions of Conservatives and Liberals*, 31 J. APPLIED SOC. PSYCHOL. 191, 212, 214–15 (2001).

42. *Id.* at 212.

43. *Id.* at 214–15.

44. Haidt, *supra* note 32, at 822–23.

45. See Haidt, *supra* note 32.

46. Haidt, *supra* note 32, at 814. By contrast, people engage in motivated cognition only if there is some plausible basis for reaching the desired conclusion. See *supra* notes 36–37.

### C. Motivated Cognition in Legal Judgments

An increasing number of psychologists and legal scholars are examining the operation of motivated cognition in the arena of law. For example, Mark Alicke has experimentally demonstrated the role that motivated cognition plays in the psychology of blame.<sup>47</sup> His culpable causation model indicates that the blameworthiness of an action—i.e., the extent to which behavior defies social norms—shapes judgments about its causal impact on a harmful outcome.<sup>48</sup> In one study, people were more likely to see a speeding driver as the cause of a car accident if the reason for his speeding was to hide a vial of cocaine, as compared to when he was speeding home to hide his parents' surprise anniversary gift, even though all of the other features of the accident were constant in both scenarios.<sup>49</sup> Other studies building upon this finding have shown that, contrary to the dictates of the criminal law, legal judgments about causation, intent, and foreseeability may arise also due to a defendant's character traits that are irrelevant to the substance of a case.<sup>50</sup>

Examining motivated cognition from another angle, Dan Kahan and colleagues have conducted a range of experimental studies to support their cultural cognition hypothesis that people “conform their own beliefs about disputed matters of fact . . . to values that define their cultural identities.”<sup>51</sup> In regard to harm, this theory posits that cognitive mechanisms motivate people to believe that “behavior they find base is also socially harmful.”<sup>52</sup>

In one experiment illustrating the theory of cultural cognition, a diverse sample of 1350 respondents evaluated a real video from the Supreme Court's 2007 *Scott v. Harris* case, which the Court had released so that the public could “see for themselves” that the footage supported the Court's holding (i.e., that a police officer did not violate the Fourth Amendment when he deliberately rammed his car into a fleeing motorist after a high-speed chase).<sup>53</sup> The researchers found that although the majority of the respondents agreed with the

47. See Mark D. Alicke, *Culpable Control and the Psychology of Blame*, 126 PSYCHOL. BULL. 556, 556 (2000) (suggesting that “evidence concerning . . . the event is reviewed in a manner that favors ascribing blame to the . . . persons who evoke the most negative affect”).

48. Mark D. Alicke, *Culpable Causation*, 63 J. PERSONALITY & SOC. PSYCHOL. 368 (1992).

49. Alicke, *supra* note 47, at 567.

50. Janice Nadler & Mary-Hunter McDonnell, *Moral Character, Motive, and the Psychology of Blame*, 97 CORNELL L. REV. 255, 291, 302 (2012) (noting, however, that the influence of an actor's motive and moral character on blame judgments is subject to “boundary conditions,” such as in cases of reckless or negligent wrongdoing, where harm caused by the transgression is severe).

51. See, e.g., Dan M. Kahan et al., “*They Saw a Protest*”: *Cognitive Illiberalism and the Speech-Conduct Distinction*, 64 STAN. L. REV. (forthcoming 2012) [hereinafter Kahan, *They Saw a Protest*]; Dan M. Kahan, *Culture, Cognition, and Consent: Who Perceives What, and Why, in Acquaintance-Rape Cases*, 158 U. PA. L. REV. 729 (2010); Dan M. Kahan et al., *Whose Eyes Are You Going to Believe? Scott v. Harris and the Perils of Cognitive Illiberalism*, 122 HARV. L. REV. 837, 852, 861 (2009) [hereinafter Kahan, *Whose Eyes*]; Dan M. Kahan, *The Cognitively Illiberal State*, 60 STAN. L. REV. 115, 116 (2007) [hereinafter Kahan, *Cognitively Illiberal*].

52. Kahan, *Whose Eyes*, *supra* note 51, at 852.

53. *Id.* at 879.

Court's conclusion in favor of the defendant police officer, "there were sharp differences of opinion along cultural, ideological, and other lines"—with African American, democrat, liberal, egalitarian, communitarian, female, and lower-income respondents reporting more pro-plaintiff perceptions of the video.<sup>54</sup>

Our studies provide a novel contribution to this existing literature by using experimental methodologies to demonstrate how the desire to criminalize can alter perceptions of harm. We suggest that decision makers will engage in motivated cognition to achieve their preferred outcomes within the technical constraints of a given law. This process may not be purposeful, but it can nevertheless undermine fundamental assumptions and principles of the legal system.

If [decision makers] unconsciously . . . favor outcomes congenial to favored ways of life, citizens who adhere to disfavored ones will suffer the same array of disadvantages for failing to conform that they would in a regime expressly dedicated to propagation of a sectarian orthodoxy. This distinctively psychological threat to constitutional ideals . . . has received relatively little attention from commentators or jurists.<sup>55</sup>

The four studies presented below provide original data to address this important concern.

### III.

#### EXPERIMENTAL STUDIES

The following experiments investigate how motivated cognition, in the context of punishment, can lead to plasticity in people's perceptions of harm. We first identified scenarios that induce defiance of the harm principle, such that people want to criminalize the acts even if they do not think the behaviors cause harm to others.<sup>56</sup> We then used the acts that people are willing to criminalize without findings of harm to test our central research query: Is the concept of harm psychologically malleable, such that people who want to penalize such acts will impute harm to the conduct if told that the law requires a finding of harm in order to criminalize?<sup>57</sup> We hypothesized that people

---

54. Kahan, *Whose Eyes*, *supra* note 51, at 838, 864.

55. Kahan, *They Saw a Protest*, *supra* note 51, at 3–4.

56. The first study was exploratory; it did not introduce an experimental manipulation.

57. We recognize, of course, that the law does not always require a showing of harm to criminalize conduct—as seen, for example, in the punishment of criminal attempts that do not result in harm. Previous studies on harm in the criminal context have reported that respondents do assign higher punishment to completed, successful offenses that cause harm than to completed but unsuccessful attempts that do not cause harm. Paul H. Robinson & John M. Darley, *Intuitions of Justice: Implications for Criminal Law and Justice Policy*, 81 S. CAL. L. REV. 1 (2007); John M. Darley et al., *Community Standard for Defining Attempt: Inconsistencies with the Model Penal Code*, 39 AM. BEHAV. SCIENTIST 405 (1996). The present research, however, investigates not how actual harm

would, without full awareness, mold their reported perceptions of harm to achieve their end punishment goals, ostensibly within the terms of the given legal constraint. We also sought to provide evidence that such plasticity of harm exists on both sides of the ideological spectrum, occurs without the decision makers' recognition, and holds significant implications for the realm of law.

### *A. Study 1: Mapping Harm*

#### *1. Methodology*

The first of our four studies sought to identify acts so contrary to widespread social values that people would want to criminalize them even if they regarded the conduct as harmless to others. The participants read short descriptions of fourteen types of conduct that were designed to transgress social values without causing direct, tangible harm to anyone. These included acts like urinating on a public memorial for victims of the 9/11 terrorist attacks; serving cat and dog meat at a restaurant (with full disclosure to the diners); having consensual and protected intercourse with an adult sibling (with no risk of bearing children); going to the supermarket in the nude; using the national flag to wipe mud off the pavement; and using cell-cloning technology to grow and eat a strip of human tissue (with no trouble digesting the meal).

The respondents were asked to answer three questions about each scenario: (1) Does this conduct violate deeply held social values? (2) Does this conduct cause any harm? (3) Should this conduct be criminalized?<sup>58</sup> For conduct that indisputably causes harm to others, such as committing a murder, we would expect people to answer all three questions in the affirmative. However, with our examples, we were seeking to identify scenarios that the respondents would say did violate social values, did not cause harm, but should nevertheless be criminalized.

#### *2. Participants*

After excluding the data of four participants who neglected to answer all of the questions, we analyzed the results from a total of ninety-eight respondents. These participants were 59 percent female, ranged in age from eighteen to forty, and had a mean age of twenty-one.<sup>59</sup>

---

outcomes drive punishment judgments, but rather, how desired punishment outcomes drive harm judgments.

58. The questions were presented in a fixed order for all of the scenarios; however, the scenarios were presented in a counterbalanced order. We intentionally left the definition of harm wide open in this first study due to its exploratory nature.

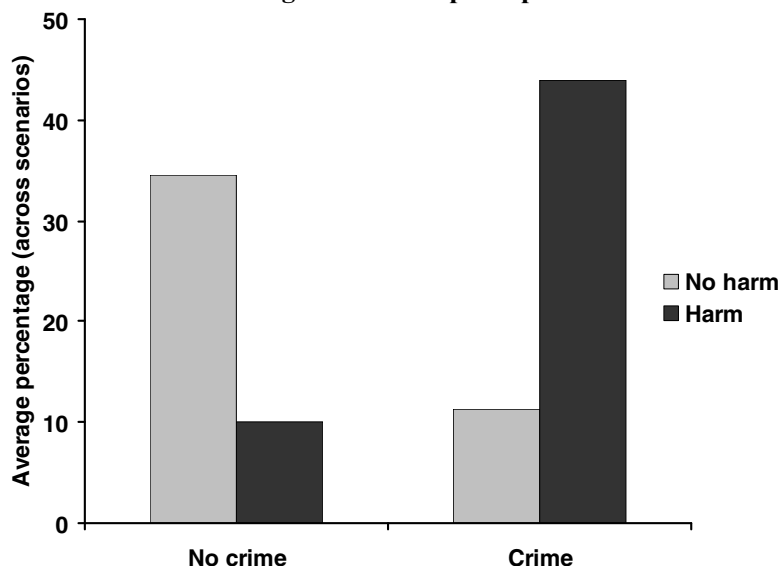
59. The respondents were primarily students at Princeton University who participated in the study in exchange for partial credit toward a course requirement.

### 3. Results

#### a. Compliance with the Harm Principle

Almost all of the respondents reported that the given scenarios violated deeply held social values. Yet, although we tried to present unusual acts that would induce defiance of the harm principle, the participants' decisions about criminalizing most of the acts followed a dominant pattern that remained consistent with the harm principle. Those who said that the conduct caused harm were almost five times more likely to say it should be criminalized; meanwhile, those who said the conduct did not cause harm tended to say that it should not be criminalized.<sup>60</sup> This symmetrical harm/crime pattern, displayed in Figure 1's composite of results, held true for eleven out of the fourteen scenarios.

**FIGURE 1: Study 1—Percentage of harm reports in relation to criminalization responses, in the eleven scenarios conforming to the harm principle**



#### b. Defiance of the Harm Principle

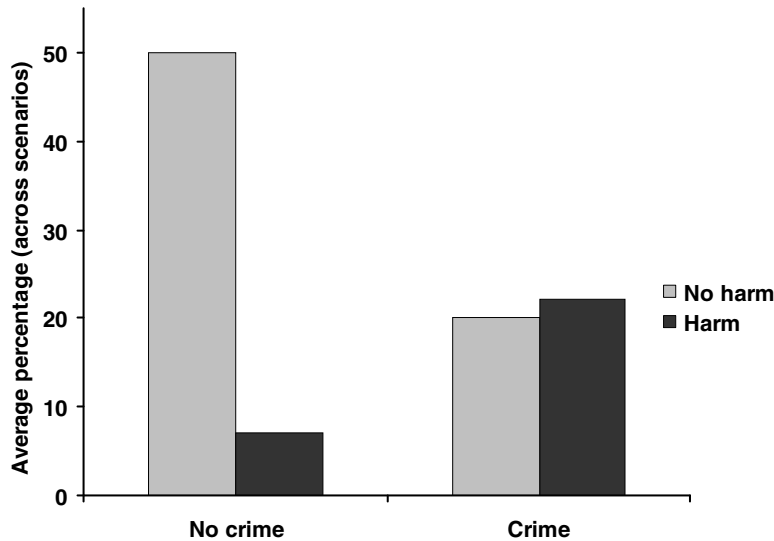
The results in the remaining three scenarios—acts of public nudity, flag defiling, and self-cannibalism<sup>61</sup>—hit closer to the pattern we were seeking to

60. This study sought to identify the relationship between judgments about harm and crime, without making any causal claims. The subsequent studies that switched or counterbalanced the order of the harm/crime questions showed no order effects. *See infra* notes 68, 95.

