

April 10, 2018

Dear colleagues,

Thank you for making the time for our workshop!

It comes at a very useful moment for the project. I've been running a series of experiments on whether the law can influence our moral intuitions. The first two rounds of experiments were reported in an earlier essay reviewing philosopher Frances Kamm's book on the trolley problem.

I'm now chasing down some questions raised by those initial findings, with a third and fourth round of experiments, and my hope is that these will coalesce into a new paper. At the workshop, I'll be showing you results from the third round and asking for your suggestions about how to set up the fourth round. And so I'm including here two documents:

[1] My earlier essay, motivating the project and describing the first two rounds of experiments. (You might skip Part I, which is very much in the weeds of the book.)

[2] A description of the third round, with a tentative discussion of its results, as well as my current ideas for designing the fourth round.

Please excuse me for not offering a proper new paper—I'm extremely grateful to get to hear your suggestions now, at this in-between moment. Any thoughts you may have are also always welcome at bhuang@law.columbia.edu. Thank you!

Looking forward,

Bert

BOOK REVIEW

LAW AND MORAL DILEMMAS

THE TROLLEY PROBLEM MYSTERIES. By F.M. Kamm. New York, N.Y.: Oxford University Press. 2015. Pp. xi, 270. \$29.95.

*Reviewed by Bert I. Huang**

INTRODUCTION

A runaway trolley rushes toward five people standing on the tracks, and it will surely kill them all. Fortunately, you can reach a switch that will turn the trolley onto a side track — but then you notice that one other person is standing there. Is it morally permissible for you to turn the trolley to that side track, where it will kill one person instead of five? Is it not only morally permissible, but even morally *required*? This classic thought experiment is a mainstay in the repertoire of law school hypotheticals, often raised alongside cases about cannibalism at sea, tossing people from overcrowded lifeboats, or destroying buildings to save a city from fire.¹

The liturgy of the trolley problem is one we all know. It is a call-and-response, suitably Socratic in style. The question above is asked. The unsuspecting subject will firmly reply: “Yes, it must be morally permissible to avoid five deaths, when the alternative is a single death.” The confidence of this reply gives the next question its zing: “What if you cannot divert the trolley, but you can stop it by pushing someone in front of it? No? Yet isn’t this also sacrificing one to save five?”²

This venerable vignette and its still darker variations have continued to capture our imagination beyond the classroom as well — informing public thought on subjects ranging from climate change, to abortion, to bioethics, to capital punishment, to takings and public ne-

* Professor of Law, Columbia Law School. I wish to thank Jessica Bulman-Pozen, Glenn Cohen, Elizabeth Emens, Shannon Fanning, Jeffrey Gordon, Scott Hemphill, Kathryn Judge, Frances Kamm, Jeremy Kessler, Benjamin Liebman, Katerina Linos, David Pozen, Eric Rakowski, Frederick Schauer, and Barbara Spellman for insightful conversations and comments on drafts. Arisa Akashi, Bram Schumer, and Yuko Sin provided excellent research assistance, and I am grateful for what they have taught me. This project has been funded by a Global Innovation Award from Columbia Law School, and I also thank the grant committee for their suggestions.

¹ See *Regina v. Dudley & Stephens* (1884) 14 Q.B.D. 273; *United States v. Holmes*, 26 F. Cas. 360 (C.C.E.D. Pa. 1842) (No. 15,383); *Bowditch v. Boston*, 101 U.S. 16 (1880); cf. Lon L. Fuller, *The Case of the Speluncean Explorers*, 62 HARV. L. REV. 616 (1949).

² We shall soon see what happens when someone answers the first question with, “No, it is *not* morally permissible to turn the trolley.”

cessity, to killing in wartime, and to torture.³ At a higher level of abstraction, it is invoked in controversies over how to mark the ethical boundaries of contested modes of regulation: When should consequentialist approaches to law or policy (such as cost-benefit analysis, optimal deterrence, and social-welfare aggregation) be trumped by deontological commands (such as “do not kill”), and vice versa?⁴

Professor Frances Myrna Kamm, a moral philosopher who for decades has been a leading analyst of this thought experiment,⁵ has now published her richly stimulating Tanner Lectures. Joining her lectures as chapters in the book are a trio of rigorous and unrelenting responses from a panel of philosophers — Professors Judith Jarvis Thomson, Thomas Hurka, and Shelly Kagan — as well as an introduction by legal scholar Professor Eric Rakowski.

The book captures a moment of intellectual unrest arising from a sudden about-face by Thomson — a “founding mother” of the trolley problem (p. 21) — concerning one of the core premises underlying a generation of trolley debates. In short, Thomson, who first established the standard view that it is morally permissible for a bystander to turn

³ See, e.g., Daniel Epps, *The Consequences of Error in Criminal Justice*, 128 HARV. L. REV. 1065, 1134–35 (2015) (relating the trolley problem to criminal punishment); Susan S. Kuo, *Disaster Tradeoffs: The Doubtful Case for Public Necessity*, 54 B.C. L. REV. 127, 137–40 (2013) (using the trolley problem to discuss the public necessity exception to just compensation for takings); Lawrence Summers & Richard Zeckhauser, *Policymaking for Posterity*, 37 J. RISK & UNCERTAINTY 115, 122–23 (2008) (using the trolley problem to discuss climate change policy); Cass R. Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs*, 58 STAN. L. REV. 703, 742–43 (2005) (drawing implications for capital punishment from the trolley problem); Carol S. Steiker, *No, Capital Punishment Is Not Morally Required: Deterrence, Deontology, and the Death Penalty*, 58 STAN. L. REV. 751, 760–62 (2005) (responding to Sunstein & Vermeule, *supra*); Suzanne Uniacke, *The Doctrine of Double Effect*, in PRINCIPLES OF HEALTH CARE ETHICS 263 (Richard E. Ashcroft et al. eds., 2007) (discussing the trolley problem in relation to questions of health care ethics and bioethics, including euthanasia); David Edmonds, *Lessons in Morality at West Point*, BBC (Sept. 18, 2010, 12:04 PM), <http://news.bbc.co.uk/2/hi/9006784.stm> [<https://perma.cc/H4RG-3NM8>] (discussing use of the trolley problem to teach cadets the ethics of collateral damage in war); David Ropeik, *The Ethics of Climate Change Denial*, BIG THINK (Feb. 2, 2012), <http://bigthink.com/risk-reason-and-reality/the-ethics-of-climate-change-denial-2> [<https://perma.cc/72KA-4LDE>] (using the trolley problem to discuss climate change).

⁴ See, e.g., DOUGLAS A. KYSAR, REGULATING FROM NOWHERE 25–36 (2010); Mark Kelman, *Saving Lives, Saving from Death, Saving from Dying: Reflections on “Over-Valuing” Identifiable Victims*, 11 YALE J. HEALTH POL’Y L. & ETHICS 51, 59–60 (2011); Eric Rakowski, *Taking and Saving Lives*, 93 COLUM. L. REV. 1063, 1063–65 (1993); Cass R. Sunstein, *Moral Heuristics*, 28 BEHAV. & BRAIN SCI. 531, 541 (2005); see also John Mikhail, *Moral Heuristics or Moral Competence? Reflections on Sunstein*, 28 BEHAV. & BRAIN SCI. 557, 557 (2005).

⁵ As one of her commentators, Professor Shelly Kagan, put it: “I think it fair to say that no one has worked harder to solve the trolley problem than Kamm has. Over the years she has probably examined hundreds of different cases, and she has struggled mightily to produce a principle that matches our intuitions about those cases . . .” (p. 155).

the trolley, now argues that it is *not*.⁶ But if turning the trolley is just as verboten as pushing someone in front of it, there is no curious contrast to reconcile — so what remains of the trolley problem?⁷

The beating heart of this book is a fresh, and often raw, analytical quarrel between Kamm and Thomson about this reversal — and about how much of the trolley problem continues to exist.⁸ As Rakowski observes, “[t]his is a truly remarkable exchange between the two leading contributors to this moral philosophical debate”; no doubt “these lectures, commentaries, and replies [will be] absolutely invaluable for future work on the trolley problem” (p. 5).⁹

The book’s timing is impeccable. It arrives just as the trolley problem is gaining newfound attention in the public discourse, due in part to its uncanny resemblance to emergent questions about how to program autonomous vehicles — such as military drones or driverless cars — to act ethically.¹⁰ If your self-driving Volvo may have to decide whether to swerve hard into a wall, killing you, in order to avoid an accident that will kill five others, what should it do?

At this moment of special vitality for the classic thought experiment, I wish to broach a topic that the current trolley debates have tended to neglect: *Can our intuitions about moral dilemmas be influenced by the presence of the law?*

This gap in the philosophical discourse is set in sharp relief by the ample attention paid by lawyers — and by the public — to questions of how and when the law can influence moral beliefs and social norms. Think of Prohibition, *Brown v. Board of Education*, or same-sex marriage. Not surprisingly, legal scholars have long theorized about the law’s role as a source of moral or ethical knowledge.¹¹ Some empirical

⁶ Her reasoning will be detailed below, in Part I, but it is useful to note here that Thomson’s reversal concerns only whether a bystander may turn the trolley. Her answer continues to be “yes” when the person facing the decision is the *driver* of the trolley.

⁷ As Thomson puts it, “I therefore now conclude that there isn’t really any such thing as the [bystander version of the] trolley problem” (p. 117).

⁸ As Kamm describes part of her motivation: “When a major figure whose past work on a problem has been admired and served as a basis for subsequent work . . . changes her mind, one wants to be sure that the reversal is justified, especially since so many still believe there is a Trolley Problem and continue to produce work dealing with it” (p. 50 n.19).

⁹ Yet as Rakowski also notes in his introduction, “[c]onsequentialist views are not discussed at any length” in the book (p. 7 n.4).

¹⁰ See *infra* notes 69–71 and accompanying text.

¹¹ The literature is vast, but for a brief review, see Kenworthy Bilz & Janice Nadler, *Law, Psychology, and Morality*, 50 PSYCHOL. LEARNING & MOTIVATION 101, 107–13 (2009), noting relevant work of legal scholars and cataloguing hypothesized mechanisms for law’s influence on moral cognition. Of course, there is also a voluminous literature debating how the law affects *behaviors* (possibly through morality-related signals), as opposed to how the law influences moral beliefs directly; two recent works offer critical surveys of the major theories. See RICHARD H. MCADAMS, *THE EXPRESSIVE POWERS OF LAW* (2015); FREDERICK SCHAUER, *THE FORCE OF LAW* (2015). For oft-cited early works, see TOM R. TYLER, *WHY PEOPLE OBEY THE LAW*

research on the potential for law to directly influence moral beliefs has also been done¹² — but it appears to be scarce in the context of moral dilemmas such as the trolley problem.¹³

By conducting a set of randomized survey experiments, I hope to jumpstart this line of inquiry.¹⁴ Moral dilemmas are an especially intriguing domain for the study of law's potential influence. They are pervasive in the real world in the form of tragic choices or other harm-harm tradeoffs and are often regulated by law or policy.¹⁵ Moreover, they have a peculiar structure: they pose a contest between deeply felt moral commands. We must save the five. We must not kill the one. These are not mild suggestions. If the law does move the moral needle in such a case, it would be doing so where potent moral intuitions normally govern — a possibility these new experiments aim to test.

Part I elaborates on the exchanges between Kamm and her commentators, highlighting questions raised by their debates about the stability of moral intuitions. Part II explains the design of two experiments I conducted, in which survey subjects are asked their moral in-

(1990); Dan M. Kahan, *Social Influence, Social Meaning, and Deterrence*, 83 VA. L. REV. 349 (1997); Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. CHI. L. REV. 943 (1995); Cass R. Sunstein, *Social Norms and Social Roles*, 96 COLUM. L. REV. 903 (1996).

¹² See Kenworthy Bilz & Janice Nadler, *Law, Moral Attitudes, and Behavioral Change*, in THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW 241, 253–58 (Eyal Zamir & Doron Teichman eds., 2014) (surveying quantitative literature and case studies). For widely cited qualitative accounts, see, for example, MICHAEL J. KLARMAN, FROM JIM CROW TO CIVIL RIGHTS (2006); GERALD N. ROSENBERG, THE HOLLOW HOPE: CAN COURTS BRING ABOUT SOCIAL CHANGE? (2d ed. 2008).

¹³ Bilz & Nadler, *supra* note 11, at 109–10 (surveying empirical literature, noting that “[u]nfortunately, the empirical evidence supporting the claim that law is a persuasive informational source that directly influences [moral] attitudes is thin,” *id.* at 109, and identifying two experimental studies from the 1960s, neither of which focuses specifically on moral dilemmas). A more recent nonexperimental study that did involve moral dilemmas found that in the Netherlands, survey subjects reported moral intuitions corresponding to the act/omission distinction in a variety of scenarios (including the trolley problem) — even though a widely supported Dutch law did not make such a distinction between active and passive euthanasia. M.D. Hauser, F. Tonnaer & M. Cima, *When Moral Intuitions Are Immune to the Law: A Case Study of Euthanasia and the Act-Omission Distinction in the Netherlands*, 9 J. COGNITION & CULTURE 149, 151, 154 (2009).

¹⁴ This approach of using randomized survey experiments follows a growing body of research in experimental philosophy and moral psychology that has studied trolley-type moral dilemmas in depth (but to my knowledge has not yet focused on the influence of law as an exogenous factor). As I observe below, several of these studies are particularly relevant to arguments raised in the book. See *infra* notes 57–58 and accompanying text.

¹⁵ The canonical reference is GUIDO CALABRESI & PHILIP BOBBITT, TRAGIC CHOICES (1978). The universe of moral and ethical dilemmas extends beyond such harm-harm tradeoffs, of course. See, e.g., I. Glenn Cohen, *Rationing Legal Services*, 5 J. LEGAL ANALYSIS 221 (2013) (applying principles developed in bioethics to ethical dilemmas arising from the rationing of legal services); Lani Guinier, *The Supreme Court, 2007 Term — Foreword: Demosprudence Through Dissent*, 122 HARV. L. REV. 4, 66 (2008) (describing a case in which “Justice Stevens admits his own moral dilemma” in that he “now believes that the death penalty is wrong, but feels bound by the Court’s precedent”).

tuitions about trolley scenarios that include randomly assigned information about what the law says. The first experiment follows the standard trolley scenario, with a bystander standing at the switch; the law variously criminalizes or justifies turning the trolley. The second replaces the bystander with a railroad engineer, whose official role includes duties regulated by the law.¹⁶ By varying what the law says, these experiments together explore the role of law — and in particular, the law of roles — in shaping our moral intuitions.

Part III reports the findings, which offer evidence that the presence of law *can* influence our intuitions about the trolley dilemma.¹⁷ Telling subjects that turning the trolley constitutes criminal homicide increases the number who say that doing so is morally prohibited. In comparison, telling subjects that the law requires the engineer to minimize casualties reduces the number who say that turning the trolley is morally prohibited; it also increases the number who say that doing so is morally required. Moreover, the content of the law appears able to exert some influence even when subjects are told that the law will not be enforced.¹⁸

Part III also suggests lines of questioning — for the trolley debates, for empirical research, and for legal design — raised by the following concern: What if the moral intuitions we can observe have already been shaped by people's impressions, however vague or subconscious, about what the *law* expects?

More pragmatic questions arise, as well. In the context of self-driving cars, for example, should regulators hasten to announce that the role of the engineers who program such cars is to minimize casualties — before the public's moral intuitions start to be forged by a series

¹⁶ This change in personnel is motivated by arguments that Kamm presents in her chapter focusing on the question “who turn[s] the trolley” (pp. 11–56), as will be explained below.

¹⁷ As will be apparent from the experiments, by the “presence of law” I mean that their trolley scenarios directly supply information about the law's imperatives and permissions. In this study, the focus is on primary rules of conduct for nonlegal actors, rules that are readily understood as part of what one might call positive law, human law, or first-stage law. Cf. SCHAUER, *supra* note 11, at 70 (borrowing the term “first-stage law” from Professor Ruth Gavison to mean “what the ordinary person and the ordinary official take to be law,” and framing the following inquiry: “whether we call the category first-stage law, positive law, human law, or something else, a persistent issue is whether people should and do act in accordance with the mandates of the components of that category”). The term “law” will take this usage throughout, setting aside usages that refer to natural law or higher law.

¹⁸ This finding is one small step toward addressing the need for research, pressed by Professor Frederick Schauer in recent work, that distinguishes between the law's influence through its substantive content and its influence through the threat of sanctions. See *id.* at 72–73 (“[W]e do not know how much of law's effect on moral and policy attitudes is a function of law's sanction-independent content and how much is a function of the emphasis supplied by the sanction.”).

of tragic accidents?¹⁹ The Conclusion points to questions for further research concerning the possible mechanisms of law's influence.

I. THE NEW TROLLEY DEBATES

The trolley problem has intrigued moral and ethical theorists ever since Thomson coined the term forty years ago.²⁰ Legal and philosophical thinkers have long puzzled over the contrasts between what our moral intuitions say about the most basic variation — whether a bystander may turn the trolley²¹ — and other dilemmas with seemingly similar structures. For example, may a transplant surgeon sacrifice one healthy patient, so as to harvest the organs needed to save five others? If not, then why is turning the trolley permissible?

Such conflicting intuitive judgments are taken by trolley problem enthusiasts as a sort of moral raw data.²² The contrast between what is so plainly right in one case and so plainly wrong in another forces us to articulate reasons — moral principles — to fit and justify how our intuitions shift so dramatically from case to case.²³ From such fertile

¹⁹ Indeed, one might readily see the newly released federal guidelines as an advance signal to the public of the moral acceptability of self-driving cars — given that 35,000 people died last year in auto accidents, 94 percent of which resulted from human error. See, e.g., Cecilia Kang, *Self-Driving Cars Gain Powerful Ally: The Government*, N.Y. TIMES (Sept. 19, 2016), <http://www.nytimes.com/2016/09/20/technology/self-driving-cars-guidelines.html> [https://perma.cc/98ZE-QK6K]; Barack Obama, *Self-Driving, Yes, but Also Safe*, PITTSBURGH POST-GAZETTE (Sept. 19, 2016, 8:00 PM), <http://www.post-gazette.com/opinion/Op-Ed/2016/09/19/Barack-Obama-Self-driving-yes-but-also-safe/stories/201609200027> [https://perma.cc/N7LG-APUR].

²⁰ See Judith Jarvis Thomson, *Killing, Letting Die, and the Trolley Problem*, 59 MONIST 204 (1976); Judith Jarvis Thomson, Comment, *The Trolley Problem*, 94 YALE L.J. 1395 (1985) [hereinafter Thomson, *The Trolley Problem*]. Thomson's articles respond to Professor Philippa Foot, who introduced one version of the trolley (or tram) scenario in making an argument about abortion to save the life of the mother. Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, 5 OXFORD REV. 5 (1967), reprinted in PHILIPPA FOOT, VIRTUES AND VICES AND OTHER ESSAYS IN MORAL PHILOSOPHY 19, 23–24 (2002).