61. Specifically, the three target scenarios were as follows: “A man goes to the supermarket in the nude”; “A man uses the national flag to wipe mud off the pavement in front of his house”; and “A

identify. In these target cases that defied the harm principle, the participants who said that the conduct should not be criminalized were still far more likely to report that it caused no harm. However, those who said that the behavior in question *should* be criminalized were divided on the question of whether or not it caused any harm, creating the asymmetrical pattern reflected in Figure 2.<sup>62</sup>

**FIGURE 2: Study 1—Percentage of harm reports in relation to criminalization responses, in the three scenarios defying the harm principle**



In sum, this exploratory first study revealed that when a conduct violated deeply held social values, one of two patterns emerged in the respondents' judgments about harm and crime. In most of the scenarios, the participants' responses were in line with the harm principle: reports of harm and decisions to criminalize seemed to go hand in hand. The dominance of this pattern shows that, as reflected in the Model Penal Code and traditional debates on criminalization, the harm principle is quite robust in people's intuitions about when the State should intervene to punish someone's conduct. However, there were a few scenarios that inspired defiance of the harm principle, such that those who criminalized the acts were divided on the question of whether or not

---

man uses cell-cloning technology to grow human cells from his own arm into a strip of human tissue and, out of curiosity, cooks and eats it. He has no trouble digesting the meal."

62. The overall rate of criminalizing was lower for the target scenarios as compared to the scenarios that followed the dominant pattern; people were more likely to recommend criminal penalties when they were intuitively following the harm principle. As we would expect, reports of harm were also lower in the target scenarios—not only because participants were generally less likely to criminalize these acts, but also because those who did criminalize them were more likely to do so without reporting harm (i.e., in defiance of the harm principle).



the acts actually caused harm. These were of primary interest for Study 2, in which we investigated whether people would recruit harm when necessary to justify their desire to punish such conduct.

### *B. Study 2: The Plasticity of Harm*

#### *1. Methodology*

The second study was a randomized controlled experiment designed to test the plasticity of harm hypothesis. Mainly, we sought to demonstrate that when people are presented with a legal constraint that requires harm in order to punish an act, and are then asked to judge an act that they want to penalize regardless of whether it causes harm to others (e.g., flag defiling or public nudity from Study 1), they will impute harm to the conduct in order to legitimately criminalize it within the terms of the given law.

The participants in Study 2 were randomly assigned to either a Constraint condition, where they received a necessity-of-harm manipulation described below, or a Control condition, where no manipulation was introduced. Specifically, those assigned to the Constraint condition were presented with the following information at the start of the study, and once again as a reminder halfway through the survey: “U.S. courts have decided that the government can impose a criminal penalty only upon conduct that is shown to cause harm.”<sup>63</sup> This manipulation essentially imposed the harm principle as a legal constraint upon the participants’ decisions to criminalize. Although the statement is not accurate, we introduced a check during the debriefing to ensure that the participants had found it credible, and only one respondent was excluded for not believing it. On the other hand, those participants assigned to the Control condition were under no legal restraint in their decision making; they did not need to find harm in order to criminalize a disfavored act.

All of the respondents in both the Constraint and Control conditions were given the following definition of harm, in order to capture thoughtful and legally relevant measures of the variable: “Harm, for these purposes, is defined as injury to a person or persons that can be clearly demonstrated. There could be types of conduct that are wrong, but do not cause harm.” By imposing these parameters, we anticipated that the overall number of reports of harm would go down as compared to the first study, in which the definition of harm had been left entirely open-ended. The greater specification of harm in this experiment created a more conservative test of our plasticity hypothesis: if, despite this

---

63. When the respondents were asked to rate on a seven-point scale the extent to which they agreed that the government should only impose a criminal penalty for conduct that is shown to cause harm (with one being “strongly disagree,” four being “neutral,” and seven being “strongly agree”), the mean was 4.14—indicating that, on average, participants reported being neutral in regard to the harm principle (although the results of Study 1 suggest that people tended to intuitively follow the harm principle in their criminalization judgments).

heightened definition, those participants faced with the legal constraint imputed harm to conduct that they wanted to criminalize, such a finding would provide strong evidence that people do in fact recruit harm to achieve their punishment goals.

After this initial information, all of the participants were presented with short descriptions of different types of conduct. Of primary interest were two of the scenarios that had shown the target pattern in Study 1 (i.e., people criminalizing without harm): the public nudity case (“A man goes to the supermarket in the nude”) and the flag-defiling case (“A man intentionally uses the national flag to wipe mud off the pavement in front of his house on a Sunday afternoon”).<sup>64</sup> Among the three target scenarios from Study 1, we chose to focus on these two acts because they similarly defied the harm principle even though they fell at different ends of the criminalization spectrum: regardless of their knowledge of the law on these issues, the participants in Study 1 were most likely to criminalize the public nudity conduct (which would, in fact, be punishable in many jurisdictions),<sup>65</sup> and least likely to criminalize the flag-defiling conduct (which would be protected under the First Amendment).<sup>66</sup> Although these scenarios are fairly trivial by way of crimes, the relative triviality was necessary for purposes of the experimental design; we needed to use acts that many would regard as harmless to others, albeit unsavory, in order to show that people will recruit harm, if needed, to punish such behavior.

To keep their judgments about harm and crime in perspective, the participants in Study 2 were also asked to rate various calibration scenarios—acts that had shown the dominant harm/crime pattern in the first study, and more typical crimes that are clearly harmful and punishable (e.g., physical assault). All the scenarios were presented in counterbalanced order, using a Latin square design.<sup>67</sup> After each scenario, the participants were asked to indicate, *inter alia*, whether or not the government should punish the conduct

---

64. The flag-defiling scenario specified that the conduct took place on a Sunday afternoon (when people are likely to be out and about in a neighborhood), in order to suggest that this conduct was likely to be seen by others—which was relevant to the given definition of harm. Nevertheless, the nudity scenario was still more public than the flag-defiling scenario, which enabled us to test for the plasticity of harm in both more and less communal spheres.

65. See *Barnes v. Glen Theater, Inc.*, 501 U.S. 560, 560 (1991) (upholding a state law regulating public nudity due to “governmental interest in protecting societal order and morality”).

66. *United States v. Eichman*, 496 U.S. 310 (1990) (striking down the Flag Protection Act that made it a crime to destroy an American flag); *Texas v. Johnson*, 491 U.S. 397 (1989) (holding that burning the U.S. flag is protected expression under the First Amendment).

67. A Latin square is “a basic tool that an experimenter uses to set up a design in which an incidental factor is systematically counterbalanced over the treatment conditions.” GEOFFREY KEPPEL & THOMAS D. WICKENS, *DESIGN AND ANALYSIS: A RESEARCHER’S HANDBOOK* 381 (4th ed. 2007). In this study, it was used to control for any variation due to the order in which the participants judged the scenarios. The scenarios were randomly assigned to rows and columns in a Latin Square table, such that each scenario appeared once per row and once per column, and this determined the order in which the scenarios were presented to each participant (i.e., with each scenario in each possible position).

through a criminal penalty; whether or not the conduct caused clearly demonstrable harm (a dichotomous yes/no measure); and how much harm, if any, the conduct caused (a continuous measure on an eight-point scale ranging from “no harm” to “a lot” of harm).<sup>68</sup>

## 2. *Participants*

Eleven participants who failed an instructional manipulation check<sup>69</sup> and one participant who indicated that he did not believe the manipulation were excluded from the analysis, leaving data from a total of ninety respondents for the analysis (forty-five respondents per condition). These participants were 52 percent female, ranged in age from eighteen to twenty-three, and had a mean age of twenty.<sup>70</sup>

## 3. *Predictions*

The participants in the Constraint condition, who were presented with the target scenarios (acts that people criminalize even without finding harm) and the necessity-of-harm constraint (a legal instruction to criminalize conduct only if it causes harm), could be expected to respond in one of three ways: (1) they could abide by the legal constraint and not punish the offensive acts when they perceived no harm, thereby causing the rate of criminalization to go down as compared to in the Control condition; (2) they could ignore the legal constraint and continue to punish the offensive acts at the same rate even when they reported no harm, which would lead to no difference in the ratings of criminalization and harm between the two conditions; or (3) they could recruit harm in order to meet their criminalization goals, ostensibly within the terms of the legal constraint.

Our hypothesis predicted the third outcome. We knew from the dominant pattern in Study 1 that people tend to intuitively subscribe to the harm

---

68. The questions were posed in a fixed order after presenting each of the scenarios. Although the participants in Study 1 were asked first about harm and then about criminalization, the respondents in this second study were asked first about criminalization and then about harm. We did not expect any differences to arise based on the order in which these questions were asked, and the data from the Control condition of Study 2 (which mirrored the methodology of Study 1, but flipped the order of the harm/crime measures) as well as Study 3 (which counterbalanced the order in which the harm/crime questions were presented) confirmed this expectation. *See infra* note 95.

69. To test whether the participants carefully and comprehensively read the instructions and information presented at the beginning of the questionnaire, which included the experimental manipulation in the Constraint condition, we inserted a manipulation check asking them to “write today’s date” after reading the material. Those respondents who failed to write the date were assumed not to have paid adequate attention, so their responses were excluded from the analysis. *See generally* Daniel M. Oppenheimer et al., *Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power*, 45 J. EXPERIMENTAL SOC. PSYCHOL. 867 (2009).

70. The respondents were Princeton University students who participated in the experiment in exchange for partial credit toward a course requirement. A follow-up experiment conducted with a larger and more representative sample of adults through the Mechanical Turk website ( $N = 208$ ) provided a theoretical replication of the primary results presented here (data on file).

principle, and prior research had shown that lay decision makers are generally motivated to comply with the law,<sup>71</sup> so we did not expect that the respondents would entirely ignore the harm principle when it was presented in the form of a legal constraint. However, we also knew from prior psychological research on motivated cognition and the social intuitionist model that people intuitively reach for ways to justify the outcomes that feel right to them. We therefore predicted that those participants who wanted to criminalize the target acts would impute harm to the conduct in order to support their desired punishment outcomes. That is, we expected the number of harm reports to be higher in the Constraint condition as compared to the Control condition, while punishment recommendations stayed constant—thereby demonstrating a plasticity of harm in the service of people’s criminalization goals.

#### 4. Results

##### a. The Rigidity of Criminalization

As predicted, the necessity-of-harm manipulation did not reduce the extent to which the participants penalized the offensive conduct in the target scenarios. When we averaged the criminalization scores for the nudity and flag-defiling scenarios, there was no statistical difference between the Constraint and Control conditions (see Figure 3 on the following page).<sup>72</sup> Chi-square analyses of the relationship between condition and criminalization scores were also nonsignificant in both scenarios, confirming that any slight variation in the criminalization count between the two conditions was attributable to chance, not the experimental manipulation.<sup>73</sup> Furthermore, a binary logistic regression analysis confirmed that the condition to which the participants were randomly assigned (i.e., whether or not they received the necessity-of-harm legal constraint) did not significantly predict whether or not they criminalized the conduct in either scenario.<sup>74</sup>

This finding demonstrates the perseverance of people’s criminalization judgments. Even though we knew from Study 1 that approximately half of the people who wanted to punish acts like public nudity and flag defiling did so despite reporting them as nonharmful, telling the participants that the law requires a finding of harm in order to punish did not reduce the rate of criminalization in these scenarios.

---

71. TOM R. TYLER, WHY PEOPLE OBEY THE LAW 46 (1990) (“Just as respondents almost universally feel that breaking the law is immoral, they feel a strong obligation to obey the law.”).