²¹ This version is often among the first thought experiments that students of law, ethics, moral philosophy, or moral psychology encounter. See, e.g., Michael Sandel, *Justice: What's the Right Thing to Do? Episode 01 "The Moral Side of Murder,"* YOUTUBE (Sept. 4, 2009), <https://www.youtube.com/watch?v=kBdfcR-8hEY> [https://perma.cc/RB23-VJK7] (opening lines in the first day of Professor Michael Sandel's moral reasoning class for college students, called "Justice"); see also BBC Radio 4, *The Trolley Problem*, YOUTUBE (Nov. 18, 2014), <https://www.youtube.com/watch?v=bOpf6KcWYyw> [https://perma.cc/2CEZ-E9B5] (video primer introducing philosophical debates about the trolley problem).

²² As Kamm once explained, "people who have responses to cases are a natural source of data from which we can isolate the reasons and principles underlying their responses." F.M. KAMM, INTRICATE ETHICS 8 n.4 (2007). "The idea [is] that the responses come from and reveal some underlying psychologically real structure . . ." *Id.* For a concise exposition of a skeptical view that such raw data have normative value (for determining what our moral principles should be), as opposed to having only descriptive value (for understanding our moral psychology), see Peter Singer, *Ethics and Intuitions*, 9 J. ETHICS 331 (2005).

²³ As Rakowski neatly explains the standard method: "The foundation of the trolley problem is a pair of firm beliefs about moral duties or permissions in particular circumstances. The trolley

juxtapositions have sprung dense forests of argument about distinctions that should (or should not) matter.²⁴ Some of these distinctions would sound familiar to any lawyer; for example, act versus omission, positive versus negative duties, and intended harms versus foreseen side effects.

One might have guessed that a generation's worth of philosophical debate ought to have settled at least the most basic principles by now. But true to its name, *The Trolley Problem Mysteries* proves the opposite. If anything, the most fundamental questions may be even more unsettled now than ever before.

A. A Surprising Turn

At the core of this book is a fierce, fundamental debate between Kamm and Thomson. The controversy arises because Thomson has recently reversed her position on the most central trolley case,²⁵ the benchmark against which all variations are measured: whether a bystander standing by the switch is morally permitted to change the trolley's course, sacrificing one to save five.²⁶ The standard view, which she herself argued when she invented this bystander variation, says "yes."²⁷

But Thomson now says "no," based on new reasoning that I will soon relate.²⁸ Let's first pause, though, to take note of what is at stake. If her new position prevails, it could upend much of the existing trolley problem literature.²⁹ Recall how the canonical contrasts of the

driver *should*, or *may*, turn the trolley from five to one. The doctor *may not* remove parts of one person's body to save five other people. What moral principle or set of principles best accounts for these beliefs? We want the principles we endorse and our reactions to cases to cohere" (p. 5). In typical trolley-style reasoning, if the judgment in an individual case turns out to be an outlier that cannot be reconciled with proposed principles that neatly fit and justify many other cases, then the analyst may consider changing her judgment in the outlier case; how readily she should revise individual case judgments, however, turns out to be a point of dispute between Kamm and her commentators. See *infra* note 51 and accompanying text.

²⁴ Meanwhile, research in experimental philosophy and moral psychology has also investigated whether such distinctions do (or do not) matter, in fact, to people's moral intuitions. See *infra* notes 54–55 and accompanying text.

²⁵ See Judith Jarvis Thomson, *Turning the Trolley*, 36 PHIL. & PUB. AFF. 359, 366 (2008).

²⁶ This is the same scenario that opened this Review; it is also the centerpiece of the experiments reported below. Throughout this Review, as in the book's core debates, it is assumed that the people whose lives are at stake are generic or equal in any possible morally relevant sense. In other words, none of them is the bystander's mother, and none of them is any more likely to cure cancer someday than any other.

²⁷ See Thomson, *supra* note 20, at 1397 ("Of course you will kill one if you [turn the trolley]. But I should think you may turn it all the same.").

²⁸ See Thomson, *supra* note 25, at 366.

²⁹ As Kamm puts it: "The most striking claim offered by Thomson in her 2008 article is that she now believes that it is not permissible for a mere bystander to redirect the trolley. Indeed, in that article Thomson reversed herself, claiming that it is not any more permissible for the bystander to redirect the trolley toward the one than to topple the fat man to stop the trolley; and

trolley problem rely on an emphatic “yes” (that turning the trolley is morally permissible) as the answer to the original case. That answer is what creates the tension with other classic scenarios in which the answer is just as emphatically “no”: pushing someone in front of the trolley, or harvesting organs from a healthy patient. But if turning the trolley is also deemed *not* to be permissible, what is left to explain?³⁰ All these judgments would be aligned, and no fancy constructs would be needed to reconcile them.

It is not much exaggeration, then, for Kamm to wryly observe that Thomson’s new position equates to saying “there is no trolley problem” (p. 21).³¹ But Kamm refuses to acquiesce in such an inglorious demise. Her hope is to “resurrect the trolley problem” (p. 14). As her book title indicates, she continues to believe that this thought experiment serves up mysteries to be resolved. Accordingly, she spends much of the book, including most of her first lecture and a good portion of her second, defending the conventional assumption against Thomson’s recantation.

In brief, Thomson’s new argument goes like this. Imagine that our beleaguered bystander now faces a moral dilemma with *three* dreadful options: he can let the trolley kill five workmen, turn the trolley to kill one other workman, or turn the trolley to kill *himself*.³² He can’t be morally required to sacrifice himself, Thomson posits.³³ But if that is true, then he also should not be permitted to sacrifice someone else. This is “because neither he nor [the lone workman] is required to pay the cost of saving the five, and therefore if he wants the cost paid, he must pay it himself” (p. 117). But then, what about the original two-options case, in which bystander *cannot* sacrifice himself? “All the same,” Thomson argues, “he may not throw the switch so that it kills Workman because neither he nor Workman is required to pay the cost of saving the five, and therefore, since he can’t pay the cost himself (whether or not he would like to), he must let the five die” (p. 117).

therefore, she, a founding mother of the trolley problem, believes there is no trolley problem . . .” (p. 21).

³⁰ Thus Kagan sees Thomson as making the case for “abandoning the very intuition that got us started thinking about the trolley problem in the first place!” (p. 164).

³¹ Thomson also characterizes her change of position as implying that “there isn’t really any such thing as the [bystander version of the] trolley problem” (p. 117). Not everyone would lament the disappearance of the trolley problem, it might be noted. See Barbara H. Fried, *What Does Matter? The Case for Killing the Trolley Problem (or Letting It Die)*, 62 PHIL. Q. 505, 506 (2012) (arguing that typical trolley-like thought experiments are limited in real-world relevance because they do not cover cases of “uncertain risk of accidental harm to generally unidentified others” created by “prima facie permissible” activities, such as driving cars, but rather focus on tradeoffs of one harm for another under hypothetical conditions of certainty about choices, causation, and victim identities).

³² See Thomson, *supra* note 25, at 364.

³³ *Id.* at 365.

The path of this argument requires several steps, and Kamm suspects that they are too loosely connected. She urges the reader to mind the gaps in Thomson's logic, and to be wary of leaping across them blithely. What makes it safe, for instance, to draw conclusions about the standard two-options case based on the three-options case?³⁴ Why is it relevant whether the bystander would be willing to sacrifice himself if he could, in a case where he simply cannot?³⁵ And why should it matter if the potential victim would in theory be required to sacrifice himself if *he* could?³⁶

But that is only one piece of Kamm's strategy for preserving the mysteries of the trolley problem. She has titled her first lecture "Who Turned the Trolley?,"³⁷ and as her strategy unfolds, we see why she is pressing this inquiry into "*who* done it" (p. 14). First, she reminds us that the original trolley problem, as proposed by Professor Philippa Foot, concerned whether the *driver* of the trolley may turn it (pp. 11–12). If we view the driver's choice as between killing one or killing five, then it seems the answer must be "yes." And if it remains an easy call that the driver may turn the trolley (never mind what Thomson now says about the bystander's choice), then certain puzzling conflicts do survive. For instance: If the driver may turn the trolley, to kill one instead of killing five, then why do we think he is barred from stopping the trolley by hitting a switch that topples someone in its path? (pp. 21–22).

Kamm further reminds us why Thomson invented the bystander variation in the first place. Unlike the driver, if a mere bystander does nothing, it seems hard to say that he has *killed* the five; rather, he has let them die. By saying that he may turn the trolley, then, we are say-

³⁴ As Kamm pointedly observes: "For example, if we have three choices, among letting a murder occur or stopping it by killing the murderer or stopping it by shooting him in the leg, it is impermissible to kill him. However, this does not mean it is impermissible to kill him if that is one's only alternative to letting the murder occur" (p. 23). Thomson's reply suggests that she does not end up resting much of her argument on such a move: "I supplied the three-options case primarily in order to bring out more vividly the consideration that makes it impermissible for the bystander to kill Workman in the two-options case — namely, that neither the bystander nor Workman is required to pay the cost of saving the five" (p. 118). Kamm is "surprised" by this reply, as the three-options case then seems to her to do very little work (p. 175); she reminds us that Thomson's three-options case also played a further role, in serving up the intuition that "if one can oneself pay a cost to do a good deed, other things equal, one may not make someone else pay it" — a principle that Kamm finds ungrounded (p. 175).

³⁵ Kamm observes, more generally: "It may, of course, be impermissible to impose costs on others without their consent, but I do not think, contrary to Thomson, that a good argument for showing this is that one will not or would not oneself *volunteer* to pay the costs" (p. 28).

³⁶ As Kamm puts it, "taking someone's life without his consent involves imposing a cost on another, and this is not the same as demanding that the other person impose the cost on himself altruistically when the bystander will not or would not impose the cost on himself altruistically" (p. 28).

³⁷ Cf. Thomson, *supra* note 25 (entitled *Turning the Trolley*).

ing that it is morally permissible to kill one in order to avoid letting five die. Maybe so, but this undermines what would have been an easy story for why we think a transplant surgeon may *not* harvest a healthy patient's organs to save five others — that killing is worse than letting someone die.³⁸ Thus Thomson's bystander case came to stand for a bedrock analytical point: that our intuitive judgments cannot be explained simply by giving more moral weight to a negative duty (not to kill) than to a positive duty (to save, or not let die).³⁹

Yet now Thomson has reversed course, saying that the bystander may not turn the trolley after all. Meanwhile, she also concludes that the driver not only may, but must, turn the trolley.⁴⁰ To justify this stark contrast between the bystander and the driver, she fully endorses the negative/positive duties distinction she herself so famously deflated forty years earlier.⁴¹

In response, Kamm offers a challenging invention of her own: suppose that the bystander is actually the driver, who was accidentally thrown from the trolley but happened to land next to the switch (p. 19).⁴² This "bystanding driver" variation, which blurs the two roles, serves as a sort of stress test for the negative/positive duties distinction. Kamm thinks the bystanding driver *is* permitted to turn the trolley — just as the bystander is, and just as the original driver is.⁴³ But for Thomson, the question is a trickier one: given the sharp divide she draws between the bystander (who may not kill one, rather than let five die) and the driver (who must kill one, rather than kill five), how would she deal with this hybrid case of the bystanding driver?⁴⁴

³⁸ Likewise, what about the bystander who pushes someone in front of the trolley to stop it — isn't he also killing one to avoid letting five die?

³⁹ For convenience of exposition, I will follow the book's lead and refer to the distinction as one between negative and positive duties (or between killing and letting die), setting aside related distinctions such as doing/allowing or act/omission.

⁴⁰ Thomson, *supra* note 25, at 372.

⁴¹ Thomson grappled with this contrast in detail in her 2008 article, explaining there that "[t]his difference between [the driver] and the bystander is obviously due to the fact that whereas [the driver] kills five if he does nothing, the bystander instead lets five die." *Id.*

⁴² Kamm labels this hybrid case as one of "killer let die," to indicate that although one might say that the driver would become in some sense the killer of the five if he did not turn the trolley, at the moment of decision he is choosing only whether to let them die (p. 20). This is in contrast to both the pure bystander case (a "let die" case) and the original driver case (a "kill" case).

⁴³ Kamm does allow for one possible difference between the driver and the bystander, in that the driver might be *required* to turn the trolley (as Thomson believes) (p. 16).

⁴⁴ Thomson does not take up the invitation to address this hybrid in her chapter but it is possible to imagine how each of the possible answers might leave her with some explaining to do. If Thomson is to say that the bystanding driver must turn the trolley, then she would have to either let up on the kill/let die distinction, or else argue that the bystanding driver would not merely be letting the five die in the same sense as the pure bystander (as Kamm posits), but would be killing them in some relevant sense — perhaps necessitating a new moral category for such "killer let die" cases. Moreover, this answer would also create a contrast to be reconciled, because Kamm and Thomson agree that sacrificing the one would be prohibited in another "killer let die" case — the

Thus, in essence, Kamm's first lecture locates some of the more intriguing trouble spots — where philosophical mysteries remain — both for those who believe the bystander may turn the trolley, and for the contrarians who do not. In that spirit, one might even suggest a way in which Thomson's reversal creates a *more* puzzling moral lineup than before: now the mere bystander is prohibited, the driver is permitted (in fact, required), and the transplant surgeon is prohibited; and yet, isn't the driver serving a special role more like the surgeon's — with certain responsibilities for the well-being of others — than the mere bystander is?⁴⁵

In her second lecture, it is Kamm who goes rogue. Rejecting the conventional view that an actor's intentions are a centrally relevant moral factor,⁴⁶ Kamm has for some time been developing a Principle of Permissible Harm that focuses instead on the causal relation between the act and its good and bad results (pp. 66–80).⁴⁷ In this book, she articulates the latest iteration of this theory, which receives its own share of challenges from the panel of commentators.

Under Kamm's theory, an act that kills one to save five is prohibited if it has a closer causal relation to the killing than to the saving — for example, the act of pushing a person into the trolley's path kills him but is only a causal means to saving the five, and would thus be impermissible. By contrast, the act is permissible if the killing is a causal effect of (what is essentially) the saving itself — for example, the act of diverting the trolley kills the one person as a causal effect of the trolley threat moving away from the five, and would thus be per-

case of the “culpable doctor,” in which the transplant surgeon must decide whether to harvest organs from a healthy patient when this surgeon herself caused the organ failures in the other five patients in the first place (p. 19). And if Thomson is to say instead that the bystanding driver may not turn the trolley, then she faces a line-drawing problem as to when the driver no longer counts as a driver — as suggested by a further “killer let die” scenario that Kamm has at the ready: What about the case of a driver who faints, thus giving up control of the trolley, but then awakens just in time to choose whether to let the trolley continue forward to hit the five, or to instead turn it toward the one (p. 18)? Finally, if Thomson is to say that the bystanding driver may turn the trolley (but need not), then she must explain why this actor differs from both the pure bystander (who may not turn) and the driver in the usual case (who must turn) — again, maybe implying a new moral category for such “killer let die” cases.

⁴⁵ The question of role-based morality will return shortly; it also motivates the “engineer” variation of the trolley scenario in the second experiment reported below.

⁴⁶ For example, Kamm rejects the “doctrine of double effect,” which would say that stopping the trolley by pushing someone into it is forbidden, because that option requires that one intend the man's death, whereas diverting the trolley is permissible because then the man's death is not intended but merely a foreseen side effect of saving the five (pp. 14–15).

⁴⁷ That is, under her theory, different acts (that cause both a greater good and a lesser evil) can be either morally permitted or prohibited — even if the actor has exactly the same intentions and achieves the same outcomes — depending on how those acts causally relate to the outcomes. An earlier development of the Principle of Permissible Harm is found in KAMM, *supra* note 22, at 24–25.

missible. (This is all a rough paraphrase of her principle; it is hard to express concisely.⁴⁸)

This might seem a less-than-intuitive way to draw sharp moral distinctions, and indeed all three commentators are somewhat mystified by Kamm's approach.⁴⁹ At various points they gamely challenge her on the metaphysics of cause and effect,⁵⁰ but it quickly becomes apparent that they also just disagree with her intuitions — both about her proposed principle and about the specific cases she has considered in deriving it.

Kagan, for example, complains that although Kamm may have found a theory that reconciles her own intuitions about a range of trolley cases, the theory itself lacks intuitive appeal (p. 162). And Hurka ventures that one reason may be Kamm's tendency to "reject an attractive-sounding principle because it conflicts with just one or a few [of her] particular judgments" about specific imaginary scenarios⁵¹ — some of which are "so far from reality" that one's intuitions about them (if any) should not be taken quite so seriously (p. 139). Pressing

⁴⁸ Kamm's own "rough description" in her lectures is as follows: "Actions are *permissible* if greater good or a component of it (or means having these as a noncausal flip side) leads to lesser harm even directly. Actions are *impermissible* if mere means that produce greater good . . . cause lesser harm at least directly, and actions are *impermissible* if mere means cause lesser harms (such as toppling people in front of a trolley) that are mere means to producing greater goods" (p. 66). Rakowski, Thomson, Hurka, and Kagan also each offer unique variations on a paraphrase (pp. 4, 122, 137, 156). My paraphrase draws on theirs, as well as on my conversations with Professor Kamm, for which I am grateful.

⁴⁹ There is much sport for philosophers, the book reminds us, in conjuring up seemingly absurd implications of a given theory. For example, Hurka urges, let's consider the case of collateral damage in war: Why should it matter morally if the debris that kills an innocent civilian is a piece of the bomb that blew up an enemy munitions factory, or a piece of the building itself? (If it's a piece of the bomb, the killing would be prohibited under Kamm's theory because the bomb's explosion is a causal means of destroying the factory. But if it's a piece of the factory, the killing would be permitted because then the death is a causal effect of the factory's destruction, which is assumed to be the greater good at stake (pp. 137–39).) Despite this challenge, Kamm remains convinced that the distinction matters — at least, in a trolley version in which a bomb can be thrown to stop the trolley (doing so would be permissible, under her theory, if a piece of the blown-up trolley flies off and kills someone, but impermissible if the bomb itself kills that person) (p. 222).

⁵⁰ Thomson briefly raises the question (p. 122), as does Kagan (p. 157). Hurka devotes more of his response to it (pp. 137–47). See *supra* note 49.

⁵¹ One can sense this tendency in how Kamm herself explains her method: "Consider as many case-based judgments of yours as prove necessary. Do not ignore some case-based judgments, assuming they are errors, just because they conflict with simple or intuitively plausible principles that account for some subset of your case-based judgments. Work on the assumption that a different principle can account for all of the judgments. Be prepared to be surprised at what this principle is. . . . Then, consider the principle on its own, to see if it expresses some plausible value or conception of the person or relations between persons." KAMM, *supra* note 22, at 5. A counterexample of sorts, however, might be found in Kamm's insistence on accepting the results that her principle dictates in the bomb hypotheticals, *supra* note 49, despite the apparently counterintuitive nature of those particular judgments.

the point, Kagan reports that based on a “very informal survey of students in my upper level normative ethics course,” half the students disagreed with Kamm’s intuitions about one case she used to develop her theory, and three-quarters disagreed with her about another such case (p. 162). And it “does seem problematic for Kamm if others fail to share her intuitions,” he observes, for she grounds her arguments in “assertions about what . . . ‘people would think’ or what ‘we think,’ and she talks about ‘our intuitive judgments’ as well” (p. 162).⁵²

And yet, not much hard evidence is mentioned in these exchanges about what “our” moral intuitions actually say, how fluid or rigid they are, or what might cause them to change.⁵³ This is not for lack of data, as a growing body of research in experimental philosophy and in moral psychology has steadily collected precisely this sort of evidence.⁵⁴ Notably, among the topics that have gained research attention is one that intersects with Kamm’s theory: how our minds assess the strength of the causal connection between an act and an outcome — including how such an assessment may be influenced by our prior moral judgments (and not only the other way around).⁵⁵

⁵² As she has previously explained in an interview, “My approach is generally to stick with our common moral judgements, which I share and take seriously.” ALEX VOORHOEVE, CONVERSATIONS ON ETHICS 20 (2009) (interviewing Kamm). She makes clear, however, that in her view, “[t]hat a lot of people agree doesn’t show that something is correct.” *Id.* at 28.

⁵³ Two exceptions are Kagan’s informal report about his students’ intuitions (p. 162), as noted above, and Rakowski’s pointer in the book’s introduction (p. 6 n.3) to empirical research by Professors Mark Kelman and Tamar Kreps, described more fully below.

⁵⁴ For a review of the growing experimental philosophy and moral psychology literature by several of its leading researchers, see Fiery Cushman & Liane Young, *The Psychology of Dilemmas and the Philosophy of Morality*, 12 ETHICAL THEORY & MORAL PRAC. 9 (2009), identifying “several instances where cognitive research has identified distinct psychological mechanisms for moral judgment that yield conflicting answers to moral dilemmas,” *id.* at 9; and Joshua D. Greene, *The Cognitive Neuroscience of Moral Judgment and Decision Making*, in THE COGNITIVE NEUROSCIENCES 1013 (Michael S. Gazzaniga & George R. Mangun eds., 5th ed. 2014), surveying recent research, including work with functional MRI brain scans showing that for some variations of the trolley problem, the brain’s emotional centers are more active, while for other versions, the brain’s cognitive centers are more active, *id.* at 1016.

Part of this literature, notably, is directed at specific arguments made in the analytical literature. See, e.g., S. Matthew Liao, Alex Wiegmann, Joshua Alexander & Gerard Vong, *Putting the Trolley in Order: Experimental Philosophy and the Loop Case*, 25 PHIL. PSYCHOL. 661, 666–68 (2012) (showing order effects in moral judgments about Thomson’s “loop” case — a famous variation that blurs the distinction between the pushing and turning cases — depending on whether the loop case is presented after a pushing case or a turning case); see also Ezio Di Nucci, *Self-Sacrifice and the Trolley Problem*, 26 PHIL. PSYCHOL. 662 (2013).

⁵⁵ See, e.g., Fiery Cushman, Joshua Knobe & Walter Sinnott-Armstrong, *Moral Appraisals Affect Doing/Allowing Judgments*, 108 COGNITION 281, 288 (2008) (reporting experiments showing “that people’s moral appraisals affect their application of the doing/allowing distinction”); Fiery Cushman & Liane Young, *Patterns of Moral Judgment Derive from Nonmoral Psychological Representations*, 35 COGNITIVE SCI. 1052, 1056 (2011) (providing experimental evidence suggesting that the act/omission distinction primarily affects moral judgment via causal attribution, but also that “the patterns of causal and intentional attribution observed in the moral condition were, in

Drawing on existing research would seem doubly useful in these debates, moreover, for it may be relevant not only to Kamm's theory but also to her case against Thomson's. Indeed, when Thomson reversed course, she thought it necessary to explain (away) the broad appeal of the view that she was recanting.⁵⁶ That apparent consensus,⁵⁷ after all, might account for why her original view was taken for granted for so long as a philosophical fixed point.

Offering an exception that proves the rule, Rakowski's introduction points to recent survey experiments by Professors Mark Kelman and Tamar Kreps,⁵⁸ showing that survey subjects are less likely to deem turning the trolley to be permissible when that case is presented alongside the contrasting case of pushing someone in front of the trolley (p.

part, a *consequence* of subjects' moral judgments," *id.* at 1062); Joshua Knobe, *Person as Scientist, Person as Moralist*, 33 BEHAV. & BRAIN SCI. 315, 315 (2010) (noting that "in recent years, a series of experimental results" have begun to show that "people's judgments about whether a given action truly is morally good or bad can actually affect their intuitions about what that action caused and what mental states the agent had"); Janice Nadler & Mary-Hunter McDonnell, *Moral Character, Motive, and the Psychology of Blame*, 97 CORNELL L. REV. 255, 292 (2012) (providing experimental evidence suggesting that "moral character, inferred independently from an actor's motive for creating harm (or the conditions that led to harm) can influence judgments about causation, intent, and blame"); Barbara A. Spellman & Elizabeth A. Gilbert, *Blame, Cause, and Counterfactuals: The Inextricable Link*, 25 PSYCHOL. INQUIRY 245, 245 (2014) (arguing that "causal attribution depends at least in part on counterfactual reasoning and that downstream assessments of intentionality, reasons, obligations, and capacity affect the invocation and content of the counterfactuals that will be used in making the causal judgment").

Extrapolating from such findings, for instance, one might hypothesize that if our moral intuitions are already predisposed more favorably (for whatever reason) to turning the trolley toward one person than to pushing one person into its path (or predisposed more favorably to someone who would do the former than someone who would do the latter), then we might also intuit a weaker intentionality and a weaker causal connection between turning and death than between pushing and death. Or one might hypothesize that, if we are already thus predisposed, we will tend to subconsciously downplay the kill/let die distinction in the turning case (thereby also weakening the sense of causal connectedness between turning and death in that case).

⁵⁶ As she put it, even if her new view means that the "trolley problem is therefore in one way a nonproblem, it is therefore in another way a real problem, for if the bystander must not turn the trolley in [the bystander case], then we need to ask why so many people who are presented with that case think it obvious that he may." Thomson, *supra* note 25, at 368. She suspects that the explanation for how our intuitions vary involves "how drastic an assault on the one the agent has to make in order to bring about, thereby, that the five live"; and in particular, that "[t]he more drastic the means, the more strikingly abhorrent the agent's proceeding . . . [and] the more striking it is that the agent who proceeds infringes a negative duty to the one." *Id.* at 374; cf. Joshua D. Greene, Fiery A. Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom & Jonathan D. Cohen, *Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment*, 111 COGNITION 364 (2009).

⁵⁷ See Cushman & Young, *supra* note 54, at 11 ("Numerous studies have demonstrated that a large majority of individuals consider it morally acceptable to use a switch to redirect a runaway trolley away from five victims and onto a single victim, but unacceptable to push a single victim in front of a runaway trolley in order to stop its progress towards five victims.").

⁵⁸ Mark Kelman & Tamar Admati Kreps, *Playing with Trolleys: Intuitions About the Permissibility of Aggregation*, 11 J. EMPIRICAL LEGAL STUD. 197 (2014).

6 n.3).⁵⁹ For one thing, this suggests that the standard intuition may in fact be somewhat malleable. Moreover, isn't this sort of gravitational pull from a contrasting case just what Thomson's argument invites?⁶⁰

In fact, a more recent study shows that people do more often say that turning the trolley is prohibited (in the standard case) if they are first presented with a version of Thomson's three-options case.⁶¹ (By contrast, Kelman and Kreps show that directly exposing subjects to the actual argument — "that no one is obliged to sacrifice his own life to save others, and that it seems immoral to force another to make a sacrifice one would not have to make oneself" — seems to make much less of a difference to their moral intuitions.⁶²) Isn't such an elision of intuitions across the two scenarios, disguised as the persuasive effect of an argument, just what Kamm is warning us to guard against?⁶³

Although "philosophical arguments . . . are not reports on opinion polls," as Rakowski notes, empirical research may nonetheless inform "what each of us should conclude about the tug of intuitions or the robustness of tentatively held principles" (p. 6 n.3). Such experiments

⁵⁹ Even moral philosophers as a class are not immune to such framing effects in judging trolley cases, it turns out. See Eric Schwitzgebel & Fiery Cushman, *Philosophers' Biased Judgments Persist Despite Training, Expertise and Reflection*, 141 COGNITION 127, 131–36 (2015) (confirming authors' earlier work documenting order effects among philosophers in the canonical scenarios, and further showing that such effects persist despite various means for debiasing). It remains possible, of course, that particular individuals would not be susceptible.

⁶⁰ As Rakowski notes, "Thomson's thinking has followed [the] same arc" as those who "abandon [the standard] view to achieve what they regard as a greater consonance with their response to the [contrasting] case" (p. 6 n.3). Recall that Thomson's argument looks first to the three-options case, in which it is intuitively attractive to say "you cannot sacrifice someone else (if you can but won't sacrifice yourself)," before turning to the standard two-options case, where a sense of consistency may lend a boost to the intuition that "you cannot sacrifice someone else (period)." See *supra* notes 32–33 and accompanying text. (These are my paraphrases.)

⁶¹ Nucci, *supra* note 54, at 668. Although the Nucci study aimed to mirror Thomson's argument, the question it asked subjects was not whether they deemed turning the trolley to be morally permissible, but rather the question of "What should you do?" *Id.* at 666–67.

⁶² Kelman & Kreps, *supra* note 58, at 217. It must be cautioned that the numbers from the Nucci and the Kelman & Kreps studies are not directly comparable, given differences in subject populations, scenario setup, and measured outcomes. Yet the shift in the Kelman & Kreps experiment (a fall from 77 percent to 67 percent saying that turning the trolley is permissible, *id.*) seems to be quite muted relative to the shift in the Nucci experiment (a fall from 67 percent to 39 percent saying they should turn the trolley). Nucci, *supra* note 54, at 668. It should also be noted that both the Kelman & Kreps and Nucci studies, as well as some of the studies cited above, were published after Kamm's Tanner Lectures and the panelists' commentaries took place, and it may well be that had Kamm and her commentators had access to this research, they might have made use of it.

⁶³ Kamm has expressly raised such a concern, as it applies to philosophers, in more general terms: "[T]he fact that philosophers often do not respond to one case without thinking of another, and especially the possibility that their considering several cases together might yield different intuitive judgments than considering each case in isolation, may threaten an assumption about the purity of a philosopher's intuitions. . . . [Such intuitions] may be impure because they are the result of a coherentist frame of mind This is a problem worth thinking about." KAMM, *supra* note 22, at 427.

may be just as revealing, as these new studies suggest, in interrogating how various forms of intuition-driven arguments may — and whether they should — change our minds.

B. New Directions

This book, with all its sophisticated frisson, could hardly have been better timed for publication. The trolley problem has enjoyed a resurgence in the public consciousness. It has not only thrived in academic debates about law and policy,⁶⁴ but it has also become the subject of a popular press book by a bestselling author,⁶⁵ of a BBC radio play by a renowned playwright,⁶⁶ and of renewed attention in other media.⁶⁷ Public commentators have invoked it in debates about Obamacare, drone strikes, and even the recent Iran nuclear deal.⁶⁸

⁶⁴ See *supra* notes 3–4 and accompanying text.

⁶⁵ DAVID EDMONDS, *WOULD YOU KILL THE FAT MAN? THE TROLLEY PROBLEM AND WHAT YOUR ANSWER TELLS US ABOUT RIGHT AND WRONG* (2014). The first chapter of the book, preprinted in *Foreign Affairs*, is available at <https://www.foreignaffairs.com/articles/2013-12-18/would-you-kill-fat-man> [<https://perma.cc/5R5Z-FCPL>]. See also Sarah Bakewell, *Clang Went the Trolley: “Would You Kill the Fat Man?” and “The Trolley Problem,”* N.Y. TIMES (Nov. 22, 2013), <http://www.nytimes.com/2013/11/24/books/review/would-you-kill-the-fat-man-and-the-trolley-problem.html> [<https://perma.cc/2WC6-9FVE>] (reviewing the book).

⁶⁶ Tom Stoppard, *Darkside* (BBC Radio 2 2013) (radio play commissioned to celebrate the 40th anniversary of Pink Floyd’s album, *Dark Side of the Moon*); Larry Rohter, *An Author Dives into Pink Floyd*, N.Y. TIMES (Nov. 25, 2013), <http://www.nytimes.com/2013/11/26/theater/tom-stoppard-gives-the-dark-side-of-the-moon-a-makeover.html> [<https://perma.cc/77BA-XXM6>] (noting that “the subject matter of ‘Darkside’ will be familiar to anyone who has ever taken a college course in ethics or moral philosophy”).

⁶⁷ See, e.g., Elisa Criado, *Would You Push a Stranger Off a Bridge? How Your Morals Depend on Language*, INDEPENDENT (Apr. 29, 2014), <http://www.independent.co.uk/news/science/to-push-or-not-to-push-how-your-morals-depend-on-language-9303510.html> [<https://perma.cc/P2VN-LR3K>]; Lauren Cassani Davis, *Would You Pull the Trolley Switch? Does It Matter?*, THE ATLANTIC (Oct. 9, 2015), <http://www.theatlantic.com/technology/archive/2015/10/trolley-problem-history-psychology-morality-driverless-cars/409732> [<https://perma.cc/P7WA-VYC8>]; Amy Dockser Marcus, *How New Technology Is Illuminating a Classic Ethical Dilemma*, WALL STREET J. (Jun. 8, 2016, 10:11 AM), <http://www.wsj.com/articles/how-new-technology-is-illuminating-a-classic-ethical-dilemma-1465395082> [<https://perma.cc/W5JV-V86Z>]; cf. Poncie Rutsch, *Men and Women Use Different Scales to Weigh Moral Dilemmas*, NPR (Apr. 3, 2015, 2:52 PM), <http://www.npr.org/sections/health-shots/2015/04/03/397280759/men-and-women-use-different-scales-to-weigh-moral-dilemmas> [<https://perma.cc/LQ5F-F4WJ>] (concerning moral dilemmas more generally). Apparently, it has also become the subject of an internet meme. See Linch Zhang, *Behind the Absurd Popularity of Trolley Problem Memes*, HUFFINGTON POST (Jun. 1, 2016, 7:15 PM), http://www.huffingtonpost.com/linch-zhang/behind-the-absurd-popular_b_10247650.html [<https://perma.cc/GJ5D-TNDF>].

⁶⁸ See Hussein Banai, *How Obama Solved U.S.-Iran Relations’ “Trolley Problem,”* ETHICS & INT’L AFF. (July 16, 2015), <http://www.ethicsandinternationalaffairs.org/2015/how-obama-solved-u-s-iran-relations-trolley-problem> [<https://perma.cc/V3XE-FBXD>]; Jonathan Chait, *Yes, the Republican Obamacare Strategy Will Kill People*, N.Y. MAG. (Nov. 14, 2014, 12:17 PM), <http://nymag.com/daily/intelligencer/2014/11/yes-republican-obamacare-plan-will-kill-people.html> [<https://perma.cc/3M8N-PKNT>]; Larissa MacFarquhar, Forum, *The Logic of Effective Altruism*, BOS. REV. (July 1, 2015), <http://bostonreview.net/forum/logic-effective-altruism/larissa>

Most vividly, it has become the touchstone for the question of how to program ethics into self-driving cars⁶⁹: In an emergency, should your Tesla be programmed to swerve to hit one pedestrian, in order to save five? Should it sacrifice *your* life, to save those five?⁷⁰ And what if the situation allowed for not two, but three options: sacrificing you, the passenger, by swerving left; hitting someone else by swerving right; or continuing straight to hit the five?⁷¹

Amidst such popularization and the attendant risk that nuanced insights might be lost, it is vital to have in hand a book that “showcases some of the best of current thinking, by the leading voices in the field” (p. 6) — even if what the book reveals is a crescendo of discord among these voices. All the more so, perhaps, as some of the fantastical figures inhabiting the philosophical mind now seem to be coming to life. Consider: Are the human backup drivers in Uber’s new self-driving cars more like the trolley driver, or more like the bystander — or more like Kamm’s creature, the “bystanding driver”?⁷²

-macfarquhar-response-effective-altruism [https://perma.cc/ZEN3-DB82] (discussing drone strikes).

⁶⁹ See, e.g., Joel Achenbach, *Driverless Cars Are Colliding with the Creepy Trolley Problem*, WASH. POST (Dec. 29, 2015), <http://www.washingtonpost.com/news/innovations/wp/2015/12/29/will-self-driving-cars-ever-solve-the-famous-and-creepy-trolley-problem> [https://perma.cc/4J69-MRWQ]; Robin Marantz Henig, *Death by Robot*, N.Y. TIMES MAG. (Jan. 9, 2015), <http://www.nytimes.com/2015/01/11/magazine/death-by-robot.html> [https://perma.cc/MA6N-X7N8] (discussing autonomous military drones and self-driving cars); Patrick Lin, *The Ethics of Autonomous Cars*, THE ATLANTIC (Oct. 8, 2013), <http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360> [https://perma.cc/UJC8-NNYC]; Gary Marcus, *Moral Machines*, NEW YORKER (Nov. 24, 2012), <http://www.newyorker.com/news/news-desk/moral-machines> [https://perma.cc/ZWV4-32YM]; Levi Tillemann & Colin McCormick, *Will Driverless Car Makers Learn to Share?*, NEW YORKER (Sept. 25, 2016), <http://www.newyorker.com/business/currency/will-driverless-car-makers-learn-to-share> [https://perma.cc/3C7X-HLPA]; Tim Worstall, *When Should Your Driverless Car from Google Be Allowed to Kill You?*, FORBES (June 18, 2014, 8:27 AM), <http://www.forbes.com/sites/timworstall/2014/06/18/when-should-your-driverless-car-from-google-be-allowed-to-kill-you>.