72. Nonsignificant t-test result:  $t(88) = .80, p = .43$ .

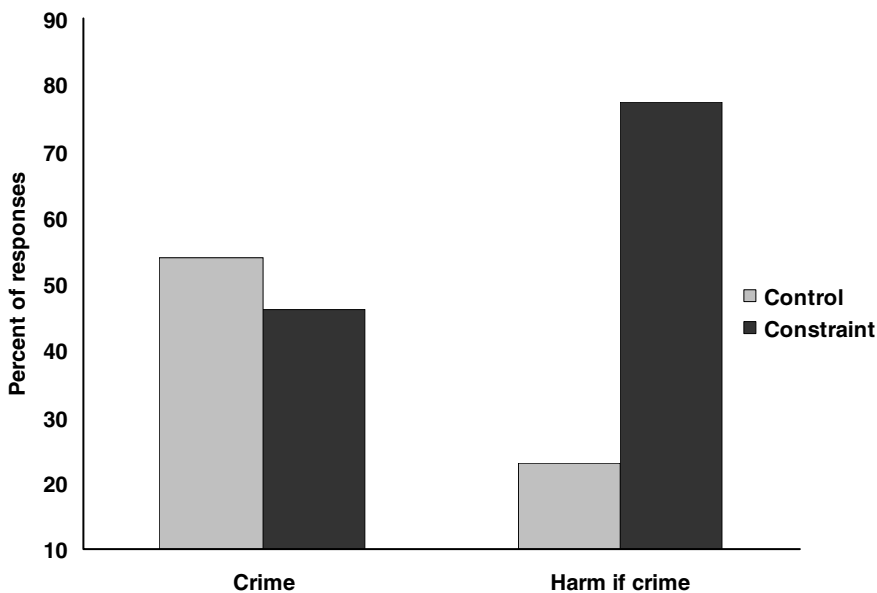
73. Nonsignificant chi-square results:  $\chi^2(1, N = 90) = .50, p = .64$  (nudity scenario);  $\chi^2(1, N = 90) = .28, p = .79$  (flag-defiling scenario).

74. Nonsignificant binary logistic regression results:  $B = -.33, S.E. = .47, p = .48$  (nudity scenario);  $B = -.28, S.E. = .53, p = .60$  (flag-defiling scenario).

*b. The Plasticity of Harm*

Yet, the participants who had been presented with the necessity-of-harm manipulation did not simply ignore the legal constraint either. Instead, they found a way to penalize these acts while seemingly complying with the law: by recruiting harm on an as-needed basis. For both scenarios, the participants who wanted to punish the conduct were significantly more likely to report harm in the Constraint condition (where the law required a finding of harm to punish), as compared to in the Control condition (where there was no legal constraint).<sup>75</sup> Figure 3 shows that the percentage of participants (averaged across the nudity and flag-defiling scenarios) who wanted to criminalize the behavior remained statistically constant across the two conditions, while the percentage of criminalizing participants who reported harm rose dramatically in the Constraint condition. This discrepancy illustrates the rigidity of people's criminalization goals, as compared to the plasticity in their reports of harm in order to achieve those goals within the terms of the legal constraint.

**FIGURE 3: Study 2—Percentage of criminalization recommendations (among all of the participants) and harm reports (among the participants who criminalized), averaged across the nudity and flag-defiling scenarios, by condition**



75. Chi-square results:  $\chi^2(1, N = 65) = 11.20, p = .001$  (nudity scenario);  $\chi^2(1, N = 18) = 7.90, p = .005$  (flag-defiling scenario). These statistics indicate that the difference in reports of harm between the participants in the Constraint and Control conditions in both scenarios is too great to be attributed to chance. In the calibration scenarios, there was no significant difference in harm reports between the conditions because, as expected, reports of harm were high even among those who criminalized the conduct in the Control condition.

More specifically, Table 1 and Figure 4 reveal that among the respondents who wanted to punish the act of public nudity (which the majority of the participants did), reports of harm more than doubled in the Constraint condition as compared to the Control condition. Although far fewer participants wanted to punish the act of flag defiling, among those who did only one reported harm in the Control condition, as compared to six in the Constraint condition.

**TABLE 1: Study 2—Frequency of harm reports, between the two conditions, among the participants who criminalized (total number of participants indicated in parentheses)**

	<i>Control condition</i>	<i>Constraint condition</i>
<i>Nudity scenario</i>	10 (34)	22 (31)
<i>Flag-defiling scenario</i>	1 (10)	6 (8)

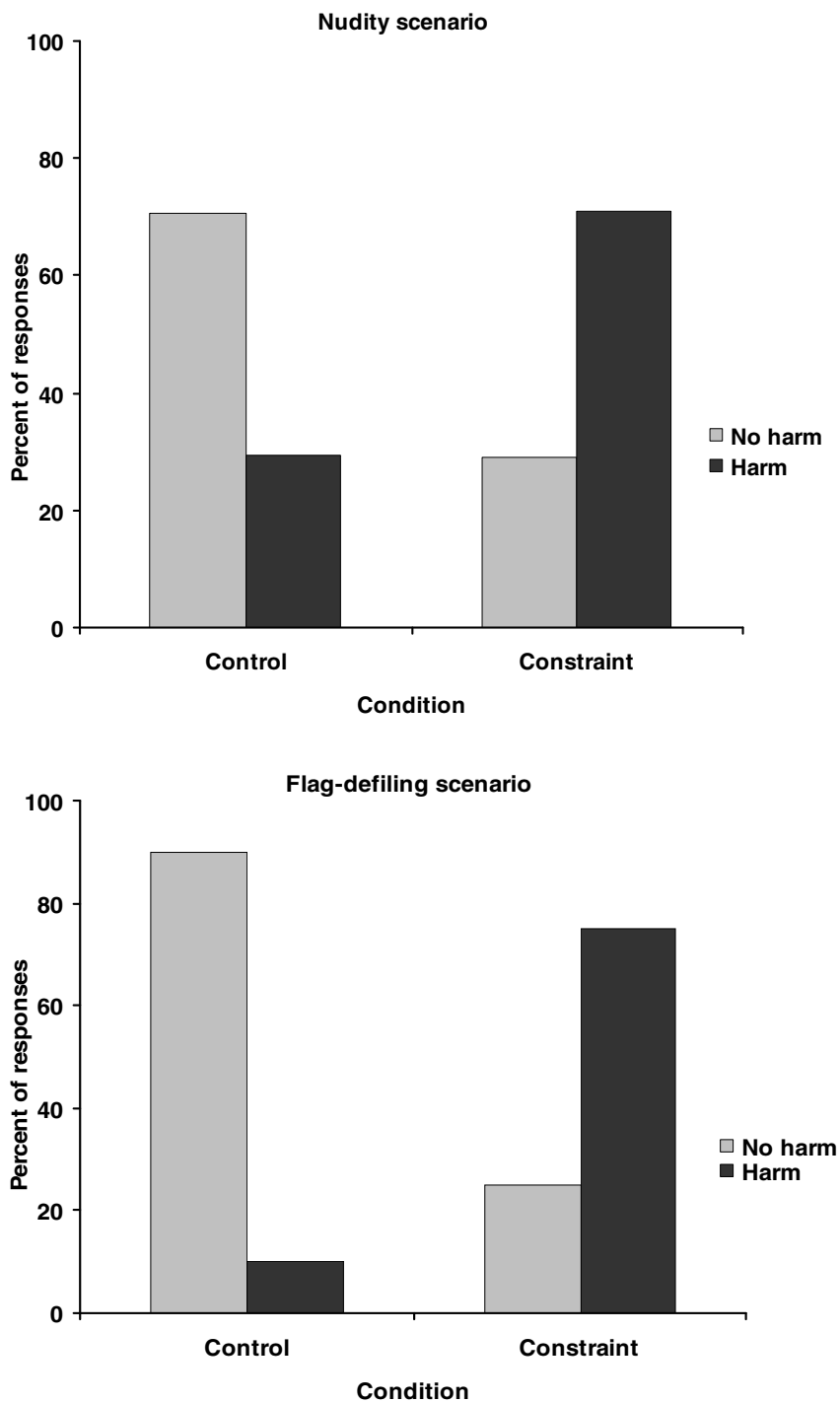
The significant growth in reports of harm between the Constraint and Control conditions supports the predicted plasticity of harm effect. In both scenarios, those who wanted to punish were significantly more likely to impute harm to the conduct when a finding of harm was required to penalize, as compared to when no constraint was imposed upon the criminalization decision. Due to the greater specification of harm in this experiment, the frequency of reported harm was lower in the Control condition than it had been for the same scenarios in Study 1, making the plasticity effect—i.e., the significantly higher frequency of harm reported in the Constraint condition—all the more striking.

A binary logistic regression analysis confirmed that criminalization was a significant predictor of harm only in the Constraint condition, and not in the Control condition.<sup>76</sup> That is, whether or not the participants punished an act was a strong predictor of whether or not they would report that act as harmful only if they had been told that the law requires a finding of harm in order to punish. Moreover, among the participants who recommended criminalization, the condition to which they had been assigned (Constraint or Control) was a significant predictor of whether or not harm was reported—with those in the Constraint condition being significantly more likely to report harm.<sup>77</sup>

76. Binary logistic regression results:  $B = 2.19$ ,  $S.E. = .76$ ,  $p = .004$  (nudity scenario);  $B = 3.21$ ,  $S.E. = .97$ ,  $p = .001$  (flag-defiling scenario).

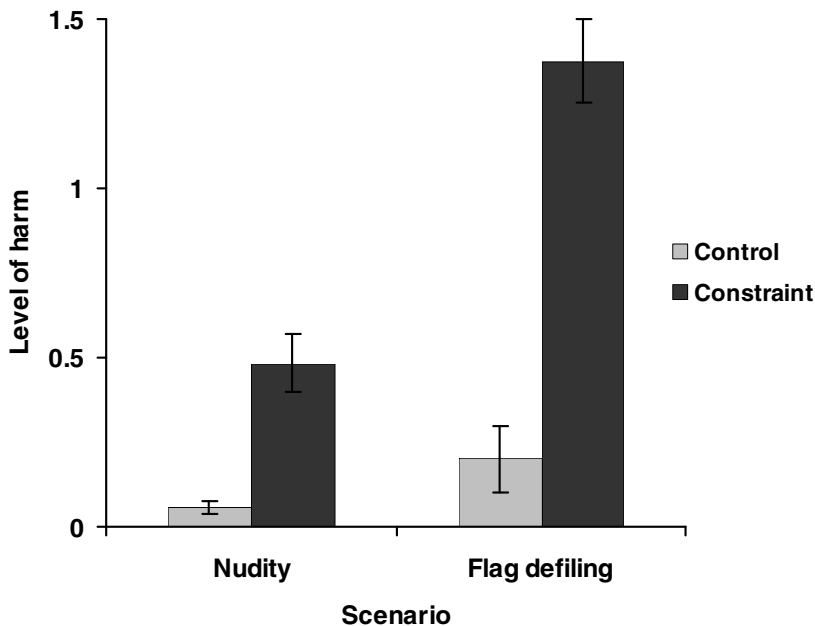
77. Binary logistic regression results:  $B = 3.30$ ,  $S.E. = 1.33$ ,  $p = .01$  (nudity scenario);  $B = 1.77$ ,  $S.E. = .55$ ,  $p = .001$  (flag-defiling scenario). Furthermore, a Pearson correlation matrix indicated that among the participants who recommended criminal penalties, the condition to which they were assigned was significantly correlated at the .01 level with reports of harm in both scenarios ( $r = .42$  for nudity scenario,  $r = .66$  for flag-defiling scenario). As expected, the condition was neither a significant predictor of harm nor significantly correlated to harm in the calibration scenarios.

**FIGURE 4: Study 2—Percentage of harm reports (among the participants who criminalized), by condition**



Finally, we analyzed the continuous measure of harm to ensure replication of the plasticity effect. In addition to being asked a yes/no question about whether the act in question caused harm, the participants had been asked to indicate *how much* harm, if any at all, the conduct caused. An analysis of variance in the responses of those who punished the conduct revealed that these participants reported significantly higher degrees of harm in the Constraint condition as compared to the Control condition, in both the nudity and flag-defiling scenarios.<sup>78</sup> Thus, those who were told that the law requires a finding of harm in order to criminalize were not only more likely to report the presence of harm if they chose to penalize, but also reported significantly greater levels of harm than those who were not informed of the legal constraint, as depicted in Figure 5.

**FIGURE 5: Study 2—Level of harm reported (among the participants who criminalized), by condition**



Although the statistical effect size of the difference between the levels of harm reported in the two conditions was large,<sup>79</sup> the average levels of harm reported in the scenarios were relatively low, even in the Constraint condition (i.e., below 1.5 on an 8-point scale). This is not surprising given that the

78. Analysis of variance results:  $F(1, 63) = 3.86, p = .05, \eta^2 = .06$  (nudity scenario);  $F(1, 16) = 6.34, p = .02, \eta^2 = .28$  (flag-defiling scenario).

79. The  $\eta^2$  statistic (eta squared) provides an estimate of effect size: an  $\eta^2$  of .01 is considered a small effect, an  $\eta^2$  of .09 is a medium effect, and an  $\eta^2$  of .25 or more is a large effect. See BARBARA G. TABACHNICK & LINDA S. FIDELL, *USING MULTIVARIATE STATISTICS* 54–55 (5th ed. 2007).



Constraint condition required harm to punish but did not suggest that any particular amount of harm was needed. Thus, the participants simply needed to cross a threshold of some de minimis harm in order to justify their criminalization decisions within the terms of the legal constraint. Consistent with the theory of motivated cognition—which posits that people only stretch their reasoning to the extent necessary to achieve their desired outcomes—the respondents did not recruit any more harm than necessary.<sup>80</sup>

In sum, the results of Study 2 indicate that the experimental manipulation had its predicted effect. When presented with acts that people criminalize without reporting harm, but informed that the law requires a finding of harm in order to punish, the participants who wanted to penalize the acts continued to do so but were significantly more likely to report the presence of harm and reported significantly greater levels of harm to support that choice. The respondents did not alter their criminalization goals in adherence to the necessity-of-harm constraint, but rather, recruited the harm that they needed to justify their desired punishment outcomes within the terms of the legal constraint. The data support our hypothesis regarding the plasticity of harm: when people want to criminalize conduct and a finding of harm is required to do so, harm is imputed to the conduct.

It is worth noting that the participants reported that the behaviors in the target scenarios violated widely held social values to a significantly larger extent than their own personal values.<sup>81</sup> Considered in light of the harm plasticity finding, people seemed motivated to recruit harm in order to criminalize conduct even when they reported being less offended by it personally as compared to the rest of society. However, in the flag-defiling scenario, the respondents' punishment recommendations were significantly correlated only with the extent to which the conduct violated their own personal values, and not societal values at large.<sup>82</sup> Thus, in the more ideologically charged scenario, punishment was fuelled primarily by people's personal disagreement with the conduct in question—a motive that is explored further in the next study.

### *C. Study 3: Motivating Role of Ideological Incongruency*

An alternative to motivated cognition that explains Study 2's plasticity of harm finding could be that, rather than imputing harm to the conduct they wanted to criminalize, those presented with the legal constraint were spurred to

---

80. Correspondingly, the participants assigned relatively minor penalties in these cases. The most common type of punishment recommended was a monetary fine, followed by a fine coupled with community service and/or public apology. For both target scenarios, the mean penalty severity was in the two-point range of a seven-point scale.

81. Paired sample t-test results:  $t(64) = -8.31, p < .001$  (nudity scenario);  $t(64) = -6.06, p < .001$  (flag-defiling scenario).

82. Correlation results:  $r(87) = .30, p = .004$ .

conduct a “more intense but essentially objective search” for harm.<sup>83</sup> Study 3 sought to rule out this possibility by explicitly demonstrating that subjective, directional goals drive the recruiting of harm in this context.

To this end, we introduced a second punishment motive: in addition to the instinct to criminalize the disliked conduct of public nudity, which we discovered in the prior studies, we also triggered a (legally impermissible) desire to punish for espousing an ideologically repugnant viewpoint. If the results in Study 2 can be explained by saying that there always was harm present in the nudity scenario but it took a harm constraint for the respondents to look hard enough to find it, then the introduction of a second legally irrelevant factor that heightens the motivation to punish should not change the rate at which people find harm stemming from the nudity itself.

We chose the issue of abortion for this purpose, because people on both sides of the debate—pro-life or pro-choice—tend to have strong views on the matter. The design of Study 3 thus also provided an opportunity to explore the universality of the harm plasticity effect by testing whether it would be manifested among people on both sides of the abortion debate.

### *1. Methodology*

The first step in Study 3 was to gauge the participants’ views on abortion. This was done through a “social values survey,” in which the respondents were asked to rate their views on various controversial social issues, such as gay marriage and the death penalty, among others. Embedded in these questions was a measure of the respondents’ positions on abortion, on a seven-point scale ranging from “strongly pro-life” to “strongly pro-choice.” The participants who responded using the extremes of the scale were then categorized as either pro-life (1-2) or pro-choice (6-7), and those with moderate views (3-5) were excluded from the analysis.

The respondents were next asked to fill out the experimental questionnaire, which began with the legal constraint used in Study 2—informing them that according to the law, the government can punish a conduct only if it causes clearly demonstrable harm to others. However, the constraint was not used as an experimental manipulation in this study because all the participants received the same legal instruction. They were then presented with one scenario about which they answered criminalization and harm questions (on seven-point scales).

The scenario retained the public nudity vignette from Study 2, because we already knew from the results of that study that people recruit harm in order to punish that conduct. However, in this experiment, we added a twist whereby the nudist in the supermarket was either a pro-choice or a pro-life advocate,

---

83. Kunda, *supra* note 33, at 489 (discussing an alternative explanation for the motivated cognition effect).

who held up a sign and handed out flyers stating why abortion should or should not be legalized. Thus, Study 3 in effect had a 2x2 between-subjects design, depicted in Table 2, in which the nudist's ideological position on abortion was, by random assignment, either congruent or incongruent with the respondents' positions on abortion.

**TABLE 2: Study 3—Between-subjects design**

	<i>Nudist with pro-life message</i>	<i>Nudist with pro-choice message</i>
<i>Pro-life respondent</i>	Congruent	Incongruent
<i>Pro-choice respondent</i>	Incongruent	Congruent

The ability of the participants to make an association between the two questionnaires (the first one gauging their positions on abortion and the second one presenting an abortion-related scenario) was diminished by the fact that this study was conducted in a public mall on a day when several other researchers were doing survey experiments, so the participants were filling out numerous questionnaires. Moreover, the abortion measure was imbedded among other “hot-button” questions in the initial survey, and the two surveys were formatted to look very different from each other.

## 2. Participants

People who expressed moderate views on abortion and who failed the experimental manipulation checks were excluded from the analysis, leaving a total of fifty-three respondents in this experiment. These participants were 70 percent female, ranged in age from eighteen to sixty-five, and had a mean age of thirty.<sup>84</sup>

## 3. Predictions

We predicted that the participants judging a nudist whose position on abortion was ideologically incongruent with their own would be more motivated to punish him, and would therefore recommend more severe penalties and be more likely to impute harm to the conduct in order to justify their punishment, as compared to when the nudist's message on abortion was congruent with their own. We expected to see this result even though the ideological position of the nudist is not only legally irrelevant, but also constitutionally impermissible to consider in the determination of punishment in this context, under the First Amendment.<sup>85</sup> However, consistent with the

84. The respondents were shoppers at a shopping mall in New Jersey, who participated in the study in exchange for a small monetary payment.

85. See U.S. CONST. amend. I; *infra* Part IV.C (discussing constitutional implications of present findings).

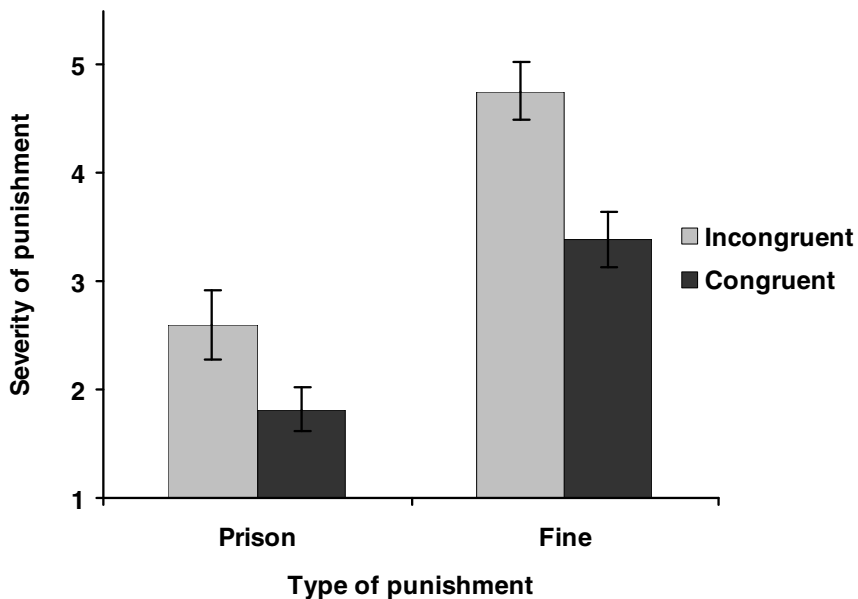
illusion of objectivity, we hypothesized that the respondents themselves would fail to recognize the legally inappropriate motive driving their punishment decisions. Furthermore, we expected to see the harm plasticity effect across ideological groups, among both pro-life and pro-choice participants.

#### 4. Results

##### a. Enhanced Punishment

As predicted, the participants recommended punishing the nudist differently, depending on whether or not they agreed with his message.<sup>86</sup> Specifically, as illustrated in Figure 6, the respondents assigned significantly higher fines and prison sentences to the nudist if his message on abortion was incongruent with their own positions on the issue.<sup>87</sup>

**FIGURE 6: Study 3—Prison and fine recommendations, by congruency with the nudist’s ideological position**



##### b. Imputed Harm

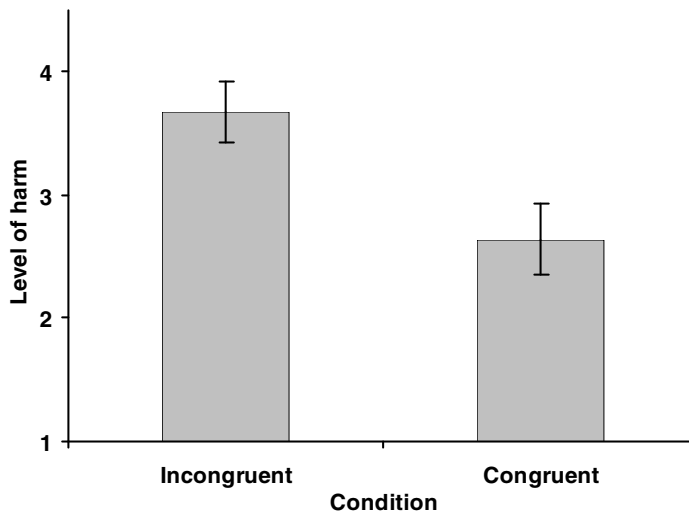
Consistent with our hypothesis, the respondents who fell into the Incongruent condition, and were thereby more motivated to punish the nudist, correspondingly reported significantly more harm in the scenario as compared

86. Analysis of variance result:  $F(1, 51) = 4.01, p = .05, \eta^2 = .07$ .

87. Analysis of variance results: fine— $F(1, 51) = 13.22, p = .001, \eta^2 = .21$ ; prison term— $F(1, 51) = 4.34, p = .04, \eta^2 = .08$ .

to those who fell into the Congruent condition (Figure 7).<sup>88</sup> The act of public nudity and the necessity-of-harm constraint were held constant across conditions and thus could not account for this significant difference in harm ratings; the only feature that differed was whether or not the nudist's message on abortion was consistent with the respondents' own views. So, the participants' reported perceptions of harm were directionally motivated not only by their desire to criminalize nudity within the terms of the given legal constraint (as evidenced in Study 2), but also by their personal ideological beliefs about a topic that was legally irrelevant to the alleged crime. This shows that the respondents did not simply uncover existing harm in the nudity scenario, but rather, imputed harm based on their punishment goals.