⁷⁰ See, e.g., Jean-François Bonnefon, Azim Shariff & Iyad Rahwan, *The Social Dilemma of Autonomous Vehicles*, 352 SCIENCE 1573 (2016); Julia Griffin, *People Want Self-Driving Cars to Value Passenger Safety over Pedestrians, Study Says*, PBS: NEWS HOUR (June 23, 2016, 2:51 PM), <http://www.pbs.org/newshour/rundown/people-want-self-driving-cars-to-value-passenger-safety-over-pedestrians-study-says> [https://perma.cc/7K85-VH6N].

⁷¹ In her 2008 article, Thomson analyzes this scenario: a driver “suddenly sees five people on the street ahead of him, but his brakes fail: he cannot stop his car, he can only continue onto the street ahead or steer to the right (killing one) or steer to the left (killing himself).” Thomson, *supra* note 25, at 369. Emphasizing that “if he simply takes his hands off the wheel, he runs the five down and kills them,” she asserts that this driver “cannot at all plausibly insist that he merely lets them die.” *Id.* Indeed, it was this assertion that prompted Kamm to respond with the case of the “bystanding driver,” as explained above (pp. 16–21).

⁷² See, e.g., Jennifer Maselli, *Uber’s “Self-Driving” Test Cars to Be Overseen by Driver and Engineer*, SCI. AM. (Aug. 25, 2016), <http://www.scientificamerican.com/article/uber-s-self-driving-test-cars-to-be-overseen-by-driver-and-engineer> [https://perma.cc/U7P8-W697]; Aarti Shahani, *Uber to Roll Out Self-Driving Cars in Pittsburgh*, NPR (Aug. 18, 2016, 4:28 PM), <http://www.npr>

So this turns out to be a useful time for the book's intricate controversies to be so richly aired. But it is also a useful time to highlight major avenues of inquiry that the book, and the trolley discourse more generally, have bypassed. In the remainder of this Review, by presenting original empirical research, I hope to turn some of our attention to two further lines of questioning: about the role of law, and in particular, about the subset of law that shapes or defines social roles.

1. *The Role of Law.* — Current philosophical debates about the trolley problem have tended not to address one basic source of possible influence on moral intuitions: what the law requires or allows. There appears to be little consideration of whether our moral intuitions and judgments could or should be shaped — or might already be shaped — by the presence of laws prohibiting, permitting, or even requiring the sacrifice of someone's life to save others.⁷³

To be clear, what has been under-studied is the potential pull of positive law, as a set of commands or permissions, rather than the possible influence of legal ideas or concepts. One need not search far at all for signs that law-like *thinking* might be influencing the trolley discourse. Recall Kamm's proposed Principle of Permissible Harm, for example. If anyone might find something intuitive about notions of the proximity of causation, it would be current and former students of the law. Love it or hate it, we do appreciate how the notion of proximate cause can meet its task of cutting off legal responsibility.⁷⁴ Why not similarly cut off *moral* responsibility for killing the one, based on a lack of causal proximity to the act? Or why not erase moral credit for saving the five, in the same way? For that matter, why not compare how relatively proximate the good and bad outcomes are? Now we are getting close to recasting Kamm's principle as a "relative proximate cause" theory of morally permissible harm.⁷⁵

.org/sections/thetwo-way/2016/08/18/490524020/uber-to-roll-out-self-driving-cars-in-pittsburgh [https://perma.cc/BR3S-XQHB].

⁷³ This may be especially surprising given how much attention, as noted above, has been paid to potential influences going in the *other* direction: how the moral intuitions revealed by trolley scenarios might influence law and policy, and whether they should. See *supra* p. 661.

⁷⁴ We also appreciate how this line-drawing problem can generate endless conceptual and pragmatic debate, when it calls for choosing a breakpoint on a continuum with no clear demarcations. This discussion does not mean to overlook the many well-known problems with the concept of proximate cause, or with the notion of a "causal chain." It's only to say that these conceptual devices (or fictions) are familiar in the law. Familiarity may not breed enduring admiration, of course; the new Restatement (Third) of Torts has abruptly purged the term "proximate cause" entirely from the lexicon in favor of a "harm within the risk" rubric. See RESTATEMENT (THIRD) OF TORTS § 29 (AM. LAW INST. 2005).

⁷⁵ Moral psychologists have proposed and tested a theory of "intervention myopia" that similarly supposes a diminished attribution of moral responsibility due to causal distance. See Michael R. Waldmann & Jörn H. Dieterich, *Throwing a Bomb on a Person Versus Throwing a Person on a Bomb: Intervention Myopia in Moral Intuitions*, 18 PSYCHOL. SCI. 247 (2007).

It should be little surprise that much trolley argumentation sounds in the same register as legal argument.⁷⁶ Moral dilemmas and legal controversies often overlap, to say the least, in substance and in normative aims; one might recall that the thought experiment originated in Foot's article, in which she expressly draws on common law cases,⁷⁷ and that legal scholars and policymakers have joined philosophers in the trolley debates.

They also overlap in modes of reasoning — the moral philosopher's "reflective equilibrium" approach for fitting and justifying case-by-case judgments is what the lawyer might recognize as a cousin of the common law method.⁷⁸ Indeed, Kamm herself has reflected on why her case-based method seems so "lawyer-like."⁷⁹ (And if one might be wondering, yes, she is also quite familiar with *Palsgraf*, the iconic case about proximate cause.⁸⁰)

Yet for all this overlap, the trolley discourse has given little attention to the potential influence of the law in the form of *laws* — as rules of conduct that might exert influence at the micro level of our moral intuitions about specific cases. This deficit is especially notable with respect to substantive criminal law, which is said to have the purpose

(providing experimental evidence, using trolley scenarios among others, consistent with the hypothesis that "in their moral evaluations, people tend to focus on the causal path of the agent or patient targeted by their intervention," *id.* at 249).

⁷⁶ It would be *no* surprise — and maybe even axiomatic — for those subscribing to the "moral grammar" theory of intuitive moral judgments urged by Professor John Mikhail. See John Mikhail, *Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge*, in 50 *PSYCHOLOGY OF LEARNING AND MOTIVATION* 27, 29 (Brian H. Ross ed., 2009) ("The moral grammar hypothesis holds that ordinary individuals are intuitive lawyers, who possess tacit or unconscious knowledge of a rich variety of legal rules, concepts, and principles, along with a natural readiness to compute mental representations of human acts and omissions in legally cognizable terms." (citations omitted)).

⁷⁷ Foot, *supra* note 20, at 30 (referring to cases including the "famous case of the two sailors, Dudley and Stephens, who killed and ate the cabin boy when adrift on the sea without food").

⁷⁸ As Kamm crisply explains, "[u]sing our intuitive judgments about which implications for cases are correct helps us decide among, and also revise, theories and principles" (p. 13). On specific varieties of the method of "reflective equilibrium," see Norman Daniels, *Reflective Equilibrium*, in *STANFORD ENCYCLOPEDIA OF PHILOSOPHY* (Edward N. Zalta ed., 2013), <http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium> [<https://perma.cc/Z928-XKLZ>]. In short:

The method of reflective equilibrium consists in working back and forth among our considered judgments (some say our 'intuitions') about particular instances or cases, the principles or rules that we believe govern them, and the theoretical considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them. The method succeeds and we achieve reflective equilibrium when we arrive at an acceptable coherence among these beliefs.

Id.

⁷⁹ F.M. KAMM, *BIOETHICAL PRESCRIPTIONS: TO CREATE, END, CHOOSE, AND IMPROVE LIVES* 553 (2013).

⁸⁰ She mentions it in defense of the value of considering seemingly far-fetched fact patterns (p. 221).

of reinforcing or even generating moral intuitions of prohibition.⁸¹ Accordingly, to begin filling this gap, each of this study's experiments will test the impact of a legal prohibition on moral intuitions, by presenting scenarios that criminalize the turning of the trolley.⁸²

2. *The Law of Roles.* — Another, more subtle way that the law might influence moral intuitions is by defining official or social roles, which in turn set our expectations about correct or blameworthy behavior.⁸³ Think again of the trolley driver. A newcomer to the thought experiment probably feels a bit of this reflex: "But of course he must turn the trolley — he's the *driver*, responsible for everyone's safety."⁸⁴ And in the case of the transplant surgeon: "But of course she can't sacrifice the patient — she's a *doctor*, after all. Do no harm."

Such a reflex is not mere naiveté. There is something more to these characters than just whether their choice is about killing or letting die. Unlike the simple bystander, the driver and the surgeon have defined roles that affect what we intuitively think they should do.⁸⁵ From the deontologist's perspective, such roles might be said to alter one's positive or negative duties. Indeed, this possibility was one of Thomson's original reasons for shifting the trolley problem discourse from the driver to the bystander.⁸⁶ At other times, however, the core

⁸¹ See, e.g., MICHAEL MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW 207–08 (1997); Donald Braman, Dan M. Kahan & David A. Hoffman, *Some Realism About Punishment Naturalism*, 77 U. CHI. L. REV. 1531, 1535–36 (2010); Paul H. Robinson & John M. Darley, *Intuitions of Justice: Implications for Criminal Law and Justice Policy*, 81 S. CAL. L. REV. 1, 18–19 (2007). Whether civil wrongs are so readily mapped onto moral prohibitions is more controversial. See, e.g., ARTHUR RIPSTEIN, PRIVATE WRONGS 6 (2016) (providing an account of tort law that "start[s] from the moral idea that no person is in charge of another"); John C.P. Goldberg & Benjamin C. Zipursky, *Torts as Wrongs*, 88 TEX. L. REV. 917, 919 (2010) (recognizing torts as "legal wrongs rather than moral wrongs").

⁸² The study design is described below, *infra* Part II, pp. 680–85.

⁸³ The question of social roles has received much attention, including empirical study, by moral philosophers, experimental philosophers, and moral psychologists; and my discussion here is not meant to suggest otherwise. See, e.g., Jonathan Haidt & Jonathan Baron, *Social Roles and the Moral Judgement of Acts and Omissions*, 26 EUR. J. SOC. PSYCHOL. 201 (1996) (surveying literature and presenting experiments varying social roles as stranger/best friend/acquaintance; boss/employee/coworker; passenger/captain; relative/unrelated; subordinate/peer).

⁸⁴ In an analogous scenario involving a lifeboat, experimental evidence shows differences in people's moral judgments when the actor is a captain versus a passenger. Haidt & Baron, *supra* note 83, at 215 (showing interaction effects between act/omission distinction and social roles, using scenarios including one in which "[a] person in a crowded lifeboat (the captain or a passenger) either fails to throw a rope to a drowning person, or else pulls a rope away from a drowning person," *id.* at 213).

⁸⁵ This concern is naturally at the front of the lawyerly mind. For a survey of contemporary debates over role-differentiated morality, especially as it relates to lawyers and legal ethics, see W. Bradley Wendel, *Professional Roles and Moral Agency*, 89 GEO. L.J. 667, 669–81 (2001) (reviewing ARTHUR ISAK APPLBAUM, *ETHICS FOR ADVERSARIES* (1999)). See also DAVID LUBAN, *LAWYERS AND JUSTICE* (1988); WILLIAM H. SIMON, *THE PRACTICE OF JUSTICE* (1998).

⁸⁶ She observes: "In the first place, the trolley driver is, after all, captain of the trolley. He is charged by the trolley company with responsibility for the safety of his passengers and anyone

trolley debates have flagged this possibility only to say that it will be ignored.⁸⁷

One of the experiments I describe below will highlight, rather than suppress, the potential for social roles to affect our moral intuitions. Moreover, it constructs a scenario in which the demands of the role can be regulated by the law⁸⁸: A railroad engineer (rather than a casual bystander) is the one standing at the switch. And the experimental conditions vary this engineer's legal obligations.⁸⁹ Will our moral intuitions be moved, to know that in such a situation the law requires the engineer to minimize casualties?

The experiment thus brings a further dimension to the inquiry of "who turned the trolley" that Kamm has pursued.⁹⁰ There are at least a couple ways to characterize the influence, if any, of such role regulation on our moral intuitions. First and most simply, one might say that the law imposes a new positive duty or strengthens the existing positive duty to save the five, by creating a "special relationship." Second, one might say that the legal duty alters the framing of the dilemma, converting it from a "killing" case into a "saving" case.⁹¹ Or to put it more elaborately, whether a given moral duty is characterized as positive or negative depends on the baseline⁹² — and in a scenario

else who might be harmed by the trolley he drives. The bystander at the switch, on the other hand, is a private person who just happens to be there." Thomson, *supra* note 20, at 1397.

⁸⁷ In her 2008 article, Thomson recognizes that it might seem extra-appealing that the trolley driver should turn the trolley because "[p]erhaps we think of a trolley driver as charged, as part of his duties, with seeing to the safety of the men who are working on the tracks"; but she assumes away this appeal, saying that "we should prescind from the possibility that the agents in the cases we are considering have special duties towards the other parties." Thomson, *supra* note 25, at 370.

⁸⁸ Cf. Sunstein, *supra* note 11, at 923 (observing that "[l]aw can help constitute roles" and that "[o]ften law tries to redefine roles").

⁸⁹ As with Kamm's invention of the bystanding driver case, one might see my engineer's case as scrambling expectations about positive/negative duties — albeit in an experimentally controlled way. To be clear, the engineer in my experiments is not the driver of the trolley.

⁹⁰ This is not to suggest that Kamm is insensitive to how social roles might affect the boundaries of moral conduct. Far from it. She touches on the role of roles in noting that a bystander who is a bodyguard, or a friend, or a promisor, might have different moral obligations than a mere bystander (pp. 33, 88). She also notes the possible view that someone who has already begun a rescue might have an obligation to continue (p. 43). And most relevant here, she contemplates that the driver might have a duty to drive in the "best possible way" (p. 35).

⁹¹ See Kelman & Kreps, *supra* note 58, at 209–12 (finding notable empirical differences in the sturdiness of intuitions between sacrificial dilemmas framed as killing cases and those framed as saving cases).

⁹² As Professor Ronald Dworkin elegantly observed: "It is unclear what it means to let nature take its course. If it is natural to try to rescue five people at the cost of one, then throwing the switch is letting nature take its course. But perhaps 'nature' means nonintelligent nature, so that a potential rescuer lets nature take its course by pretending that he is not there. But why should he?" RONALD DWORKIN, *JUSTICE FOR HEDGEHOGS* 298–99 (2011). One might read this observation to suggest that even in the original bystander case, the proper baseline is open to question.

where the engineer is said to have a legal duty to minimize casualties, the salient counterfactual may be the death, rather than the survival, of the one.⁹³

II. INVESTIGATING LAW'S INFLUENCE

How much can the law shift our intuitions about moral dilemmas? Do our moral opinions, about whether it is right to kill one to save five, change if we are told that the law prohibits such a sacrifice? Or if we are told that the law will allow it as justified by the circumstances? What if we are told that the law *requires* the sacrifice?

The following randomized survey experiments seek to address these questions in a novel way: by varying the information stated about the law, within trolley scenarios, to reveal how moral intuitions respond.

A. Research Method

Randomized survey experiments are now quite common in the experimental-philosophy and moral-psychology literatures. In the past twenty years, they have become a familiar means for studying the trolley problem and other moral dilemmas,⁹⁴ including among legal scholars.⁹⁵ The main advantage of using randomized survey experiments is that causal inferences can be drawn from the results in a straightforward way, as with a randomized controlled trial in science or medicine. This method, it is worth noting, also parallels Kamm's technique for identifying which specific factors seem to matter to moral intuitions.

⁹³ One might even stretch to say that killing and letting die are flipped: letting the five die (when they should not, given a baseline of minimizing casualties) more resembles killing; and killing the one (when it should be done) more resembles letting die.

⁹⁴ See *supra* note 14 and sources cited *supra* note 54. Survey subjects are randomly assigned to read moral dilemma scenarios that vary in a specific way chosen by the experimenter (for example, a single fact is changed — in my case, a fact about the law). The outcomes of interest are the subjects' expressed moral opinions. Due to the initial random assignment of scenarios, the observed differences in moral opinions (as compared among the scenarios) can be attributed causally to the factor that has been varied by the experimenter.

⁹⁵ See JOHN MIKHAIL, ELEMENTS OF MORAL COGNITION 319–60 (2011); Kelman & Kreps, *supra* note 58, at 203–09. Legal scholars have used similar experimental survey methods to test other psychological effects that relate to moral judgments, such as what factors affect moral outrage about eminent domain and whether the inclusion of a liquidated damages clause in a contract can alter a party's willingness to breach. Janice Nadler & Shari Seidman Diamond, *Eminent Domain and the Psychology of Property Rights: Proposed Use, Subjective Attachment, and Taker Identity*, 5 J. EMPIRICAL LEGAL STUD. 713, 742–47 (2008); Tess Wilkinson-Ryan, *Do Liquidated Damages Encourage Breach? A Psychological Experiment*, 108 MICH. L. REV. 633, 655–56 (2010).

tions — by varying one factor at a time, to see whether one’s moral judgment changes.⁹⁶

1. *Introducing the law.* — Survey subjects are randomly assigned to read variations of the same basic trolley scenario, differing only in the information given about what the law prohibits, permits, or requires. In other words, the potentially pivotal factor being isolated in these experiments is information about the law. They are designed to generate evidence about this primary research question: Can informing people about the *law* influence their *moral* intuitions about the trolley problem?

If the answer is “yes,” then a number of further questions arise. What are the potential mechanisms of the law’s influence? Which moral intuitions are more (or less) susceptible? Which kinds of legal commands exert more (or less) influence? A secondary aim of the present study is to draw out preliminary, suggestive evidence about these subsequent questions and to motivate their further investigation.

2. *Law or liability?* — Of particular note, the experimental scenarios here are designed to distinguish between the presence of law and the threat of liability.⁹⁷ Some scenarios describe what the law says but assure the reader that the actor will not face any actual liability because the law will not be enforced. Other scenarios, by contrast, make clear that the actor will be held liable.

Making this distinction in the scenarios allows us to address one question about the possible mechanisms of influence: When the presence of law moves the moral needle, is it due only to the presence of a threat of liability? Is it, for example, because people think that the punitive consequences for the actor ought to count in the moral calcu-

⁹⁶ As she explains it, her trolley variations “are specifically constructed, like scientific experiments, to distinguish among and test theories and principles” (p. 13). Elsewhere, she likens her use of thought experiments to how “[s]cientists use experiments in which they can change one variable at a time, holding everything else constant, in order to see if that variable is crucial to an explanation of a phenomenon.” KAMM, *supra* note 79, at 579. Further, she elaborates that “just as artificially controlled conditions in a lab can lead to results that are applicable to real life, the results of artificial thought experiments might help us explain intuitive responses in ‘messier’ cases closer to real life or in real life.” *Id.* Kagan also observes that Kamm’s method begins with “psychological reconstruction,” in which “[s]he is doing her best to identify the various features that actually influence our intuitions about the different cases” (p. 157). He hastens to add that this descriptive inquiry is only a prelude to the normative, for, “of course, Kamm is interested in more than psychology,” in that “[s]he is looking for the *correct* moral principle” (p. 158).