**FIGURE 7: Study 3—Level of reported harm, by congruency with the nudist's ideological position**



Yet, when the participants were asked to describe the harm they perceived in the scenario, their explanations focused exclusively on the public nudity, and not on the nudist's position on abortion.<sup>89</sup> This finding suggests that, consistent with the illusion of objectivity in motivated cognition, the respondents seemed to be unaware of the ideologically biased motives that drove their judgments.

### *c. Individual Difference Findings*

The results of Study 3 presented thus far were grouped and analyzed according to the congruency between the views of the participants and the

88. Analysis of variance results:  $F(1, 51) = 7.26, p = .01, \eta^2 = .13$ .

89. Only one (pro-choice) participant in the study explicitly referenced the nudist's (pro-life) position on abortion in her explanation of harm.

message of the actor they were judging, rather than by each respondent's particular position on abortion (i.e., pro-life or pro-choice). However, we also examined the effect of the latter, in order to test for the universality of the plasticity effect across ideological views.

Psychologists, including Haidt, have identified five moral foundations upon which people construct their moral systems: harm/care, fairness/reciprocity, in-group/loyalty, authority/respect, and purity/sanctity.<sup>90</sup> They posit that the morality of political liberals is driven more by the first two foundations, whereas political conservatives rely on all five foundations.<sup>91</sup> Either way, harm is recognized as a motivating factor for moral convictions on both sides of the political spectrum. Given that our study focused on perceptions of harm, we predicted that respondents who were faced with a message that clashed with their own ideological views would be vulnerable to the motivated cognition effect regardless of whether they were pro-life or pro-choice on the issue of abortion.

Indeed, the analysis of this individual difference measure revealed a significant two-way interaction between the ideological position of the participants and the position of the actor on judgments of harm. Pro-life respondents reported significantly more harm when judging a nudist with a pro-choice message as compared to a nudist with a pro-life message, and vice versa for pro-choice respondents.<sup>92</sup> We did additionally find a main effect of abortion position, whereby pro-life participants (who were more politically conservative than pro-choice participants<sup>93</sup>) generally assigned more severe punishment and reported greater perceptions of harm in the nude activism scenario, regardless of the nudist's position on abortion. These results are illustrated in Figure 8 on the following page.

In sum, Study 3 demonstrated that a legally irrelevant factor of ideological incongruency increased the motivation to punish. This augmented imputations of harm, seemingly without the decision makers' recognition or acknowledgment since they maintained that the harm was caused only by the public nudity. Furthermore, our results revealed universality in people's tendency to impute harm in this context; motivated cognition operated in the legal judgments of both pro-life and pro-choice respondents, depending on

---

90. Jesse Graham et al., *Liberals and Conservatives Rely on Different Sets of Moral Foundations*, 96 J. PERSONALITY & SOC. PSYCHOL. 1029 (2009).

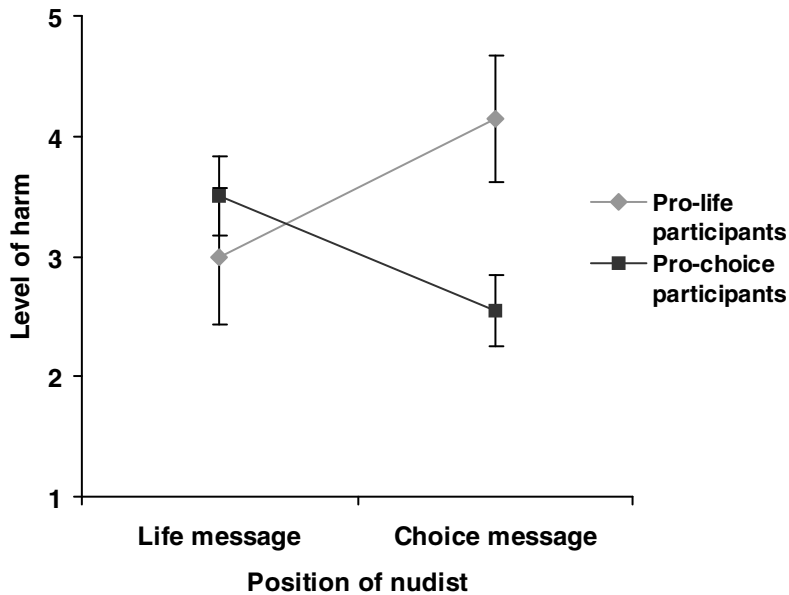
91. *Id.* at 1040 ("Across all four studies, liberal morality was primarily concerned with harm and fairness, whereas conservative moral concerns were distributed more evenly across all five foundations.").

92. Analysis of variance results:  $F(1, 49) = 5.42, p = .02, \eta^2 = .10$ .

93. There was a significant correlation between the participants' positions on abortion and their self-reported political views: those who were less in favor of legalized abortion (i.e., "pro-life") were more conservative in their political views; those who were more in favor of legalized abortion (i.e., "pro-choice") were less conservative in their political views. Correlation results:  $r = -.40, p = .003$ .

whether or not the conduct in question conflicted with their particular position on the issue of abortion.

**FIGURE 8: Study 3—Level of harm reported, by ideological positions of the participants and the nudist**



#### *D. Study 4: Confronting the Illusion of Objectivity*

Study 3 presented some indirect evidence that those participants who recruited harm to justify their ideologically driven punishment goals were operating under an illusion of objectivity, since their explanations of the alleged harm focused exclusively on the nudity aspect of the conduct. The final study sought to investigate this important assumption of the theory of motivated cognition. If, as we suggest, the motivated recruiting of harm is a nondeliberate process, then getting people to confront the illusion of objectivity should reduce the harm plasticity effect—thereby also pointing toward a potential remedy for outcome-driven legal judgments.

##### *1. Methodology*

The methodology of Study 4 was the same as that of Study 3, with a critical variation: in order to more directly explore the illusion of objectivity, Study 4 used a within-subjects design—presenting all the participants with the legal constraint requiring harm in order to punish, followed by *both* the pro-life

and the pro-choice nudist scenarios of Study 3.<sup>94</sup> The respondents were then asked to rate the severity with which each actor should be punished and the extent of harm, if any, caused by each actor's conduct.<sup>95</sup>

## 2. Participants

Excluding the data of the respondents who failed manipulation checks or reported moderate abortion views left a total of thirty-seven participants in this experiment. These participants were 68 percent female, ranged in age from eighteen to twenty-two, and had a mean age of nineteen.<sup>96</sup>

## 3. Predictions

We expected the within-subjects design to peel away the illusion of objectivity that amplified the recruiting of harm in Study 3. Since it would be obvious to the participants that the public nudity component was the same in both scenarios, irrespective of the nudist's position on abortion, they could no longer cite harm caused by the nudity as the justification for more severely punishing the nudist whose message was incongruent with their own views on abortion (as they did in Study 3). So, if the plasticity of harm resulted from an unintentionally biased cognitive process, the transparency in Study 4's presentation of scenarios could be expected to remove ideologically motivated discrepancies in punishment decisions, thereby curtailing the need to recruit harm. We predicted that when confronted with both the pro-choice and pro-life nudists, the respondents' punishment judgments and their corresponding perceptions of harm in the two scenarios would not differ based on their own ideological positions on abortion.

## 4. Results

As anticipated, the participants in this experiment did not punish the nude protester whose message was incongruent with their own views on abortion any differently than the nude protester whose message was congruent with their personal beliefs. Although pro-life respondents did generally punish more severely and assign more jail time to the nude protester than pro-choice

---

94. The pro-life and pro-choice scenarios were presented in counterbalanced order across the participants and the study was conducted using a paper questionnaire (as opposed to electronic), so that the respondents could see both scenarios before making their decisions. Moreover, the participants were instructed that they would be asked to make judgments about two scenarios, and that they should consider both of them carefully. These measures reduced the risk that the within-subject design's confrontation of the illusion of objectivity would impact only the second scenario that the respondents judged.

95. The order of the punishment and harm measures was also counterbalanced across the participants in this study in order to explicitly test for order effects, which were found not to be present.

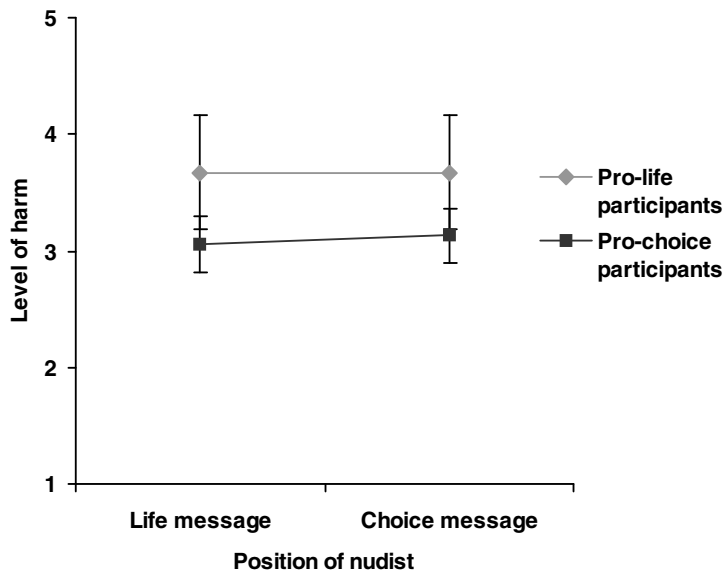
96. The respondents were students at Princeton University who participated in the study in exchange for partial credit toward a course requirement. Within-subject designs have more experimental power, and therefore allow for hypothesis testing with a smaller number of participants.



respondents (as seen in Study 3), they did so in both scenarios, regardless of whether the protester was advocating a pro-life or a pro-choice message.<sup>97</sup> Likewise, there was no significant difference in pro-choice respondents' punishment ratings of the pro-life and pro-choice nudists.

The absence of ideologically-driven differences in punishment obviated the need to impute harm based on this factor. As Figure 9 demonstrates, the congruency between the nudist's message and the participants' own positions on abortion had no effect on the levels of harm reported in each scenario (in contrast to the significant interaction seen in Study 3's Figure 8).<sup>98</sup>

**FIGURE 9: Study 4—Level of harm reported, by ideological positions of the participants and the nudist**



In fact, the mean levels of harm reported in both the scenarios in Study 4 ( $M = 3.16$  for the pro-life nudist and  $M = 3.22$  for the pro-choice nudist on 7-point scales), which were *not* significantly different from each other, fell right in between the mean levels of harm reported in Study 3's congruent ( $M = 2.64$ ) and incongruent ( $M = 3.68$ ) conditions, which *were* significantly different from each other (Figure 7). The presentation of two scenarios that differed only in the ideological content of the actor's message thus forced the respondents to confront potential biases stemming from this legally irrelevant factor, thereby

97. The pro-life participants did appear to assign slightly higher fines to pro-choice protesters as compared to pro-life protesters, but this interaction between fine severity and the participants' ideological positions did not reach statistical significance. There were no ideologically-based differences in respondents' self-ratings of objectivity.

98. Analysis of variance result (statistically nonsignificant):  $F(1, 35) = .39, p = .54, \eta^2 = .01$ .

eliminating the motivated imputation of harm based on their own ideological positions.<sup>99</sup> The results of this final experiment also addressed a potential concern that the participants in these studies may not have understood, despite the given definition of harm, that offense to one's values did not count as harm under the terms of the legal constraint.

#### IV. DISCUSSION

##### *A. Review of Findings*

The present research sought to scientifically explore outcome-driven perceptions of harm in the context of the long-debated role of harm in criminal regulation. We found that although the harm principle dominates people's intuitions about punishment, there are acts—like public nudity and flag defiling—that people want to criminalize even without finding harm. Yet, when informed that the law requires a finding of harm in order to penalize such conduct, those who want to punish are significantly more likely to report the presence of harm, and report significantly greater levels of harm. The participants in our studies did not alter their criminalization goals in adherence to the necessity-of-harm constraint, but rather, recruited the harms that they needed to justify their desired punishment outcomes, ostensibly within the terms of the law. Our data thereby provided an experimental demonstration of the predicted plasticity of harm effect.