⁹⁷ As Schauer observes, “even when law is in its most overtly regulatory and commanding mode, the existence of legal obligation is logically distinct from the sanctions and threats that law employs to enforce its commands and the obligations that law creates.” SCHAUER, *supra* note 11, at 31.

lus?⁹⁸ Or, instead, can the law's instruction affect moral intuitions even when there are *no* legal consequences for the actor?⁹⁹

B. Survey Population

The experiments described below include the answers from 1400 subjects in surveys conducted in April and May 2016. All are adults living across the United States;¹⁰⁰ they are volunteers recruited by the survey design and polling firm SurveyMonkey.¹⁰¹ Unlike other common sources of experimental subjects, such as online labor pools,¹⁰² the subjects in this study were neither paid a piece rate for each survey they took, nor a time-based wage.¹⁰³ In addition, all samples exclude subjects who reported that they could not take the exercise seriously;

⁹⁸ This study is not designed to compare criminal sanctions with civil or regulatory sanctions, however, and nothing in the following discussion should be read as implying a parity between them. As will be evident in the descriptions below, the scenarios used in this experiment specify only that the actor will be held liable but not what the precise sanctions would be. Survey subjects are thus left to imagine for themselves both the sanctions that will result from liability and why such consequences might enter into the moral calculus. One might imagine, for instance, the potential costs to the actor's family if she were convicted of manslaughter (or had to pay impoverishing amounts of civil penalties). Maybe not everyone would describe such an effect as a mechanism of law's influence on moral judgment, preferring instead a description like "how the costs of acting enter into a moral calculus." The possibility of this conceptual distinction amplifies the motivation for separating law from liability in the experiments.

⁹⁹ It is certainly possible to interpret the difference between these two types of scenarios as a matter of degree, rather than kind. One might say that a law that gets enforced sends a stronger message than one that doesn't get enforced — whether in general, or in a specific actor's case. One might then pose the question more subtly, as Schauer does: "Could law have the opinion-forming or opinion-influencing it has, however much that may be, without the way in which the sanction arguably underlines the importance of the legal norm itself?" SCHAUER, *supra* note 11, at 73.

¹⁰⁰ Among them, 58% are women.

¹⁰¹ This study does not make claims about representativeness, but SurveyMonkey does "run regular benchmarking surveys to ensure our members are representative of the U.S. population." *Our Audience*, SURVEYMONKEY, <https://www.surveymonkey.com/mp/audience/our-survey-respondents> [<https://perma.cc/KR8N-9N2R>].

¹⁰² In particular, use of Amazon's Mechanical Turk online labor pool for trolley experiments has been criticized. See Christopher W. Bauman et al., *Revisiting External Validity: Concerns About Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology*, 8 SOC. & PERSONALITY PSYCHOL. COMPASS 536, 548 n.2 (2014) (noting that heavy use of Mechanical Turk for trolley problem experiments has resulted in high levels of familiarity with the thought experiment within the labor pool).

¹⁰³ As SurveyMonkey explains: "We reward members with non-cash incentives to discourage rushing through surveys just for the reward." SURVEYMONKEY, *supra* note 101. The standard reward is a donation of \$0.50 to a charity and entry into a sweepstakes for a small prize. This method of recruitment raises the possibility that these survey subjects may be more charitably minded than other samples one might draw from the general population (although the donations are extremely small). In this study, however, the question of interest is how answers change *within* the same population of survey subjects. It remains possible that the findings may not generalize to other samples or to the general population (for instance, if this group reacts differently to the legal stimuli than other groups do).

who have taken another moral-dilemma survey in the past year; or who have gone to law school or formally studied moral philosophy. Moreover, a factual review question was used to screen out subjects who did not understand the information about the law given in the scenario. Finally, these surveys were conducted with the approval of the Columbia University institutional review board.

C. Survey Design

The trolley scenarios used for this study are all variations of the most basic bystander case. This case has been widely studied in prior experimental work and is thought to induce a moral judgment that is very broadly shared.¹⁰⁴ And yet, as is evident in the book, this seemingly well-settled scenario has become an epicenter of analytical debate.

For the sake of continuity with prior literature, the scenarios I use are adapted from ones recently used by Kelman and Kreps, which bear a resemblance to scenarios previously used by Mikhail, which in turn track Thomson's original version.¹⁰⁵ The key difference between these earlier studies and mine, of course, is that they focused on factors other than information about law or liability.

Following the usage in the standard literature, including in the book's debates, this study posits that an act can be deemed "morally prohibited," "morally permissible," or "morally required."¹⁰⁶

D. Experiment 1: Introducing a Legal Prohibition

The first experiment adopts the classic bystander scenario, in its usual form, with five lives saved by sacrificing one. Let's call it "Bystander Saves Five" for short. The scenario begins as follows, for all survey subjects:

¹⁰⁴ See *supra* note 57.

¹⁰⁵ See Kelman & Kreps, *supra* note 58, at 204–05 tbl.1; MIKHAIL, *supra* note 95, at 78–80 (listing scenario variations from experiments Mikhail has run).

¹⁰⁶ Among deontologists such as Kamm and Thomson, the term "morally permissible" is often used to mean "not morally prohibited"; when it is used this way, the term includes "morally required." But they also use the term "morally required" when it is useful to distinguish it from permissible-but-not-required. (In the survey questions here, these three choices are laid out together, and it is obvious to the subject that she must pick only one choice. Thus, in this context, the subject would not read "morally permissible" to include "morally required.") Also following the literature, this study does not allow for an act to be simultaneously "morally prohibited" and "morally required" (and so forth). It may seem sensible to suggest that the whole point of imagining a moral dilemma is to raise such a possibility (that an act is at once both prohibited and required) — but that's just not how the terminology works in the literature. Rather, each of the three labels is assumed to refer to the (unique) conclusion that one reaches after grappling with the dilemma. One further note: following the book's analysis, this study will not consider the category of "supererogatory" judgments (deeming the act praiseworthy but not required).

The Story

Michael is taking his daily walk near the railroad tracks when he notices a freight train approaching. The train seems out of control. Michael can see that the driver is slumped over, unconscious.

There are five workmen on the tracks ahead. The train is now rushing towards them. They don't see the train coming, and Michael can't get their attention. He yells at them, but the construction noise is too loud. He waves his arms, but they are looking the other way.

Michael happens to be standing near a railroad switch. He can reach it easily, and he knows that pulling the switch will turn the train onto a side track before it hits the five workmen.

However, there is another workman on this side track. He also doesn't see or hear the train coming, and Michael can't get his attention either.

Michael can choose to pull the switch, or he can choose not to do anything.

If Michael pulls the switch, he knows that the man on the side track will be killed by the train.

If Michael doesn't do anything, he knows that the five men on the main track will be killed by the train.

This story then concludes with one of the following passages, randomly assigned. (The labels in brackets are for this essay's expository purposes and are *not* shown to the reader.)

{Criminal & liable}

Under the law, pulling the switch would count as criminal homicide, and Michael will be held liable for doing it.

Michael won't be held liable if he doesn't do anything.

{Criminal & not liable}

Under the law, pulling the switch would count as criminal homicide. But the local prosecutor won't bring charges, and so Michael won't be held liable for doing it.

Michael also won't be held liable if he doesn't do anything.

{Justified & not liable}

Under the law, pulling the switch would be considered a justified act.
This means that Michael won't be held liable for doing it.

Michael also won't be held liable if he doesn't do anything.

E. Experiment 2: Introducing a Legal Duty

In the second experiment, the scenario is designed to test for the law's potential influence through another mechanism — by imposing a role-based legal duty.¹⁰⁷ The scenario tells the reader at the outset that the character Michael is a railroad engineer,¹⁰⁸ and the experimental conditions vary whether this role entails a duty to “minimize casualties.” Note that, in a departure from the literature, the tradeoff is set at two lives saved by sacrificing one.¹⁰⁹ The reason is that with a five-to-one tradeoff, the baseline percentage of subjects who would say it is “morally prohibited” for a railroad engineer to pull the switch would likely be too low to allow detection of any (further) suppression of that moral intuition.¹¹⁰ Let's call this the “Engineer Saves Two” scenario.

The legal duty to act is introduced via a set of randomly assigned story endings. As with the criminalization scenarios, there are two versions of the duty condition, one with liability and one without. In addition, note that the {Justified & not liable} scenario now says that *both* choices are deemed “justified.”¹¹¹ The two criminalization conditions are also included; they are identical to those listed above. Altogether, the five randomized story endings are as follows:

¹⁰⁷ The reason that legal duty is not included in the first experiment is that subjects may balk at the suggestion that a casual bystander who happens upon a railroad switch could possibly have a legal duty to use it to kill someone, even if it saves net lives. Excluding such a legal condition helps to avoid distortions that might arise from prompting an oppositional attitude (one might call them backlash or outrage effects) among subjects who view such a legal duty as plainly unjust or absurd.

¹⁰⁸ The first line of the story begins: “Michael, who is a railroad engineer, is standing near the railroad tracks . . .”

¹⁰⁹ The reason for the two-to-one tradeoff is to make the case as hard a call as possible (while still being more than one-to-one).

¹¹⁰ This caution is suggested by the results from Experiment 1, as well as those from prior work, showing that typically a very low percentage of subjects will say that even a casual bystander (never mind a railroad employee) is morally prohibited from pulling the switch.

¹¹¹ One might imagine this presents a more “neutral” middle option than in the analogous condition in the Bystander Saves Five scenario. The reason that the {Justified & not liable} condition in the Bystander Saves Five scenario does not say that inaction is “considered justified” is that such a statement about the law would not make as much sense when it is not plausible that inaction could be penalized in the first place. See *supra* note 107 (explaining why the first experiment does not include treatments that impose a legal duty to save).

{Criminal & liable}

Under the law, pulling the switch would count as criminal homicide, and Michael will be held liable for doing it.

Michael won't be held liable if he doesn't do anything.

{Criminal & not liable}

Under the law, pulling the switch would count as criminal homicide. But the local prosecutor won't bring charges, and so Michael won't be held liable for doing it.

Michael also won't be held liable if he doesn't do anything.

{Justified & not liable}

The law neither prohibits nor requires either choice, because the law considers both choices to be justifiable.

This means that Michael won't be held liable, either way.

{Duty & not liable}

The law requires Michael, who works for the railroad, to try to reduce casualties from accidents.

This means that Michael won't be held liable, if he pulls the switch. But he also won't be held liable, if he does nothing, because the law won't be enforced in this case.

{Duty & liable}

The law requires Michael, who works for the railroad, to try to reduce casualties from accidents.

This means that Michael won't be held liable, if he pulls the switch. But he will be held liable, if he does nothing.

F. Further Preliminaries

Several additional features and limitations of these experiments are worth noting at the outset. First, although an air of unreality may be

unavoidable in any trolley problem scenario,¹¹² these variations include a few factual details meant to preempt ways for subjects to dodge the hard moral question (for example, by imagining creative means by which everyone can be saved, or by protesting that a mere bystander couldn't possibly operate a railroad switch).¹¹³ Second, all subjects are required to write out a brief explanation for the moral judgment that they report; thus subjects are encouraged to grapple with the moral dilemma directly.¹¹⁴ Third, to minimize the concern that some survey subjects just cannot take a thought experiment seriously, the survey expressly offers all subjects the option (at the end of the survey) to say "ignore my answers" for that reason.¹¹⁵ Fourth, immediately following the presentation of the scenario, subjects are also required to answer a reading comprehension question that asks them what the scenario said about the law;¹¹⁶ those who answer incorrectly are excluded, but it remains possible that some subjects may remain unclear about the law even if they answer correctly.¹¹⁷ Fifth, these experiments only present two fairly similar scenarios, and one

¹¹² Because this study focuses on the turning-the-trolley scenarios, it also happens to avoid the criticism that some survey subjects may be prompted to laugh, or otherwise toggle out of a mindset of moral reasoning, by scenarios that seem awkward or uncomfortable (most notably, the classic so-called "Fat Man" scenario — also known as the "Footbridge" or "Drawbridge" scenario — in which the train can be stopped by pushing a large man in front of it). See Bauman et al., *supra* note 102, at 541–42 (citing Sandel, *supra* note 21, at 4:33).

¹¹³ The text also makes clear that the actor must choose (only) among the specified options.

¹¹⁴ Survey subjects see this question immediately following (and on the same survey page as) the main question of the survey, which asks them to choose among "morally prohibited," "morally permissible," and "morally required." How to interpret the actual content of the subjects' explanations for their judgments is a much more complicated question; the analysis below will not address that content. See, e.g., Fiery Cushman, Liane Young & Marc Hauser, *The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm*, 17 PSYCHOL. SCI. 1082 (2006); Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOL. REV. 814 (2001).

¹¹⁵ Nearly 5% of subjects chose this option. I borrow this technique from Greene et al., *supra* note 56, at 366. Moreover, subjects are able to quit the survey at any point (and are thus excluded from the data) by closing their browser or their app. They are expressly reminded of this option in the initial consent page of the survey.

¹¹⁶ Subjects are able to reread the scenario while answering this comprehension question. It must be noted that for some subjects the presence of this question may be a demand characteristic suggesting either that the experimenter wants them to report moral intuitions that accord with the law, or contrarily, that the experimenter wants to see if they can set aside the law in thinking about morality. Such a concern cannot be ruled out, although it may help that these surveys were answered anonymously online (and distributed with randomization by a third-party survey firm), and that each subject saw only a single scenario.

¹¹⁷ For example, although the no-liability scenarios clearly state that Michael will not face liability, some subjects may not believe such an assertion, and instead suspect that some legal consequences might nonetheless befall Michael. It is also possible that some subjects may feel that they should make their moral judgments based on the fact that sanctions exist and might befall someone (other than Michael) who faced such a situation.

ought not to assume that the results generalize to other moral dilemmas or other moral problems.

Finally, it should be noted that this study asks subjects simply to report their moral judgments. It does *not* ask people how they would act, if they themselves were standing at the switch.¹¹⁸ Thus, a typical criticism of survey experiments — that in a laboratory setting we cannot learn what people would actually do in a real setting, but only what they *think* they would do — has little relevance here. These surveys directly document people's moral judgments about someone else's actions, and those moral judgments are the main object of study.

III. EVIDENCE OF LAW'S INFLUENCE

A. *Types of Influence*

It will be convenient, for interpreting the survey results, to categorize the various possible effects on moral intuitions into the following two sets. The primary set of effects might be called the law's "directional influence" on moral intuitions, whereby the law's command pulls moral intuitions in that same direction. A second set might be called "the pull of neutrality," whereby the law's express permission — here, saying that the act is justified — pulls intuitions inward toward a "morally permissible" judgment and away from the two extremes.

1. *Directional Influence.* — If telling subjects that pulling the switch amounts to criminal homicide has any effect on moral intuitions, one might expect it to reinforce a sense that the act is (also) morally prohibited.¹¹⁹ One might also expect it to suppress a sense that the act is nevertheless morally permitted or morally required. Altogether, we should expect the introduction of the law's prohibition to push resulting judgments toward a "morally prohibited" judgment and away from "morally required." In contrast, one might expect the law to push in the other direction if it imposes a duty to rescue or a duty to minimize casualties.

The various effects of a given legal treatment are not necessarily linked; all or none or only a subset may be at work for any given legal

¹¹⁸ The scenarios and the questions make clear that a fictional character, not the survey subject, is the one standing at the switch. Undoubtedly, for many people, empathy feeds into moral judgment; it is utterly natural for us to try to imagine ourselves in the situation of the fictional character. But one might argue that using such imagination or introspection to form a moral opinion does not imply that the resulting opinion isn't a true moral opinion. In any event, on this matter of methodology, this study follows much of the experimental philosophy literature, in which studies ask people directly for their moral judgments about a fictional character's choices within an imaginary moral dilemma scenario. See sources cited *supra* note 54.

¹¹⁹ This is not to suggest in any way that it's a simple matter to sort out which psychological mechanisms might be partly responsible for this influence. Questions of mechanism are raised (but not answered) in the Conclusion.

treatment, for any given intuition, on any given person.¹²⁰ For example, a criminal prohibition may strongly boost one person's aligned sense of moral prohibition, while not putting much of a dent in her competing sense that the act may be morally required. For a different person, the relative strengths of the influences might be reversed; for a third person, the law might exert no influence whatsoever on her moral intuitions. The surveys will measure, of course, only the aggregate of these varied effects and noneffects.

These predictions follow: (1) Relative to a scenario in which the law *criminalizes* the act, a scenario in which the law *requires* the act should show fewer subjects reporting "morally prohibited." (2) The legal requirement scenario should also show more subjects reporting "morally required." (3) The number reporting "morally permitted" could go either way, depending on the first two effects; that is, on whether more subjects switch in from "morally prohibited" or more switch out to "morally required."¹²¹

2. *The Pull of Neutrality.* — The law's neutrality may also have its own influence on moral intuitions. Specifically, it might exert an inward pull toward *moral* neutrality as well, and away from both moral poles (that is, a subject might think, "If the law permits both choices, then I suppose both choices are morally permissible, too."). Such an effect may be enhanced if the law is thought to express neutrality as a substantive choice — as when it declares that an otherwise prohibited act is "justified."¹²²

This complication does not affect the interpretation of any comparisons between a criminalization condition and a legally *required* condition. The first experiment, however, does not include a legally required condition; only comparisons between the criminalized and

¹²⁰ What is not contemplated in these predictions is that the law's instruction might boost the *opposed* moral intuition (criminalizing the act should not make *more* subjects say "morally required"), or suppress the aligned intuition (making *fewer* subjects say "morally prohibited"). One can imagine contrarian subjects who take an oppositional attitude to the law — perhaps because they view a given law as illegitimate or absurd, or perhaps because they have suffered bad experiences with the legal system. I have sought to minimize the chances of inviting such an oppositional attitude by using legal conditions that should *not* be broadly unbelievable (unlike imposing a general duty on a bystander to pull the switch) or patently unjust (such as a scenario where the actor would be held liable for either choice). It may be interesting for future work to address the possibility of backlash or other such oppositional effects, as well as possible crowding-out effects.

¹²¹ Thus the effects on the extreme choices, (1) and (2), are the unambiguous predictions.