We ruled out the nonmotivational alternative—that the legal constraint inspired a more intense but essentially objective search for harm—by showing that the imputing of harm is directionally exacerbated by a legally irrelevant ideological factor. Disagreement with a public nudist's message on abortion enhanced people's motivations to penalize the nudity, which led to higher punishment recommendations and correspondingly more recruiting of harm in order to justify punishing within the terms of the necessity-of-harm constraint. However, the respondents' explanations of the perceived harm focused only on

---

99. In both the Incongruent and Congruent conditions of Studies 3 and 4 (in which all the respondents received the legal constraint), mean reports of harm were higher than those in the Constraint condition of Study 2—possibly indicating that an ideologically motivated nudist was seen as more harmful than a nudist without an expressive message, regardless of the position he was advocating. The higher reporting of harm in Study 4 as compared to Study 2 could also suggest that, while confronting the potential of an ideological bias curtailed the recruiting of harm based on ideological opposition, it did not reduce the underlying imputation of harm to public nudity when people wanted to criminalize the conduct (i.e., the original harm plasticity effect demonstrated in Study 2). However, the capacity for directly comparing the data of Studies 3 and 4 with that of Study 2 is limited, because Studies 3 and 4 used different scales for the continuous measure of harm and the participants were specifically selected for their strong pro-life or pro-choice positions on the issue of abortion (whereas the results of Studies 3 and 4 can be directly compared to each other because both these experiments used the same scales and ideologically selective samples).

the nonideological aspect of the conduct, which was constant across conditions and thus could not account for the significant difference in reports of harm.

The effect of a legally extrinsic motive directionally driving harm judgments was seen among both pro-choice and pro-life participants, suggesting an ideological universality of the tendency to recruit harm in the service of preferred outcomes. However, when the ideological factor in the scenarios was made transparent, the participants refrained from basing their punishment decisions on it, and their reports of harm were no longer motivated by their personal views on abortion. Although the respondents still imputed harm in order to criminalize nudity within the terms of the legal constraint, blocking the illusion of objectivity regarding the expressive message of the nudist eliminated the plasticity of harm that stemmed from ideological incongruency—providing some evidence for the nondeliberate nature of this motivated cognition effect.

### *B. Implications for Harm and Beyond*

The harm principle has played a pivotal role in curtailing the legal enforcement of morality through state regulation, and our work is not intended to suggest that it is now useless or obsolete. To the contrary, the first study demonstrated the dominant role the harm principle implicitly plays in guiding people's instincts about whether or not to criminalize. This in itself is an important finding; it suggests that to some extent, people tend to intuitively follow a utilitarian "harm goes with crime" logic when making these types of legal judgments. We were, however, able to identify a few acts—including public nudity and flag defiling—that people wanted to criminalize even without finding harm, and it was in these scenarios that people imputed harm if necessary to legally justify their moral intuitions about punishment.

Our results suggest that the concept of harm may not be as cognitively stable or reliable as the legal system assumes. This calls into question the laws and policies that are based on perceptions of harm, to the extent that those perceptions may be endogenous to the desired outcomes of legal decision makers. In the current U.S. justice system, "[t]he duty of lawmakers, judges, and citizens to justify their positions on grounds susceptible of affirmation by persons of diverse moral persuasions—paradigmatically, the prevention of harm—is deeply woven into prevailing norms of legal and political discourse."<sup>100</sup> However, our studies demonstrate that while supposedly objective standards like the harm principle appear to provide a way to overcome sectarian biases in legal decision making, the rhetoric of harm can covertly become a conduit for morally or ideologically motivated agendas.

In fact, rather than enforcing objectivity, seemingly neutral legal constraints like the harm principle could ironically just exacerbate people's

---

100. Kahan, *Cognitively Illiberal*, *supra* note 51, at 116.

*illusions* of objectivity about their own judgments. Telling decision makers that they are “satisfying the duty of impartiality when they *sincerely* articulate a secular justification . . . can’t make those persons genuinely impartial. It can only make them less *aware* of the influence that our cultural commitments exert . . . a form of self-misunderstanding to which persons in general are already vulnerable.”<sup>101</sup>

Moreover, our finding that people with conflicting positions on abortion significantly differed in their perceptions of the harm caused by the scenario involving an abortion-related message—while neglecting to recognize the influence that this ideological factor had on their judgments—helps explain the intensity of the legal conflicts that arise over divisive issues like abortion or gay marriage. The drive to comply with a “neutral” legal constraint (e.g., a requirement of harm in order to punish) can seemingly transform such conflicts into debates about allegedly objective “facts” (e.g., whether or not an act causes harm to others), when perceptions of those “facts” are actually cognitively motivated by people’s subjective moral intuitions. Differing viewpoints may therefore be seen as irrational or dishonest, because they appear to be contrary to objective truths rather than subjective preferences.<sup>102</sup>

The implications of our results are not confined to perceptions of harm in criminal judgments. Similar forms of motivated cognition could influence perceptions of other elements of an alleged crime, such as causation and intent, so our results dovetail well with the empirical research on criminal blame discussed in Part II.<sup>103</sup> Furthermore, motivated cognition driven by punishment goals could bias assessments of harm in other realms of law beyond criminal regulation—such as torts litigation, where a requirement of harm is “firmly embedded”<sup>104</sup> and punitive damages are used to punish certain conduct.<sup>105</sup>

---

101. *Id.* at 144.

102. *See generally id.* (discussing people’s dissatisfaction with laws that are allegedly based on generally accepted notions with which they disagree).

103. *Supra* Part II.C.; *see* Alicke, *supra* note 47; Nadler & McDonnell, *supra* note 50.

104. John M. Darley et al., *Doing Wrong Without Creating Harm*, 7 J. EMPIRICAL LEGAL STUD. 30, 53 (2010). In a series of experiments on people’s responses to inchoate torts, Darley and others reported that retributive sanctions (punitive damages, fines, prison terms, etc.) were “largely unaffected by whether the harm materialized, but were instead sensitive to whether the wrongdoer exhibited negligent or reckless conduct.” *Id.* at 30. This is consistent with our suggestion that judgments in our scenarios were driven more by the desire to punish the actors’ negligent or reckless disregard for societal or personal values than by actual harm caused.

105. *See Exxon Shipping Co. v. Baker*, 554 U.S. 471, 492 (2008) (“Regardless of the alternative rationales over the years, the consensus today is that punitives are aimed not at compensation but principally at retribution and deterring harmful conduct.”); Cass R. Sunstein, *Moral Heuristics*, 28 BEHAV. & BRAIN SCI. 531 (2005) (suggesting that people are retributively motivated when assigning punitive damages, based on the moral outrage they feel toward the transgression in question).

*C. Constitutional Consequences*

Some legal scholars have suggested that “constitutional theorists have paid too much attention to explicating the normative content of various . . . standards and too little to the psychology of enforcing them. . . . The fit—or lack thereof—between [legal] doctrines and the psychological dispositions of constitutional decisionmakers has been almost entirely neglected.”<sup>106</sup> Our findings speak to this concern.

Study 2, for example, demonstrated that the *Lawrence* Court’s “constitutionalizing” of a harm requirement for state intervention may be circumvented by the cognitive recruiting of harm when decision makers are motivated to punish. Outcome-driven cognition can also directly undermine constitutional principles like the First Amendment, as illustrated in Study 3. The respondents assigned higher punishment to the nudist when his message was incongruent with their own personal beliefs—a feature of the scenario that was not only legally irrelevant, but also unconstitutional to factor into their judgments. As the Supreme Court stated in *Texas v. Johnson* (holding that burning a U.S. flag is protected expression): “If there is a bedrock principle underlying the First Amendment, it is that the government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable.”<sup>107</sup> Even if public nudity were to objectively cause harm, the degree of harm should not differ based on whether the nudist was pro-choice or pro-life; the Constitution would require uniform applications of the law to all such nudists, regardless of the ideological position they advocate and independent of the consonance between their message and the views of those judging them.<sup>108</sup>

The constitutional consequences of motivated cognition are all the more insidious for the legal system given that the process occurs without the decision makers’ full awareness and can produce biased judgments that outwardly appear to have been made within the parameters of the law. Furthermore, people’s adherence to their punishment instincts leads to a powerful psychological effect; in our studies, the respondents’ motivation to penalize morally offensive behavior made it worth the cognitive effort, albeit nondeliberate in nature, of recruiting the harm necessary to remain in seeming compliance with the law.

---

106. Kahan, *They Saw a Protest*, *supra* note 51, at 34, 46. Our findings complement Kahan and his coauthors’ recent study demonstrating “the unconscious influence of individuals’ group commitments on their perceptions of legally consequential facts.” *Id.* at 1. The researchers found that participants’ motivation to reach an outcome that was congruent with their own cultural outlook “eviscerated” the significant legal line between constitutionally protected speech and legally answerable conduct, leading them to conclude that “what people see will often be a reflection of what they value.” *Id.* at 31, 44.

107. *Texas v. Johnson*, 491 U.S. 397, 414 (1989); *see* U.S. CONST. amend. I.

108. *See* U.S. CONST. amends. I, IV.

*D. Contribution to Theories of Punishment and Moral Reasoning*

The finding that people will impute harm to a conduct in order to fulfill their desire to criminalize it within the utilitarian boundaries of a given law offers a new perspective to the discourse on theories of punishment. The present studies were not designed to demonstrate that the participants' punishment motives were retributive (i.e., punishment as an end in itself, because the transgressor "deserves" it), as opposed to utilitarian (i.e., punishment as a means to an end). In fact, since the participants reported that the given scenarios violated their personal and social values, one could make a utilitarian argument for deterring the acts even if they did not cause any "harm" to others. However, without a finding of harm, penalizing the conduct would violate the specific utilitarian goal of the harm principle (i.e., to prevent harm to others), which was imposed as a legal constraint in these experiments.

Prior research has suggested that utilitarian legal constraints that fail to reflect the public's intuitions of justice bear a risk of defiance.<sup>109</sup> In the context of criminal law, it has been argued that the "failure to criminalize certain conduct, which the community finds morally offensive, . . . call[s] into question the moral judgment of the code drafters, which in turn may undercut the law's moral voice . . . [and] reduce the criminal law's compliance power."<sup>110</sup> However, our results indicate that rather than intentionally defying a law that would lead to an undesirable outcome, people may less than consciously reason their way toward meeting the requirements of the legal rule in a manner that supports their criminalization goals. The participants who wanted to penalize the moral violations in our studies did not blatantly ignore the utilitarian constraint that required a finding of harm in order to punish; instead, they recruited the harm necessary to fulfill their desired outcomes while appearing to remain in compliance with the given law. Thus, motivated cognition could enable the fulfillment of popular punishment impulses seemingly within the terms of utilitarian criminal justice policies.

Our findings also offer a new perspective to dual-process theories of moral reasoning that differentiate between deontological judgments that are "driven by automatic emotional responses" and utilitarian judgments that are "driven by controlled cognitive processes."<sup>111</sup> In our studies, respondents

---

109. See TOM R. TYLER, *WHY PEOPLE OBEY THE LAW* (1990) (finding that procedural justice is the best predictor of voluntary compliance with the law); John M. Darley et al., *The Ex Ante Function of the Criminal Law*, 35 L. & SOC'Y REV. 165 (2001) (arguing that people are more likely to comply with the law when it accords with their intuitions of justice); Paul H. Robinson & John M. Darley, *The Utility of Desert*, 91 NW. U. L. REV. 453, 454, 477-78 (1997) (arguing for "a criminal law based on the community's perceptions of just desert" because "every deviation from a desert distribution can incrementally undercut the criminal law's moral credibility").

110. PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* 202 (1995).

111. Joshua D. Greene, *Dual-Process Morality and the Personal/Impersonal Distinction: A Reply to McGuire, Langdon, Coltheart, and MacKenzie*, 45 J. EXPERIMENTAL SOC. PSYCHOL. 581,

eliminated the potential conflict between their deontologically-driven desire to criminalize the offensive behavior (i.e., the public nudist or flag defiler “ought” to be punished) and the utilitarian legal constraint (i.e., punishment requires a finding of harm) by imputing the harm necessary to ensure that “the right course of action morally becomes the right course of action practically as well.”<sup>112</sup> Thus, people’s deontological impulse to punish the offensive behaviors fuelled their utilitarian assessments about the presence of harm.<sup>113</sup> This result accords with other recent research indicating that “utilitarian and deontological rationales are often more complementary than hydraulic, and . . . individuals may construct utilitarian support for what have typically been taken as deontologically-based moral stands.”<sup>114</sup>

### *E. Legal Applications*

#### *1. Jury Decision Making*

Since the participants in our studies were ordinary citizens, the implications of our results are particularly worth considering in contexts of legal decision making by lay people. A robust body of empirical research on jury decision making, for example, has shown that jurors’ “naive representation of legal concepts” can trump courts’ instructions on the law,<sup>115</sup> that lay adjudicators will “alter their own interpretations of the facts to satisfy both the law and their own theories of justice”;<sup>116</sup> and that mock jurors rely on their own theories of criminal responsibility to reach verdicts independent of a judge’s legal instructions.<sup>117</sup> These findings are consistent with our demonstration that even when people are trying to abide by the constraints of a law, their end criminalization goals might drive their perceptions. Applied to jury decision making, our results suggest that jurors may be motivated, without full awareness, to mold factual determinations (such as whether or not an act causes harm) to justify a desired punishment outcome based on legally irrelevant or inappropriate factors, while appearing to remain within the constraints of the

---

581 (2009). *But see* Joshua D. Greene et al., *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*, 44 NEURON 389, 397 (2004) (recognizing that “all action, whether driven by ‘cognitive’ judgment or not, must have some affective basis. Even a cold, calculating utilitarian must be independently motivated, first, to engage in the reasoning that utilitarian judgment requires.”).

112. Brittany Liu & Peter H. Ditto, *What Dilemma? Moral Intuitions Shape Factual Beliefs*, 13–14 (June 28, 2011) (unpublished manuscript), available at <http://ssrn.com/abstract=1829825>.

113. Retributive grounds for punishment are typically justified “on the basis of fulfilling a deontological moral mandate”—that is, the notion that certain behavior should not be tolerated regardless of its practical consequences. Robinson & Darley, *supra* note 109, at 454.

114. Liu & Ditto, *supra* note 112, at 6.

115. Vicki L. Smith, *When Prior Knowledge and Law Collide: Helping Jurors Use the Law*, 17 L. & HUM. BEHAV. 507, 508 (1993).

116. Amiram Elwork & Bruce D. Sales, *Jury Instructions*, in THE PSYCHOLOGY OF EVIDENCE AND TRIAL PROCEDURE 280 (S.M. Kaplan & L.S. Wrightsman eds., 1985).

117. Richard L. Wiener et al., *The Psychology of Jury Nullification: Predicting When Jurors Disobey the Law*, 21 J. APPLIED SOC. PSYCHOL. 1379, 1383 (1991).

jury instructions (loosely analogous to our legal constraint requiring harm in order to punish).

In fact, the psychological effect shown using hypothetical scenarios in our studies could be even stronger when there is a clash between real-life punishment motives and legal constraints, and when jurors' decisions have tangible consequences. In one study demonstrating how people bias their interpretation of evidence in favor of their preferred verdicts during the course of a trial, researchers reported that this "predecisional distortion" was twice as high among prospective jurors as compared to a student sample—with the prospective jurors exhibiting "greater reliance on their prior beliefs, and more confidence in their tentatively leading verdicts."<sup>118</sup>

The present research (i.e., Study 3) suggests that juries may be particularly vulnerable to outcome-driven cognition in cases involving the First Amendment. When motivated to punish a message they dislike, jurors may subjectively reason their way toward a judgment that violates the constitutional right to free speech. A real-life example could perhaps be seen in *Newton v. National Broadcasting Company*, a defamation action that the popular entertainer Wayne Newton brought against NBC for airing a news broadcast that alleged Mafia involvement in his purchase of a gaming casino.<sup>119</sup> The jurors found in favor of Newton, and the district court upheld the jury's award of \$225,000 for "physical and mental suffering" and \$5 million in punitive damages.<sup>120</sup> The district court, however, overturned the \$5 million award for injury to plaintiff's reputation, concluding that Newton's continued success and accolades indicated that his reputation had not suffered any harm.<sup>121</sup>

On appeal, the Ninth Circuit identified a risk arguably akin to motivated cognition in the jurors' judgments: "Wayne Newton's case poses the danger that First Amendment values will be subverted by a local jury biased in favor of a prominent local public figure against an alien speaker who criticizes that local hero."<sup>122</sup> Noting that the jury needed to find "actual malice" in order to punish the defendant, of which there was "almost no evidence . . . much less clear and convincing proof," the court reversed the judgment against NBC.<sup>123</sup> The jurors in that case may have been motivated to impute malice to NBC and harm to

---

118. Kurt A. Carlson & J. Edward Russo, *Biased Interpretation of Evidence by Mock Jurors*, 7 J. EXPERIMENTAL PSYCHOL.: APPLIED 91, 99 (2001). See also Wiener et al., *supra* note 117, at 1396 (noting that community resident participants relied more on their own attributions of responsibility as compared to college student participants, because "college students are accustomed to following directions in the academic environment," whereas real jurors, like community residents, "are out of the habit of following instructions"). Given that the participants in our first two studies were college students, this suggests that our harm plasticity finding may be even stronger with a participant pool more akin to adult jurors.

119. 677 F. Supp. 1066 (D. Nev. 1987).

120. *Id.* at 1069.

121. *Id.* at 1068–69.

122. *Newton v. Nat'l Broad. Co.*, 930 F.2d 662, 671 (9th Cir. 1991).

123. *Id.* at 687.



Newton's reputation in order to punish the offensive portrayal of a beloved entertainer within the given legal standard.

Application of the present research to real jury decision making is limited, however, by the fact that these studies illustrate motivated cognition only at the individual level. There is some evidence that the process of group deliberation among jurors may actually exacerbate the effect,<sup>124</sup> but various factors influence the relative magnitude of individual versus group biases.<sup>125</sup> Therefore, further studies need to more specifically investigate how the dynamics of group decision making affect motivated cognition in legal contexts.

## 2. Voter-Based Initiatives

Motivated cognition can endanger constitutional principles not only through jury decision making, but also through voter-determined referenda. In the 2010 *Perry v. Schwarzenegger* case with which we introduced this Essay, a California district court unmasked possible attempts by proponents of Proposition 8 to impute legally cognizable harm to same-sex marriage in order to pass and defend the ballot initiative restricting marriage to opposite-sex couples.<sup>126</sup> Voters were not explicitly told that they needed to find harm in order to eliminate same-sex marriage, but campaign messages for Proposition 8 that evoked fears about same-sex relationships seemed targeted toward recruiting harms. Yet, when the district court directly asked the proponents of Proposition 8 to substantiate the alleged harms caused by same-sex marriage—to show that the initiative was within the constitutional constraints of the Due Process and Equal Protection Clauses—they were unable to do so.

The official advertising campaign in favor of Proposition 8 “never articulated” why the public needs to be protected from same-sex marriage.<sup>127</sup> Rather, the campaign “relied heavily on negative stereotypes . . . and . . .

---

124. See, e.g., David G. Myers & Martin F. Kaplan, *Group-Induced Polarization in Simulated Juries*, 2 PERSONALITY & SOC. PSYCH. BULL. 63 (1976) (finding that jury deliberations polarized mean judgments, whereby, after group discussion in simulated jury settings, respondents recommended more lenient punishment in low guilt cases and more severe punishment in high guilt cases). But see Norbert L. Kerr et al., *Bias in Judgment: Comparing Individuals and Groups*, 103 PSYCHOL. REV. 687 (1996) (providing overview of empirical literature revealing no clear pattern as to whether decision making is more biased among individuals or groups).

125. Kerr et al., *supra* note 124, at 687 (noting that factors influencing individual versus group bias include “group size, initial individual judgment, the magnitude of bias among individuals, the type of bias, and most of all, the group-judgment process”).

126. See *Perry v. Schwarzenegger*, 704 F. Supp. 2d 921, 938 (N.D. Cal. 2010); *supra* Introduction (introducing the *Perry* case).

127. *Id.* at 988; see also *id.* at 937 (describing plaintiffs’ historian expert’s testimony that “the Proposition 8 campaign did not need to explain what children were to be protected from; the advertisements relied on a cultural understanding that gays and lesbians are dangerous to children”); *id.* at 1003 (describing video advertisement that depicted a young girl asking viewers if they had considered the consequences to her of Proposition 8, without explaining what those consequences were).

inchoate threats vaguely associated with gays and lesbians,” such as fears that religious individuals would “somehow be harmed” by same-sex marriage<sup>128</sup>; that gays and lesbians are more likely to molest children;<sup>129</sup> and that exposure to same-sex relationships “would turn children into homosexuals.”<sup>130</sup> One of the plaintiffs testified:

[P]rotect[ing] the children [was] a big part of the [Proposition 8] campaign. And when I think of protecting your children, you protect them from people who will perpetrate crimes against them, people who might get them hooked on a drug, a pedophile, or some person that you need protecting from. . . . You protect [them] from things that can harm you physically, emotionally.<sup>131</sup>

Another gay plaintiff testified that the campaign attempted to “convince people that there was a great evil to be feared and that evil must be stopped and that evil is us.”<sup>132</sup>

During oral arguments for a summary judgment motion in the case, the proponents’ counsel was asked to explain how permitting same-sex marriage harmed the state’s interest in marriage. The court described the attorney’s response as follows: “[W]hen pressed for an answer, counsel replied: ‘Your honor, my answer is: I don’t know. I don’t know.’ . . . [P]roponents in their trial brief promised to ‘demonstrate that redefining marriage to encompass same-sex relationships’ would effect some twenty-three specific harmful consequences.”<sup>133</sup> This exchange illustrates a shift from the harm-versus-morality discourse on same-sex relationships seen in the Supreme Court’s *Bowers* and *Lawrence* cases discussed in Part I,<sup>134</sup> to the couching of aggrieved moral sensibilities as allegedly cognizable harms.

The district court judge ultimately determined, however, that the proponents of Proposition 8 “provided no credible evidence to support any of the claimed adverse effects [they] promised to demonstrate.”<sup>135</sup> Although their advertising campaign “ensured California voters had . . . fear-inducing messages in mind,” those fears were shown at trial to be “completely unfounded.”<sup>136</sup> The desire to pass Proposition 8 may have led its supporters to impute harm to same-sex marriage in their attempt to eliminate it within the

128. *Id.* at 1003.

129. *Id.* at 937 (describing testimony about a website that “encouraged voters to support Proposition 8 on grounds that homosexuals are twelve times more likely to molest children”).

130. *Id.* at 1003 (describing video of mother responding with horror to her daughter saying she can marry a princess).

131. *Id.* at 939.

132. *Id.* at 938.

133. *Id.* at 931 (citing Doc. # 228 at 23; Doc. # 295 at 13–14).

134. *See supra* Part I.A (discussing *Bowers* and *Lawrence*).

135. *Perry*, 704 F. Supp. 2d at 931.