¹²² See, e.g., Gabriella Blum & John C.P. Goldberg, *War for the Wrong Reasons: Lessons from Law*, 11 J. MORAL PHIL. 454 (2014) (articulating a three-part legal design common to criminal law and tort law that, first, sets out a duty, second, specifies when breach is deemed privileged — such as under a doctrine of justification — and third, specifies when privilege is inapplicable due to abuse, *id.* at 462; and suggesting reasons for potential gaps between legal and moral permissibility, *id.* at 458–59).

legally neutral conditions are possible. Thus it is useful to state these predictions, for purposes of the first experiment:

(1) Relative to a scenario in which the law considers the act to be *justified*, a scenario in which the law *criminalizes* the act should show more subjects reporting “morally prohibited.”¹²³ (2) The number reporting “morally required” could go either way, depending on whether the directional influence of criminalization (a decrease) outweighs relief from the pull of neutrality (an increase).¹²⁴ (3) The number reporting “morally permitted” could go either way, depending on the first two effects; that is, on whether more subjects switch out to “morally prohibited” or more switch in from “morally required.”¹²⁵

¹²³ But this difference should not be attributed entirely to the influence of criminalization because some of the difference may be due to the pull of legal neutrality. Consider a subject who is on the fence between “morally prohibited” and “morally permissible.” She might be pulled toward the former by a criminalization condition, and she might be pulled toward the latter by a legally neutral condition.

¹²⁴ To see this, consider a subject who is on the fence between “morally permissible” and “morally required.” He might be pulled toward the former by *both* the criminalization *and* the legally neutral scenarios.

¹²⁵ In the first experiment, then, effect (1) is the sole unambiguous prediction.

TABLE 1. BYSTANDER SAVES FIVE

	Morally prohibited	Morally permissible	Morally required	<i>N</i>
Criminal & liable	16.0%	60.3%	23.7%	156
Criminal & not liable	8.8%	70.8%	20.4%	137
Justified & not liable	4.4%	71.1%	24.4%	180

FIGURE 1A. BYSTANDER SAVES FIVE

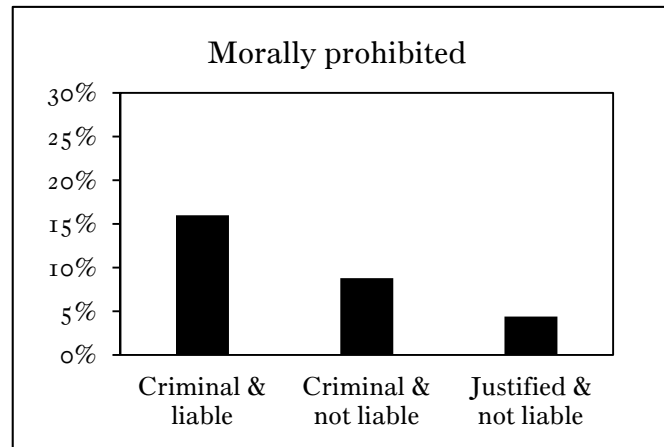
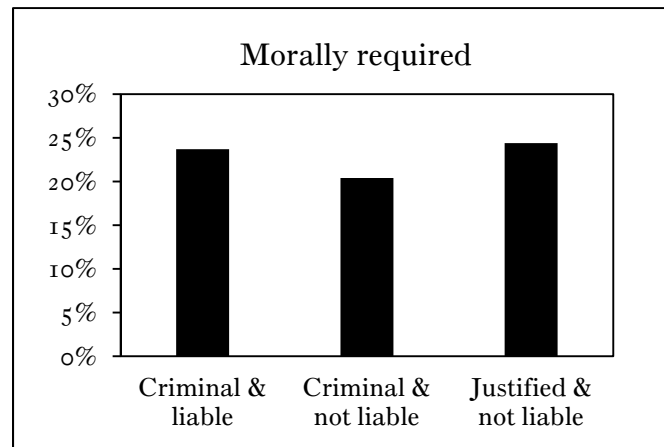


FIGURE 1B. BYSTANDER SAVES FIVE



B. Findings from Experiment 1: The Role of Law

The reported moral intuitions resulting from the varying conditions in the first experiment are shown in Table 1 and Figures 1A and 1B. We may begin by asking whether a legal prohibition, operating at full force, can influence moral intuitions. In this experiment, the most relevant comparison is between the {Justified & not liable} condition and the {Criminal & liable} condition. The share of subjects who deem pulling the switch to be “morally prohibited” is considerably higher when that choice entails criminal liability: the 16.0% share in the {Criminal & liable} condition is more than three times the 4.4% share in the {Justified & not liable} condition.¹²⁶

As explained above, this observed difference should not be attributed entirely to the moral pull of criminalization, because some of the gap might be due to a suppression of the “morally prohibited” responses in the legal neutrality condition. Rather, the difference can be interpreted as resulting from the combined effects of this change of legal regime.

We then turn to whether the fact of criminalization might influence moral intuitions even if it is known that this actor will not be held liable. The relevant comparison is between the {Justified & not liable} condition and the {Criminal & not liable} condition. There may be some change in the share reporting “morally prohibited,” in the expected direction, but the difference is not statistically significant.¹²⁷

Meanwhile, recall the prediction that the number reporting “morally required” could go either way, when comparing the legally neutral and the criminalization conditions; and in the data, there seems to be no notable difference among these conditions. One simple interpretation would be that, relative to the {Justified & not liable} condition, the directional influence of criminal liability and (relief from) the pull of neutrality roughly cancel each other out.¹²⁸ But these data do not reveal whether each influence alone is large, small, or nonexistent.

¹²⁶ $\chi^2(1, N = 336) = 12.66, p < 0.001$. Note that the standard error for any individual proportion \hat{p} is given by the usual formula, the square root of $\hat{p}(1 - \hat{p}) / N$.

¹²⁷ The comparison is between 4.4% among the {Justified & not liable} subjects and 8.8% among the {Criminal & not liable} subjects. $\chi^2(1, N = 317) = 2.45, p = 0.117$.

¹²⁸ In other words, it remains possible that some moral pull away from “morally required” does exist but is masked because the legally neutral condition also pulls in the same direction. Equivalently, there may be a pull of neutrality that is masked by the pull of criminalization.

TABLE 2. ENGINEER SAVES TWO

	Morally prohibited	Morally permissible	Morally required	<i>N</i>
Criminal & liable	25.5%	55.2%	19.3%	145
Criminal & not liable	15.3%	67.4%	17.4%	144
Justified & not liable	9.4%	78.9%	11.7%	180
Duty & not liable	3.3%	62.1%	34.6%	153
Duty & liable	1.9%	56.9%	41.3%	160

FIGURE 2A. ENGINEER SAVES TWO

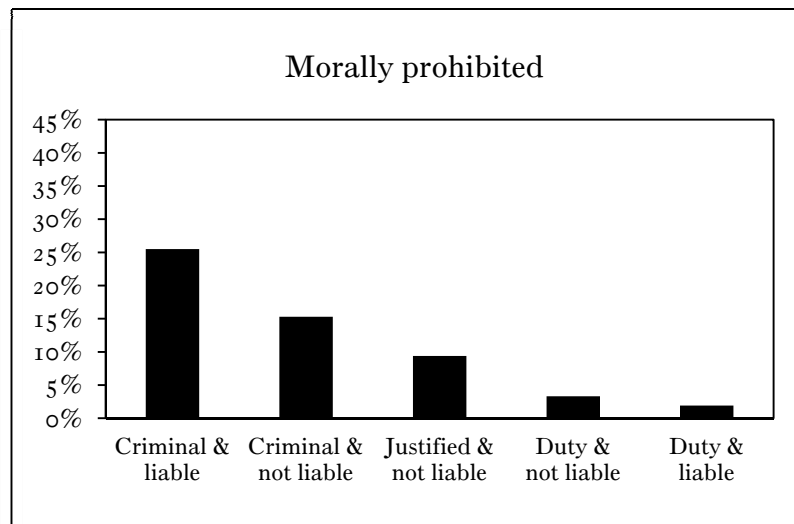
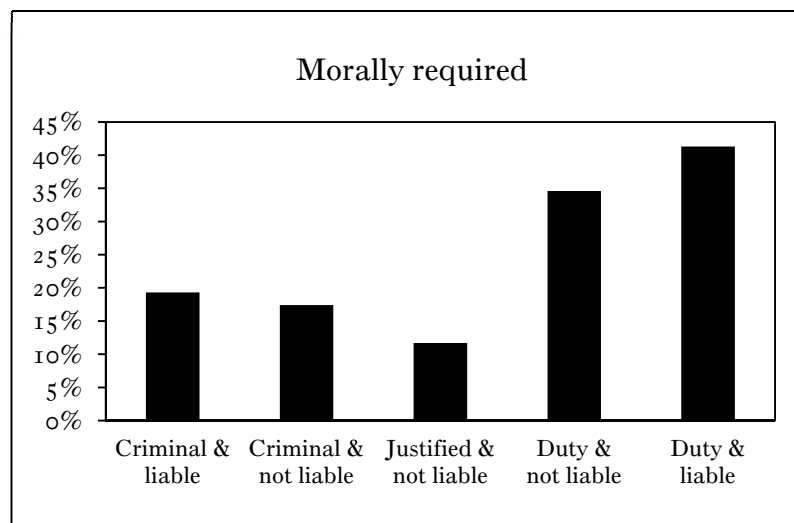


FIGURE 2B. ENGINEER SAVES TWO



C. Findings from Experiment 2: The Law of Roles

The reported moral intuitions resulting from the varying conditions in the second experiment are shown in Table 2 and Figures 2A and 2B. In this experiment, unlike the first, we are able to compare the two extremes of legal prohibition and legal requirement.¹²⁹ We may begin with both legal commands operating at full force. Between the {Criminal & liable} condition and the {Duty & liable} condition, the differences in both the “morally prohibited” and the “morally required” responses are quite large. The share of subjects who say that pulling the switch is “morally required” roughly doubles, from 19.3% in the {Criminal & liable} condition to 41.3% in the {Duty & liable} condition.¹³⁰

Even more notable, perhaps, is the nearly total suppression of the “morally prohibited” response: it falls from over 25% in the {Criminal & liable} condition to about 2% in the {Duty & liable} condition.¹³¹ Apparently, when the law requires the engineer to triage casualties, and is backed up by liability, almost nobody will say that it is morally prohibited. This collapse may seem all the more remarkable when one remembers that the tradeoff in lives here is only two-to-one, not five-to-one.

A similar set of patterns is evident when there is no liability accompanying either legal command. This second comparison of interest is between the {Criminal & not liable} condition and the {Duty & not liable} condition. Here, too, a sizeable difference appears in both the “morally prohibited” and the “morally required” responses. Again, subjects are about twice as likely to say “morally required” in the legal duty condition; 17.4% say so in the {Criminal & not liable} condition as compared with 34.6% in the {Duty & not liable} condition.¹³² And again, the share saying “morally prohibited” collapses, from over 15% in the {Criminal & not liable} condition to around 3% in the {Duty & not liable} condition.¹³³ Although these differences are smaller than in the comparison between the two conditions with liability,¹³⁴ they may

¹²⁹ This experiment’s setup has the advantage that we need not rely on comparisons with the {Justified & not liable} condition, with the attendant complications for interpretation. A brief word may be of interest, however, concerning the seeming dip in the share reporting “morally required” in the {Justified & not liable} condition relative to the criminalization conditions. Such a drop might be seen as evidence that the pull of neutrality away from the “morally required” judgment was even stronger than the pull of criminalization, on average; however, the observed difference is not quite large enough to say so with much confidence.

¹³⁰ $\chi^2(1, N = 305) = 17.17, p < 0.001$.

¹³¹ $\chi^2(1, N = 305) = 37.31, p < 0.001$.

¹³² $\chi^2(1, N = 297) = 11.44, p < 0.001$.

¹³³ $\chi^2(1, N = 297) = 12.95, p < 0.001$.

¹³⁴ Again, it should be noted that some subjects presented with the no-liability condition may not fully believe it or may think they should be answering the morality question not based on the consequences for Michael specifically but rather on the assumption that sanctions generally back

be more impressive in the sense that they are induced only by a change in what the law *says*.

D. Some Questions Raised

While it would be too hasty to suggest broad implications on the basis of this initial set of experiments, the findings here do help to motivate several lines of inquiry for law and policy design, for the growing body of experimental research focusing on moral dilemmas, and for the analytical trolley discourse, including for arguments raised in the book.¹³⁵ Most of these questions revolve around this central problem: What if the moral intuitions we can observe have already been shaped by people's impressions, however vague or subconscious, about what the *law* expects?

1. *Law and Policy Design.* — Supposing that there are good reasons to design new laws to reflect our moral beliefs in some contexts, should we (and how could we) correct for the feedback loops that occur when observable moral beliefs are already influenced by existing laws?¹³⁶

Or to take a more pragmatic perspective: In newly emerging contexts, should lawmakers aim to move speedily in promulgating rules, in order to nudge the initial orientation of these feedback loops? And if the mutual influence of law and moral intuitions is sufficiently strong, could such a first-mover approach even be used to induce a specific self-reinforcing equilibrium?

For instance, consider driverless cars. Should regulators quickly announce principles for answering the various dilemmas that Google Cars, Teslas, or other autonomous cars will likely face¹³⁷ — before that first tragic accident in the news begins to steer the public's moral intuitions in a specific direction? Should regulators hasten to make clear that the role of the car's programmer (like the role of the engineer in

up such laws. See *supra* note 117. If so, then caution is warranted in interpreting the difference between the {Criminal & not liable} and {Duty & not liable} conditions as representing the influence of the law absent sanctions.

¹³⁵ The Conclusion will point to questions, concerning the possible mechanisms of law's influence, that these findings raise for future empirical investigation.

¹³⁶ Needless to say, although I have posed the question here about moral intuitions writ large, the two reported experiments concern only moral dilemmas — and, indeed, only two specific trolley scenarios. Further investigations would be needed to know whether, and to what extent, the present findings are generalizable to other moral dilemmas or moral concerns.

¹³⁷ The newly released guidelines from the U.S. Department of Transportation highlight the need for a process for addressing sacrificial dilemmas, but do not specify substantive moral or ethical principles. U.S. DEP'T OF TRANSP., NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., FEDERAL AUTOMATED VEHICLE POLICY 26–27 (2016) (“Algorithms for resolving these conflict situations should be developed transparently using input from Federal and State regulators, drivers, passengers and vulnerable road users, and taking into account the consequences of an [automated vehicle’s] actions on others.”).

my second experiment) entails a duty to minimize casualties?¹³⁸ And after moral intuitions have already started down a given path,¹³⁹ would a more attention-getting regulatory intervention be needed to “reset” the feedback loop, so that intuitions and regulations begin to realign in favor of overall safety?¹⁴⁰

2. *The Trolley Debates.* — These experiments’ findings also motivate related questions for the standard method of trolley thinking, which takes our intuitions about individual cases as the raw moral data to be fit-and-justified through the careful crafting of more general principles.¹⁴¹ How should one proceed in crafting such general principles, if these supposedly “moral” data actually reflect some uncalibrated mix of our moral sense and our impressions about the positive law?

Or, to put it in the language of reflective equilibrium,¹⁴² could there be *multiple equilibria* — some in which purely moral intuitions and certain moral principles align, and others in which law-influenced intuitions and different moral principles align? And if so, what is the normative theory that explains how to choose among them?¹⁴³

3. *Experimental Research.* — A version of this issue also extends to the empirical work in the experimental-philosophy and moral-psychology literatures. What impressions about the law might already be reflected in the moral intuitions that these studies have been mea-

¹³⁸ It is possible, of course, that the first accident might steer moral intuitions in favor of such aggregation or triage — for example, if the tragedy is that the car did *not* minimize casualties, but instead saved the passenger’s life at the cost of many others’ lives.

¹³⁹ Consider, for example, the possible effects of widespread news coverage of a recent survey showing that people were less keen to buy a self-driving car that minimized casualties (median reported likelihood of 19, on a scale of 1 to 100) than a car that prioritized the owners’ and their families’ lives over the lives of others (median of 50). See Bonnefon et al., *supra* note 70, at 1574. Although the same study also showed widespread support for the principle that self-driving cars should be programmed to minimize casualties (especially if many people would be saved), both the study itself and the following news coverage sought to emphasize that people were hesitant to put their money where their morality is. See, e.g., Jacqueline Howard, *Driverless Cars Create a Safety “Dilemma”: Passengers vs. Pedestrians*, CNN (June 23, 2016, 3:31 PM), <http://www.cnn.com/2016/06/23/health/driverless-cars-safety-public-opinion> [https://perma.cc/W9RQ-YXKP]; Karen Kaplan, *Ethical Dilemma on Four Wheels: How to Decide When Your Self-Driving Car Should Kill You*, L.A. TIMES (June 23, 2016, 11:05 AM), <http://www.latimes.com/science/sciencenow/la-sci-sn-autonomous-cars-ethics-20160623-snap-story.html> [https://perma.cc/5XFG-29EK]; John Markoff, *Should Your Driverless Car Hit a Pedestrian to Save Your Life?*, N.Y. TIMES (June 23, 2016), <http://www.nytimes.com/2016/06/24/technology/should-your-driverless-car-hit-a-pedestrian-to-save-your-life.html> [https://perma.cc/T3Q9-SJMH].

¹⁴⁰ Note that the present experiments resemble this sort of re-intervention, if one assumes that people already held some impressions about what the law expects before reading the scenario’s new information about what the law expects.

¹⁴¹ See *supra* notes 51–52.

¹⁴² On the method of reflective equilibrium, see Daniels, *supra* note 78.

¹⁴³ This question would seem relevant even for those who might find it unremarkable (and unobjectionable) that our moral judgments might take the law into account to some degree.

suring? Might there be interactions between these legal impressions and the experimental factors whose effects the researchers have sought to isolate?¹⁴⁴

4. *The Role of Roles.* — One further point has already been suggested by the second experiment's focus on roles. The finding that varying the duties of the engineer's role (via the law) can affect our moral intuitions suggests that we should also be more attentive of changes in the actor's role across scenarios. Consider, for instance, a comparison between Kamm's bystanding driver case and the mere bystander case¹⁴⁵: an observed difference in moral judgments between these two cases might not be due to the distinction between what she calls a "killer let die" case and a simple "let die" case,¹⁴⁶ as these cases also differ in that one involves the role of a driver and the other does not.¹⁴⁷

Such a confounding factor might be present in any comparisons across roles — not least, in comparing the bystander with the original driver (who we think is responsible for safety),¹⁴⁸ and in comparing either of them with a transplant surgeon (who we think must do no harm). It may be possible for a determined philosopher to will away any such expectations;¹⁴⁹ it may even be possible, one supposes, to wipe survey subjects' minds clean of any such taint. But if role expect-

¹⁴⁴ I assume that this possibility of interaction effects is what should be of greatest interest for experimenters, because a sensible answer to the prior question is that it does not matter if subjects hold impressions about the law as long as those impressions on average do not differ across the randomized treatment and control groups.

¹⁴⁵ On Kamm's presentation of the bystanding driver case as a challenge to Thomson's new position, see *supra* notes 42–44 and accompanying text.

¹⁴⁶ See Kamm, *supra* note 42 and accompanying text. Likewise, a comparison between the bystanding driver case and the original driver case might not detect a difference in moral intuitions between someone whose status is a "killer letting die" and someone whose status is "killing" because each case still involves a driver, with role-based expectations that might dominate the effect of status.

¹⁴⁷ As other commentators on Thomson's new position have noted, the role of driver may be accompanied by duties that the driver assumed in taking on the role: "Would a driver have to turn into a wall in order to avoid killing others, as Thomson contends? That surely depends on what level of responsibility the driver assumed. It could be imagined that he assumes such responsibility in virtue of taking the job of driving the trolley." Alec Walen & David Wasserman, *Agents, Impartiality, and the Priority of Claims over Duties: Diagnosing why Thomson Still Gets the Trolley Problem Wrong by Appeal to the "Mechanics of Claims,"* 9 J. MORAL PHIL. 545, 569 n.50 (2012) (citation omitted).

¹⁴⁸ Kamm does recognize, in the context of another argument, the possibility that the driver might be thought to have a special "duty . . . to drive . . . in the best possible way" (p. 35); but for the most part, this possibility is set aside throughout the book.

¹⁴⁹ As Thomson puts it, "we should prescind from the possibility that the agents in the cases we are considering have special duties towards the other parties." Thomson, *supra* note 25, at 370.

tations or impressions about the law pull at our intuitions even subconsciously, then in such comparisons, all else is not equal.¹⁵⁰

CONCLUSION

The experiments reported here show that information about the law *can* exert an influence over our moral intuitions in the most basic trolley dilemma. Moreover, the content of the law appears able to exert an influence on our moral intuitions even when it is not enforced. The next question, naturally, is *why*?

What are the possible mechanisms of law's halo?¹⁵¹ Do some subjects consciously look to the law as a source of moral guidance? If so, do they adopt its articulated purpose and even its reasoning? Or does the law serve as a sort of tiebreaker for the morally hard cases?¹⁵²

Do some see following the law as just a morally good thing to do — and do some think so even when the law goes unenforced?¹⁵³ Do some see the law as an indicator of social proof, a sign that many others in society hold the consonant moral opinion?¹⁵⁴

Does being informed about different legal regimes amplify different mental processes, such as the cognitive as opposed to the emotional?¹⁵⁵ Does the law's expectation create a framing effect, perhaps altering the baseline by which consequences are measured?

¹⁵⁰ For example, all is not equal if our impressions about what the law might say varies across scenarios — say, if one holds an impression that it might be criminal homicide for the transplant surgeon to plan the sacrifice, but surely not for a trolley driver facing an emergency decision.

¹⁵¹ This evocative term is credited to Professor Donald Regan. See Donald H. Regan, *Law's Halo*, in *PHILOSOPHY AND LAW* 15 (Jules Coleman & Ellen Frankel Paul eds., 1987).

¹⁵² Or one might say, do some respect the “settlement function” of the law and yield to its choice? See Larry Alexander & Frederick Schauer, *Law's Limited Domain Confronts Morality's Unlimited Empire*, 48 WM. & MARY L. REV. 1579 (2007) (explaining law's settlement function for morally contested questions about which practical decisions need an answer); cf. Larry Alexander & Frederick Schauer, *On Extrajudicial Constitutional Interpretation*, 110 HARV. L. REV. 1359 (1997) (arguing that judicial supremacy provides such a settlement function).

¹⁵³ As Schauer reminds us, “[a]s far back as Socrates and his insistence on acknowledging his obligation to the law even as he believed it to have condemned him unjustly, philosophers and ordinary people have argued that there is a content-independent moral obligation to obey the law.” SCHAUER, *supra* note 11, at 55 (footnote omitted) (reviewing theories of obligation due to social contract, fairness and reciprocity, consent, coordination, and respect for the demos and the lawmaking process). But he also reminds us of the need for better empirical evidence about this possibility, as distinct from evidence about behavior that is consistent with the law or about compliance that is motivated by sanctions. *Id.* at 148–49 (“[T]hat law actually does activate in people what they understand to be a moral reason to obey the law turns out to be a claim with far less empirical support than the conventional wisdom in both jurisprudence and the literature on norms has typically assumed.”).

¹⁵⁴ See Robert Cialdini & Linda Demaine, *The Expanding, Lop-Sided Universe of Social Influence and Law Research*, in *THE OXFORD HANDBOOK OF SOCIAL INFLUENCE* (Stephen G. Harkins et al. eds., 2016).

¹⁵⁵ See generally Cushman & Young, *supra* note 54; Greene, *supra* note 54; Sunstein, *supra* note 4.

Assuming that a variety of such mechanisms might well coexist and might also vary among individuals: Which are more common, and which tend to be stronger? And how can these mechanisms be suppressed, or strengthened?

This study is not designed to answer these questions; its aim has been to present an initial round of findings that might motivate such further inquiry. Nonetheless, the data do suggest certain inferences. First, the law-without-liability effects in the second experiment suggest that law's influence does not work solely through a consideration of the liability costs that the actor faces due to sanctions. At the same time, because the effects of law *with* liability are more pronounced, the threat of sanctions appears to add to law's influence.¹⁵⁶ Beyond these clues, the potential pathways of influence must remain for now among the trolley problem's mysteries.

¹⁵⁶ Note that the moral weight of liability costs for the actor is not the only possible explanation for the extra influence of law-with-liability. Some subjects might view the fact of nonenforcement in the experimental scenario as a sign that the law is a "less serious" law; or, others might even view it as a declaration by the enforcer that the action is justified (even if it violates the law on the books). These possibilities mean that the observed influence of law-without-liability may be understated, relative to an idealized condition conveying that "the law is indeed serious, and nobody is saying that the act is justified or a low enforcement priority; rather, it just so happens that there will be no legal costs for this actor." (Perhaps a future experiment might be able to approximate such a condition — or more generally, to unbundle enforcement, liability, and sanctions.)

Experiment 3

Does criminal liability exert more influence on our moral intuitions than civil liability? Does a bureaucratic law, turning on a technicality unrelated to the moral dilemma, exert less influence? Experiment 3 explores whether the degree of influence depends on how much the law itself aims to provide moral guidance.

This line of inquiry, left open by the earlier experiments,¹ can offer insight into how people's moral intuitions interact with a law's moralizing ambitions. It can also reveal clues about the psychological pathways of influence, as certain comparisons can be understood as suppressing a key characteristic—a legal technicality doesn't purport to offer any moral wisdom, for instance—to see if the law's influence nonetheless survives. An array of possible mechanisms, as well as my assumptions about how specific comparisons might help to sort among them, are detailed below.

Experiment 3 also replicates the earlier experiments while altering a few details. Most notably, now the harm on both sides of the tradeoff is that the workers will be “seriously injured” rather than “killed.”² Otherwise, the scenario is the same as in Experiment 2. The phrasing of the law conditions has also been adjusted to allow a somewhat more critical stance by the subject toward the law.³

The scenario now reads:

Michael is a railroad engineer. One day, while he is working near the train tracks, he notices a freight train approaching. The train seems out of control. Michael can see that the driver is slumped over, unconscious.

There are two workmen on the tracks ahead. The train is now rushing towards them. They don't see the train coming, and Michael can't get their attention. He yells at them, but the construction noise is too loud. He waves his arms, but they are looking the other way.

¹ The prior experiments did not include a non-moralistic law like the bureaucratic condition here. They also did not allow the civil-criminal comparison because they did not include both civil and criminal versions of the same legal command.

² This change is meant to allow a touch more generality by going beyond the classic killing scenario. It may also make the scenario more easily imaginable to subjects, assuming that it's less artificial to hear that the workers will be seriously injured with certainty, than to hear that they will be killed with certainty.

³ Each condition now begins with “There is a law saying that...” rather than the more definitive-sounding “Under the law...” or “The law requires...” as in the earlier experiments. The new articulations also avoid the words “prohibit” and “require,” which mirror two of the deontological choices in the moral judgment question.

Michael happens to be standing near a railroad switch. He can reach it easily, and he knows that pulling the switch will turn the train onto a side track before it hits the two workmen.

However, there is another workman on this side track. He also doesn't see or hear the train coming, and Michael can't get his attention either.

Michael can choose to pull the switch, or he can choose not to do anything.

If Michael pulls the switch, he knows that the man on the side track will be seriously injured by the train.

If Michael doesn't do anything, he knows that the two men on the main track will be seriously injured by the train.

The Experimental Conditions

The five randomized law conditions included in Experiment 3 are: criminal assault and battery for harming someone by pulling the switch, or {Criminal harm}; civil liability for the same reason, or {Civil harm}; criminal negligence for failing to try to reduce casualties, or {Criminal duty}; civil liability for the same reason, or {Civil duty}; and civil liability for pulling the switch without prior authorization, or {Bureaucratic}.

The exact phrasing of each condition is as follows. As in prior experiments, a single randomized condition is appended to the end of the scenario (without the title shown in brackets here).

{Criminal harm}

There is a law saying that Michael (who works for the railroad) must not cause harm to anyone on the tracks. This means that if he pulls the switch, he will be held liable for criminal assault and battery. If he doesn't do anything, he won't be held liable.

{Civil harm}

There is a law saying that Michael (who works for the railroad) must not cause harm to anyone on the tracks. This means that if he pulls the switch, he will be held liable for breaking the law. If he doesn't do anything, he won't be held liable.

{Bureaucratic}

There is a law saying that Michael (who works for the railroad) must not change the path of a train without prior authorization. This means that if he pulls the switch, he will be held liable for breaking the law. If he doesn't do anything, he won't be held liable.

{Civil duty}

There is a law saying that Michael (who works for the railroad) must try to reduce casualties from accidents. This means that if he does nothing, he will be held liable for breaking the law. If he pulls the switch, he won't be held liable.

{Criminal duty}

There is a law saying that Michael (who works for the railroad) must try to reduce casualties from accidents. This means that if he does nothing, he will be held liable for criminal negligence. If he pulls the switch, he won't be held liable.

Thus the {Criminal harm}, {Civil harm}, and {Bureaucratic} conditions are legal prohibitions against pulling the switch, and the {Civil duty} and {Criminal duty} conditions are legal requirements to pull the switch. This group of treatments affords a fuller set of comparisons than Experiment 2, in which there was only a criminal harm prohibition (but no criminal duty) and a civil duty (but no civil harm prohibition). In particular, the new setup allows comparisons of {Criminal harm} against {Criminal duty} and of {Civil harm} against {Civil duty}.⁴

A further condition included in Experiment 3 is the train scenario on its own, without any mention of the law. The subjects' moral judgments in this {No mention} condition can be understood as reflecting the background impressions about the law (however faint, vague, or subconscious) they may already be holding even when not told anything. The differences between the moral judgments in this condition and those in the various law conditions can be simply interpreted as the impact of telling subjects what the law says, relative to not telling them anything.

Distinguishing the Mechanisms

Comparisons among these law conditions can offer insights about the psychological mechanisms of law's influence, assuming the mechanisms differ in strength among the conditions. What follows is a rough-and-ready articulation of possible mechanisms, divided into groups based on my own assumptions about how they

⁴ I did not include a bureaucratic duty to set against the bureaucratic prohibition because I did not think of a phrasing that the subjects would likely find plausible. One might imagine, for instance, that the central character Michael was already tasked with pulling the switch because the side track is actually the correct path for the train; however, this would sound odd in the context of a train running out of control and might introduce a mystery about why the "correct" track still has a worker obliviously standing on it.

might differ in strength across conditions.⁵ (Your suggestions of conceptual refinements, better labels, missing mechanisms, or sharper distinctions would be most welcome!)

1. *Moral Wisdom.* — The first group we might call the “moral wisdom” pathway, for short. These should only be active when the content of the law has to do with the core concern of the dilemma, namely, harm:

* *Law as direct moral guidance*

* *Law as supplier of morally relevant reasons*

* *Law as social proof or as a proxy for societal norms*

2. *Better Obey.* — A second group, which we might call the “better obey” pathway, are those that operate because the law is the law, even if its content does not address the dilemma or the question of harm:

* *Moral duty to obey the law*

* *Moral weight of sanctions and other consequences of liability*

* *Law as coordination device or as role definition*

3. *Arbitrary Anchoring.* — A third possibility is that the law condition does not have any special status as law, but merely breaks the tie as a coin flip might, or merely primes the subject to give the answer that sounds similar. We might call this “arbitrary anchoring”:

* *Law as priming or arbitrary anchor*

To see how comparisons across conditions might suggest which pathways are more important, consider this pair of predictions and the assumptions underlying them. First, let’s assume that the {Bureaucratic} prohibition is unlikely to be viewed by subjects as providing any moral wisdom about the dilemma, in contrast to the {Criminal harm} and {Civil harm} prohibitions.⁶ On this assumption, if the primary mechanisms of

⁵ It is worth emphasizing upfront that individual people likely vary in their responsiveness to the various mechanisms. Moreover, the subjects who are near one margin (say, those torn between saying “morally prohibited” and “morally permitted”) may respond differently than those at the other (those torn between saying “morally permitted” and “morally required”). These experiments only measure aggregate net effects, and only at those two margins.

⁶ It is possible that some subjects may yet imagine that even a legal technicality must have some unstated rationale relevant to the moral dilemma. If so, this sort of projection or rationalization would be an intriguing phenomenon to explore further.

law's influence are those in the moral-wisdom group, then the {Bureaucratic} condition should show less influence than {Criminal harm} and {Civil harm}, or possibly none at all. And if the {Bureaucratic} condition does show some influence, that would be a sign that other pathways are at work.

Second, let's assume that a criminal law, compared to its civil parallel, more vigorously activates the moral-wisdom pathway because criminal law generally is thought to both reflect and shape societal norms. If the mechanisms in this pathway are the primary ones, then we should expect to observe greater effects for {Criminal harm} than for {Civil harm}, and likewise for {Criminal duty} than for {Civil duty}. But if the criminal and civil counterparts show similar degrees of influence, then on the given assumption there would be reason to doubt the dominance of the moral-wisdom pathway.⁷

Results

There are two approaches to reading the results shown in Table 1 and Figures 1A and 1B. The first offers a way to see whether varying the content of the law makes a difference, while the second offers a way to see if supplying specific information makes a difference, given the subjects' background impressions about the law.

1. Content of the Law. — The first approach is to compare moral judgments among law conditions. As in Experiments 1 and 2, the observed differences can be interpreted as evidence that informing people about different legal commands results in different distributions of moral intuitions—in other words, what the law says, matters. Of special interest here is seeing whether varying the content of the criminal law matters more than varying the content of civil law. In the figures, to aid such comparisons, the criminal conditions are shaded differently from the civil conditions.⁸

There is about a 7-point gap between the {Criminal harm} and {Criminal duty} conditions in the share saying “morally prohibited.” On the civil side, there is about a 14-point gap between {Civil harm} and {Civil duty}; and about a 13-point gap between

⁷ This would not necessarily mean that the moral-wisdom pathway is inactive; it would also be possible that the other pathways are so highly active for both the criminal and civil conditions, that the moral-wisdom mechanisms (though also active) contribute no further influence. An entirely alternative explanation would be that the underlying assumption is simply wrong—that a criminal law does not activate the moral-wisdom pathway more powerfully than its civil counterpart.

⁸ In the table and in the figures, I have ordered the conditions based on what I had originally expected to be the most influential condition against turning to the most influential condition in favor of turning. As is evident, some of the results were rather unexpected.

{Bureaucratic} and {Civil duty}.⁹ (Reported differences are statistically significant unless otherwise noted.) The straightforward interpretation is that, whether the form of liability is criminal or civil, for some subjects their willingness to say that pulling the switch is morally prohibited is influenced by what the law says.

As for the share saying “morally required,” there is about a 7-point gap between {Criminal harm} and {Criminal duty}, which is not statistically significant. On the civil side, there is about a 20-point gap between {Civil harm} and {Civil duty}, and about a 17-point gap between {Bureaucratic} and {Civil duty}. The straightforward interpretation is that under civil liability, quite a few subjects’ willingness to say that pulling the switch is morally required is influenced by what the law says. The notably smaller gap between the opposite criminal conditions is considered in more detail below.

2. *Telling Them the Law.* — The second approach is to compare each law condition with the {No mention} condition. These differences in moral judgments can be interpreted as the impact of informing subjects about that specific legal command, relative to a baseline of whatever background impressions they may hold about the law when not told anything. In the figures, the {No Mention} response is shown as the dashed horizontal line.

From this perspective, for the share saying “morally prohibited,” the {Criminal harm} condition shows no significant effect; {Civil harm} and {Bureaucratic} show positive effects; and {Civil duty} and {Criminal duty} show negative effects. For the share saying “morally required,” only the {Civil duty} condition shows a significant effect; curiously, its counterpart {Criminal duty} shows little if any influence on this measure, despite showing influence on the other.

The directions of the observed effects are as expected, given what Experiments 1 and 2 showed. What may be unexpected is the weakness of the {Criminal harm} condition’s influence, if any.¹⁰ This puzzle is examined more closely in the following section. But it may be useful to draw one initial inference from this observation: the stronger positive effects of the counterpart conditions, {Civil harm} and {Bureaucratic}, are not likely due to priming or arbitrary anchoring. After all, if telling subjects about the law primes them to choose the aligned moral judgment (or otherwise arbitrarily anchors

⁹ There is no bureaucratic duty condition, and hence the most sensible comparator is the civil duty condition (because the bureaucratic law is not criminal but civil).

¹⁰ The standard caveat applies, about declaring non-effects (even when the point estimates are virtually the same, as is the case here), due to the margin of error.

their moral answers), then the {Criminal harm} condition should also have shown a similar positive effect on the “morally prohibited” answer.¹¹

TABLE 1. ENGINEER SAVES TWO (INJURIES)

	Morally prohibited	Morally permissible	Morally required	<i>N</i>
<i>Law prohibits turning</i>				
Criminal harm	12.9%	63.2%	23.9%	209
Civil harm	18.8%	56.8%	24.4%	176
Bureaucratic	17.5%	55.2%	27.3%	194
<i>Law requires turning</i>				
Civil duty	4.1%	51.8%	44.1%	195
Criminal duty	5.8%	62.8%	31.4%	191
<i>No mention of law</i>				
No mention	12.9%	59.1%	28.1%	171

¹¹ Similarly, the weakness of the {Criminal duty} condition’s influence (if any) on the share saying “morally required” may also be unexpected—given its evident influence on the share saying “morally prohibited,” and given the notable influence of {Civil duty} on both margins. This differential, too, offers evidence against the priming or arbitrary anchoring mechanism, which should have generated a positive influence on the share saying “morally required” in both the {Criminal duty} and {Civil duty} conditions, and not only in the latter.