136. *Id.* at 1002–03 (“The evidence at trial regarding the campaign to pass Proposition 8 uncloaks the most likely explanation for its passage: a desire to advance the belief that opposite-sex couples are morally superior to same-sex couples.”).

boundaries of the Constitution. Meanwhile, the plaintiffs presented social epidemiology, political science, economics, history, and psychology experts to testify about specific and concrete harms that Proposition 8 caused to same-sex couples.<sup>137</sup> In addition, they provided evidence that permitting same-sex marriage would *not* harm the state's interest in opposite-sex marriage.<sup>138</sup> To this point, the judge noted that the state "ha[d] already issued 18,000 marriage licenses to same-sex couples and ha[d] not suffered any demonstrated harm as a result."<sup>139</sup>

The district court concluded that Proposition 8 violated the Constitution's Equal Protection and Due Process Clauses, because there was no compelling state interest that justified denying same-sex couples the right to marry and no rational basis for limiting marriage to opposite-sex couples.<sup>140</sup> Explaining its scrutiny of the voter-determined referendum, the court stated: "[T]he voters' determinations must find at least some support in evidence. . . . Conjecture, speculation and fears are not enough. Still less will the moral disapprobation of a group or class of citizens suffice, no matter how large the majority that shares that view."<sup>141</sup>

On appeal, the Ninth Circuit agreed that the proponents of Proposition 8 had not substantiated the harms of same-sex marriage that they implied in their advertising campaign and alleged in court to pass the bar of constitutionality,<sup>142</sup> whereas "[r]estricting access to the designation of 'marriage' did . . . 'work[] a real and appreciable harm upon same-sex couples and their children.'"<sup>143</sup> Suggesting that imputed harm cannot survive rational basis scrutiny, the

---

137. Some of the plaintiffs' enumerated harms included the allegations that Proposition 8 "stigmatizes gays and lesbians"; "provides state endorsement of private discrimination"; "increases the likelihood of negative mental and physical health outcomes for gays and lesbians"; and "[creates] tangible economic harms . . . from being unable to marry, including lack of access to health insurance and other employment benefits, higher income taxes and taxes on domestic partner benefits." *Id.* at 935, 961. The plaintiffs' experts also explained how Proposition 8 caused economic harm to the State of California: lowering revenue by "slash[ing] the number of weddings performed"; decreasing the number of married couples "who tend to be wealthier than single people because of their ability to specialize their labor, pool resources and access state and employer-provided benefits"; and increasing "costs associated with discrimination against gays and lesbians." *Id.* at 938.

138. *See, e.g., id.* at 934–35 (describing testimony from a psychologist expert on couple relationships who testified that permitting same-sex marriage would not harm opposite-sex marriage).

139. *Id.* at 1003.

140. *Id.* at 995, 997. In fact, the opinion concluded that Proposition 8 actually "*harms* the state's interest in equality." *Id.* at 998 (emphasis added).

141. *Id.* at 938.

142. *Perry v. Brown*, 671 F.3d 1052, 1092 (9th Cir. 2012) (observing that the ballot initiative "could not have reasonably been enacted to promote childrearing by biological parents, to encourage responsible procreation, to proceed with caution in social change, to protect religious liberty, or to control education of schoolchildren").

143. *Id.* at 1067 (emphasis added). Explaining the balancing of harm in its determination, the court noted: "Proposition 8 works a meaningful harm to gays and lesbians . . . and this harm must be justified by some legitimate state interest. *Id.* at 1081. "Proponents need not show that they would suffer any personal injury from the invalidation of Proposition 8. That the *State* would suffer an injury is enough . . ." *Id.* at 1074.

appellate court noted: “While deferential, the rational-basis standard ‘is not a toothless one.’ [E]ven the standard of rationality . . . must find some footing in the realities of the subject addressed by the legislation.”<sup>144</sup>

Ultimately, both the district and appellate court decisions in the *Perry* litigation circled back to the question addressed by the Supreme Court in *Lawrence v. Texas*, which has been central to legal applications of the harm principle: “[W]hether a majority of citizens could use the power of the state to enforce ‘profound and deep convictions accepted as ethical and moral principles’ through the criminal code. The question here is whether California voters can enforce those same principles through regulation of marriage licenses. They cannot.”<sup>145</sup>

### 3. *Judicial Safeguards?*

The judiciary intervened in the *Newton* and *Perry* cases to prevent lay decision makers from undermining important constitutional principles. However, this safeguard is only as strong as the ability of judges themselves to resist the unintentional psychological tendency to reason toward desired legal outcomes. As observed by Kahan, although courts are required “to ‘flush out’ the impact, conscious or unconscious, of regulators’ animosity toward those whose identity or values defy dominant norms,” if professional legal decision makers are cognitively motivated, then “they—like everyone else—are more or less likely to *see* challenged laws as contributing to attainment of secular ends depending on whether those laws affirm or denigrate their own cultural commitments.”<sup>146</sup>

In *Perry*, for example, a judge with a directional motivation akin to that of the proponents of Proposition 8 might have subconsciously joined them in imputing a certain type of “harm” to same-sex marriage. Actually, the proponents in that case alleged motivated cognition in the other direction: after deciding the case, the district court judge disclosed that he was gay, and the proponents of Proposition 8 filed a motion to vacate his judgment on the basis that his potential interest in marrying his partner may have led to biased reasoning.<sup>147</sup> The motion was denied, and the Ninth Circuit agreed, holding that it is not “‘reasonable to presume’ . . . ‘that a judge is incapable of making an impartial decision about the constitutionality of a law, solely because, as a citizen, the judge could be affected by the proceeding.’ To hold otherwise would demonstrate a lack of respect for the integrity of our federal courts.”<sup>148</sup>

---

144. *Id.* at 1089 (internal citations omitted).

145. *Perry*, 704 F. Supp. 2d at 1002 (citing *Lawrence v. Texas*, 539 U.S. 558, 571 (2003)); *see id.* at 1003; *Perry v. Brown*, 671 F.3d at 1094; *supra* Part I.A (discussing *Lawrence*).

146. Kahan, *They Saw a Protest*, *supra* note 51, at 33.

147. *Perry v. Brown*, 671 F.3d at 1095.

148. *Id.* at 1096. Noting that “[c]ourts have repeatedly held that matters such as race or ethnicity are improper bases for challenging a judge’s impartiality,” the Second Circuit has asserted: “A suggestion that a judge cannot administer the law fairly because of the judge’s racial and ethnic

Given, however, that the motivated cognition process operates under an illusion of objectivity, the integrity of the decision maker is not necessarily in question; a preferred outcome could inadvertently influence even someone with the best of intentions.<sup>149</sup> Moreover, legal training does not necessarily provide inoculation against this cognitive pull. In fact, researchers have presented some evidence that motivated cognition may be “stronger and more consistent” among legally trained participants (law students), as compared to those without legal training.<sup>150</sup> This may not be surprising given that lawyers in the U.S. adversarial system are trained to engage in outcome-driven reasoning in order to provide their clients with the strongest advocacy for their cases. When those lawyers become law clerks or judges, it may not be easy for them to discard the reinforced inclination to garner directional arguments in favor of a desired outcome.

Experimental studies have provided some evidence for reduced cognitive biases among judges as compared to lawyers and lay people, but not for a lack of judicial bias altogether.<sup>151</sup> In one experiment that examined decision makers’ evaluations of social science evidence in death penalty cases, motivated cognition was seen more extensively in decisions of law students than judges, but judges did exhibit the same biases when it came to the “more subjective and value laden” task of determining the weight of the evidence once it had been

---

heritage is extremely serious and should not be made without a factual foundation going well beyond the judge’s membership in a particular racial or ethnic group. Such an accusation is a charge that the judge is racially or ethnically biased and is violating the judge’s oath of office.” *MacDraw v. CIT Grp. Equip. Fin.*, 138 F.3d 33, 37 (2d Cir. 1998); see also *Blank v. Sullivan & Cromwell*, 418 F. Supp. 1, 4 (S.D.N.Y. 1975); *Pennsylvania v. Local Union 542, Int’l Union of Operating Eng’rs*, 388 F. Supp. 155, 163 (E.D. Pa. 1974)).

149. This is not to say that there was any inadvertent bias operating in the *Perry* district court decision or that the judge should have recused himself. Highlighting the futility of trying to recuse judges on this basis, when Judge Constance Baker Motley was petitioned to recuse herself from a sex discrimination case on the grounds that her background as a black woman and civil rights litigator could make her identify with the plaintiff and thus be motivated to rule in her favor, she rejected the motion, stating: “Indeed, if background or sex or race of each judge were, by definition, sufficient grounds for removal, no judge on this court could hear this case, or many others, by virtue of the fact that all of them were attorneys, of a sex, often with distinguished law firm or public service backgrounds.” *Blank*, 418 F. Supp. at 4.

150. Eileen Braman & Thomas E. Nelson, *Mechanism of Motivated Reasoning? Analogical Perception in Discrimination Disputes*, 51 AM. J. POL. SCI. 940, 952 (2007).

151. See, e.g., Andrew J. Wistrich et al., *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding*, 153 U. PA. L. REV. 1251, 1259 (2005) (reporting experimental findings that judges were able to avoid being influenced by only some but not most kinds of inadmissible information in their decision making); Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777 (2001) (reporting that all five of the cognitive illusions that were tested for influenced judicial decision making—less so for framing effects and the representativeness heuristic, but at comparable levels to lay people for anchoring effects, hindsight bias, and egocentric bias); Jeffrey J. Rachlinski et al., *Context Effects in Judicial Decision Making* 18, 31 (Aug. 3, 2009) (unpublished manuscript), available at <http://ssrn.com/paper=1443596> (painting “intricate portrait” of how judges “navigat[e] the risk of cognitive errors in the courtroom,” both successfully and unsuccessfully).

admitted, which can be critical to the outcome of cases.<sup>152</sup> Moreover, consistent with the illusion of objectivity in motivated cognition, the judges were more confident than the law students that “other lawyers would agree with” their judgments, even though there was actually greater variability in their ratings.<sup>153</sup> Judicial samples are difficult to come by, but further experimental research is needed to explore the extent to which judges may be subconsciously susceptible to the type of motivated cognition demonstrated in the present studies.

#### *F. Potential Remedies*

Although the implications of the present research are sobering, the results of the final experiment provide some cause for optimism by demonstrating people’s unwillingness to *purposefully* engage in the motivated recruiting of harm to achieve their desired punishment outcomes. Furthermore, this finding suggests a potential pathway to remedies for the problem of outcome-driven reasoning: calling attention to potentially biasing extralegal factors could better equip decision makers to resist their motivating influence.

For example, in a case like *Newton*,<sup>154</sup> discussed above, jurors could be instructed to consider that their personal feelings toward the plaintiff may inappropriately color their judgments. Indeed, recent research indicates that piercing the illusion of objectivity merely by generating awareness of the decision makers’ susceptibility to potentially biasing motives through a simple instruction can curb the motivated cognition effect.<sup>155</sup> Alternatively, to combat the illusion of objectivity that might persist in the face of hypothetically phrased instructions, the legal system could institute a process of citizen appellate review to be invoked in particularly risky cases, whereby lay decision makers are presented with the relevant facts of a case minus the identities of the parties (or the ideological positions, or whatever factor that might trigger the potential for motivated cognition), and then asked to review the original jury’s verdict in this anonymized context. Further work, however, is required to test the value and feasibility of these and other possible ways in which to alleviate the risk of motivated cognition in legal judgments.

#### CONCLUSION

The program of research described in this Essay offers a novel contribution to over a century of discourse on the criminal enforcement of

---

152. Richard E. Redding & N. Dickon Reppucci, *Effects of Lawyers’ Socio-Political Attitudes on Their Judgments of Social Science in Legal Decision Making*, 23 L. & HUM. BEHAV. 31, 48 (1999).

153. *Id.* at 41.

154. *See supra* Part IV.E.1 (discussing *Newton*).

155. Avani Mehta Sood & Joel Cooper, *Debiasing the Motivated Reasoner* (unpublished manuscript in preparation) (on file with author); *see* Richard E. Petty et al., *Flexible Correction Processes in Social Judgment: Implications for Persuasion*, 16 SOC. COGNITION 93 (1998).

morality and to the emerging body of experimental work on motivated cognition in the realm of law. The outcome-driven perception suggested by our results is at odds with the objective, perception-driven outcomes toward which our justice system strives. With these demonstrations of the motivated plasticity of harm in the service of criminalization goals, we hope to encourage a more honest confrontation of the covert psychological biases that can inadvertently govern legal decision making.