FIGURE 1A

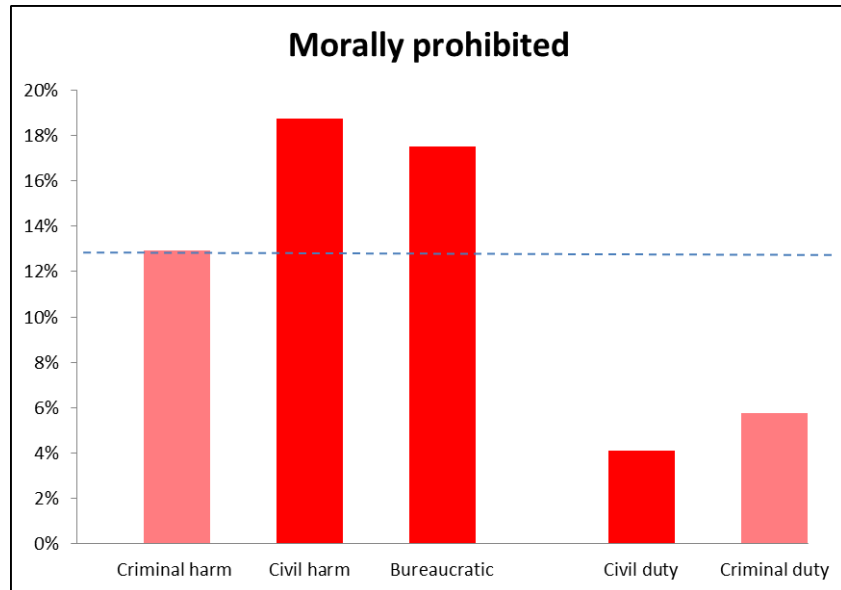
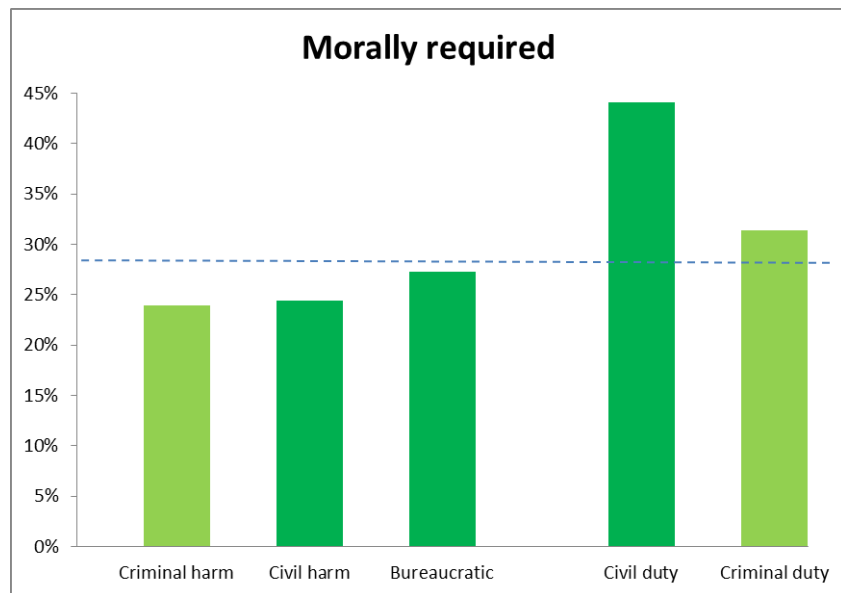


FIGURE 1B



Note: The dashed lines represent the {No Mention} condition.

Discussion: Two Unexpected Findings

Two possibly surprising findings are worth exploring in more depth. First, the {Bureaucratic} condition showed about as strong an influence as {Civil harm}. Second, the {Criminal harm} condition showed a weaker influence, if any, relative to both its {Civil harm} and {Bureaucratic} counterparts.

1. Bureaucratic Law. — I had expected the {Bureaucratic} condition to show little or no influence, because it does not even purport to offer any moral guidance about harm, much less about the dilemma. But to the contrary, the {Bureaucratic} condition does influence the “morally prohibited” answer as much as the {Civil harm} condition does, and more than the {Criminal harm} condition.¹² This suggests that informing subjects about the law can affect moral intuitions even when the moral-wisdom pathway is inactive.

The influence of the {Bureaucratic} condition is thus likely working through the better-obey pathway.¹³ For instance, some subjects may defer to such a law as defining the actor’s role within the system, or as a coordination device; such deference might seem especially sensible in the unfamiliar context of a railroad engineer’s decision. Or, some subjects may count the threat of sanctions or other collateral consequences as a proper part of the actor’s own moral calculus (maybe thinking of the harm to his family should he lose his job). Or, some subjects may feel it is moral to obey the law, even when the law is based on a technicality; notably, for some this technicality (of not having prior authorization to pull the switch) may also engage the moral value of respect for authority or hierarchy,¹⁴ thus bringing in an orthogonal moral factor (about authority/hierarchy) to break a moral tie (about harm).

2. Criminal Prohibition. — I had also expected the expressly moralistic nature of the criminal law to enhance the law’s influence. To the contrary, the {Criminal harm} condition shows little difference if any relative to the {No mention} condition, and shows a lower share saying “morally prohibited” than in the {Civil harm} and {Bureaucratic} conditions. This contrast does not square with any of the ex ante predictions noted above, in which a criminal condition might have more or the same, but not less, influence relative to its civil counterpart.

¹² Here, I am referring to the comparisons with the {No Mention} condition.

¹³ Again, priming or arbitrary anchoring is unlikely to be the explanation, given that the {Criminal harm} condition has little or no effect. Further evidence against this type of mechanism is offered below.

¹⁴ To be clear, here I mean the authority who did or didn’t authorize pulling the switch (not the authority of the law in a more generic sense, or the authority who made the law).

Trying to rationalize this unexpected result may generate hypotheses for future inquiry. One possibility is that {Criminal harm} and {No mention} are coming out about the same because {Criminal harm} is also what subjects would have expected when they aren't told anything about the law. Yet if that were the case, one should not expect the {Civil harm} condition to have a positive effect relative to {No mention}, as it does.¹⁵

Maybe criminalization induces some sort of reactance, resentment, or backlash. Yet a blanket aversion to criminalization is an unlikely explanation, given that the {Criminal & liable} condition in Experiment 2—stating that killing the worker counts as criminal homicide—did show a positive influence on the “morally prohibited” answer.¹⁶

Rather, something about telling subjects that seriously injuring the worker on the side track will result in liability for criminal assault and battery seems to have dampened the responsiveness of their moral intuitions. Could it be that this description of liability seemed especially disproportionate or illegitimate?¹⁷ For example, did more subjects think it morally unimportant to obey such a harsh criminal law? Did more subjects discount the moral wisdom (or social proof) to be gained from it? Did they not actually expect such a law to be enforced? These possibilities are entirely speculative. But as noted below, one of the follow-up questions included in the survey may offer some support—showing that 63 percent of subjects described this {Criminal harm} condition as “unfair.”

Discussion: Distinct Moral Margins

Another noteworthy pattern that has emerged is the difference in responsiveness at the “morally prohibited” and “morally required” margins.¹⁸ This can be seen in both Experiments 2 and 3. The “morally prohibited” margin appears potentially responsive to legal commands in both directions. Yet, although a legal duty can increase the share

¹⁵ In theory it is possible that subjects hold a general background impression that pulling the switch would be criminal assault and battery, while not also holding any such impression about civil liability; but this seems a stretch.

¹⁶ Magnitudes should not be directly compared between Experiments 2 and 3, because the former uses a scenario involving killing, and the latter uses a scenario involving serious injuries. Thus, in theory it is possible that there is no conflict between what looks like a sizeable effect in Experiment 2's killing scenario and little influence in Experiment 3's injuring scenario.

¹⁷ To put this in terms of the contrast with the {Criminal & liable} condition in Experiment 2: Does the {Criminal harm} condition here seem especially ill-fitting because the term “assault and battery” implies a malicious intent, whereas the earlier term “homicide” does not, to a general reader?

¹⁸ By margins I mean the boundary between “morally prohibited” and “morally permitted,” and the boundary between “morally permitted” and “morally required.” It may be helpful to think of two groups of subjects: those who are torn between saying “morally prohibited” and saying “morally permitted,” and those torn between “morally permitted” and “morally required.”

saying “morally required,” these experiments have not shown that a legal prohibition can reduce it. The straightforward interpretation is that those subjects already saying “morally required” (even when not told of a legal duty) are unmoved by hearing that the law prohibits saving the two on the main track.¹⁹

This asymmetric pattern may also be instructive about mechanisms. First, the pattern shows that informing subjects about the law does not simply shift the entire distribution of moral intuitions in one direction or the other—in which case one would expect to see changes at both measured margins, not the observed immobility at one of them. Second, this asymmetric pattern offers further evidence against priming or arbitrary anchoring. Such a mechanism might induce subjects to choose the aligned moral judgment—saying “morally required” just because the extra information also says it’s required—but it is unlikely to account for the asymmetry in observed influence of the laws on the opposite moral judgments.²⁰

Rather, what is generating the asymmetry—with some subjects thinking that “it can’t be morally prohibited if it’s legally required” but others rejecting that “it can’t be morally required if it’s legally prohibited”—would seem to be a mechanism that involves some discernment about how much moral weight to give to unwelcome legal commands.

Discussion: Further Hints?

Experiment 3 also included several follow-up questions that may allow us to pick up some clues about the nature of the observed influence.²¹

1. Stickiness. — One follow-up question asked what the subject would have answered, had the story not said anything at all about the law. This is meant to measure whether subjects can consciously undo the law’s influence when expressly asked to do so

¹⁹ This is not saying that consequentialists are resolute in some general sense. The data are only informative about subjects whose views are close to the two measured margins.

²⁰ That is, it cannot account for the duty conditions affecting the “morally prohibited” margin, but the prohibition conditions not noticeably affecting the “morally required” margin. For one thing, it’s not obvious that a priming or arbitrary anchoring mechanism should much affect the opposite moral margin. And even it could (say, by shifting the entire distribution), it should do so for law conditions going in both directions. And this is inconsistent with the observations. All told, given this asymmetry, and considering also the lack of influence of the {Criminal harm} condition noted above, such an automatic mechanism seems unlikely to explain the data.

²¹ It should be emphasized that the answers to any such follow-up questions should be interpreted as endogenous to the scenario and to the law condition, both of which the subject has already encountered by that point in the survey.

(one might say, to “forget” the law) or whether instead the law’s influence is subconsciously sticky.²²

Stickiness should vary among the different law conditions, depending both on the salience of the law information and on which mechanisms are activated. Because undoing the law’s influence requires self-awareness of it, stickiness might suggest that the influence is of a more subtle form. For example, the subject may have subconsciously adopted the law’s reason as her own, or she may have been nudged by the social proof that the law seemingly provides.

In contrast, a bouncy response would suggest that the law’s content was more salient to the subject and that the mechanism of influence was more consciously available or recognizable to her. For example, the subject may feel that she had responded originally based on the moral weight she gives to obeying the law or to the legal consequences for the actor, perhaps because the law was opposed to her initial instinct.

The difficulty of making such distinctions in the data is statistical power; the present sample is too small to say much with confidence. The few detectable instances of stickiness or bounciness do suggest a possible difference between the laws requiring turning and the laws prohibiting turning. The former tend to have a stickier influence, with subjects able to only partially undo the influence of {Civil duty} and {Criminal duty}. That is, on average, subjects underestimated how much influence these laws had, on their judgments. By contrast, there was more bounciness for the legal prohibitions—with even some overshooting in the {Bureaucratic} condition—suggesting that the law was more salient and the mechanism a more conscious one.

2. *Unfairness.* — Another follow-up question asked subjects how they felt about the law information they were given, including whether “the law was unfair in this situation.” Here, a large gap appears between the law conditions that prohibit versus require turning the train. Far fewer say that the laws requiring turning are unfair: 25 percent for {Civil duty} and 25 percent for {Criminal duty}. Many more say that the laws prohibiting turning are unfair: 63 percent for {Criminal harm}, 53 percent for {Civil harm}, and 48 percent for {Bureaucratic}.

This difference seems consistent with the differential stickiness noted above, if laws perceived as unfair are also more salient and exert influence mainly through the more conscious mechanisms, whereas laws not seen as unfair have a better chance of

²² The {No mention} condition supplies the “true” distribution of what subjects would have said, if the story had said nothing about the law.

operating through more subtle and less conscious mechanisms. The especially strong reaction against {Criminal harm} as being unfair may also offer some support for the speculative reasons noted above for why its influence is diminished relative to its civil counterparts, although a considerable (if lower) share of subjects also deemed those civil prohibitions to be unfair.

What may also be notable is that the share saying that any given law is unfair exceeds the share who would choose the opposite moral judgment if not told about the law. To speak in terms of the point estimates, only 13 percent in the {No mention} condition say that turning the train is “morally prohibited,” and yet 25 percent say that a law requiring it is unfair. Meanwhile, only 28 percent say that turning the train is “morally required,” and yet 48 to 63 percent say that a law prohibiting it is unfair. These responses suggest that there may be some general reactance against legal interference—about the very presence of law’s command (with no indication of relief via justification, excuse, or non-enforcement) in the context of a difficult moral dilemma.

Tentative Summary

Altogether, given my initial interpretive assumptions, the results seem to support these inferences about the pathways of influence: The better-obey pathway is an important one, as seen in the influence of the {Bureaucratic} condition.²³ Meanwhile, the moral-wisdom pathway may or may not be at work.²⁴ If it is, its influence seems more likely for the legal duty conditions, given the clues from the stickiness and unfairness questions; also, its strength is not necessarily enhanced by the moralizing ambitions of criminal law or criminal punishment. Finally, it is safe to say that the arbitrary-anchoring pathway does not help to explain the data.

Several limitations are worth noting. First, it should go without saying that individuals surely vary in their responsiveness to law information, and the operative mechanisms probably vary among them too. These aggregate data allow inferences only about patterns of influence, and thus only about the macro-level prevalence of the various pathways. Second, the weak effect of the {Criminal harm} condition lacks an explanation that fits squarely with the experiment’s ex ante predictions, although some speculative ex post possibilities may be imagined. Third, it may be that the initial

²³ Recall that this pathway is shorthand for a group of mechanisms including: feeling a moral duty to obey the law; giving moral weight of sanctions and other consequences of liability; and seeing law as coordination device or as role definition.

²⁴ Recall that this pathway is shorthand for a group of mechanisms including: law as direct moral guidance; law as supplier of morally relevant reasons; and law as social proof or as a proxy for societal norms.

interpretive assumptions are unsound. For example, if the {Bureaucratic} condition is understood by many subjects as being relevant to the harm-based concerns in the moral dilemma because it instructs against creating a further risk of harm, then the interpretation above should be revised, with some credit shifted from the better-obey to the moral-wisdom pathway.

* * *

Experiment 4

I am now designing Experiment 4 with the aim of learning more about the possible mechanisms of influence. As of now, the leading candidates for new law conditions to include are listed on the pages that follow. Your suggestions about which of these might be more useful would be most welcome! I can't use all of them, as the total sample size would be prohibitive. I also welcome your suggestions about how to improve the phrasing of them, or about other conditions or questions to include.

New Outcome Measure

In Experiment 4, I intend to use again the Engineer Saves Two scenario. But I am considering using a different metric for moral judgment. The earlier experiments have asked subject to choose among three options by saying whether pulling the switch is morally prohibited, morally permitted, or morally required. For Experiment 4, an alternative would be to ask, "Which would you say is the morally better decision?"—with the two choices being pulling the switch or not doing anything.

Posing the question this way alters the nature of the judgment: whereas the original three-option format maps onto standard deontological categories (prohibited, permitted, or required), the two-option format invites a moral balancing, in order to say which action is "morally better."²⁵ Subjects may think about degrees of overall moral goodness differently from how they think about what is categorically wrong.²⁶

The new binary measure also directly changes the margin of observation within the population. With the three-options format, the most popular category is the middle category of "morally permitted," which is not surprising if the middle option is thought to serve as a default option; thus the analyses so far have focused on how many subjects were decidedly willing to say that pulling the switch is "morally prohibited" or "morally required." With the two-options format, the new margin of observation is more centrally located, so to speak. If it is possible to map the new measure onto the old, the new boundary probably falls somewhere within the "morally permitted" band.²⁷

²⁵ It thus should be distinguished from other possible binaries, such as asking whether a choice is morally "wrong," which is again more obviously deontological.

²⁶ It is also possible that asking for a moral balancing may lean subjects toward consequentialism, whereas the earlier phrasings might have leaned subjects toward deontological thinking.

²⁷ As mentioned, the nature of moral judgments being called for by these two metrics is different. But it seems intuitive that a subject categorizing a choice as morally required would tend to also say that it is morally better, and

Secondary Measures

In addition to changing the primary outcome measure, I may also include these secondary outcome measures.

1. *Moral Certainty*. — The first is to ask, “How sure do you feel, about your answer?”²⁸ This question can be useful in two ways: First, it may allow some insight into inframarginal effects; that is, for those subjects who do not switch their answer, does the law nonetheless have some influence on their intuitions?²⁹ Second, it may also detect reactance or backlash against an unwelcome law—for instance, if subjects confronted with a law condition express *more* certainty in saying that the *opposite* action is morally better.³⁰

2. *Support for Law*. — Second, I may include a question asking subjects which kind of law they would be most likely to support, among a menu of choices that correspond with the various law conditions (with one of the choices being “there should be no laws about a situation like this”). This question may allow an initial sense of the potential for a feedback loop in which what subjects are told about the law influences their moral intuitions, thereby boosting their support for that same law. It might also allow some detection of reactance or backlash, for example, if subjects under a given law condition more enthusiastically supported the “no laws” option in the menu.

Basic Conditions

Experiment 4 will again seek to replicate the earlier experiments while altering a few things. The most notable change would be the new primary outcome measure. In addition, I intend to make the following changes: (1) The {Bureaucratic} condition will be more generic in its command: “must not operate any equipment without authorization.” This change addresses the possible concern that the bureaucratic

one categorizing a choice as morally prohibited would tend to also say that the other choice is morally better. And probably some of those who would have said that pulling the switch is permitted would also say that it is morally better, while others among them would say that doing nothing is better.

²⁸ The options would be “Not sure at all,” “Somewhat sure,” and “Very sure.”

²⁹ For example, if a greater share (of all subjects) were to say they are “very sure” that pulling the switch is morally better, when told of a law requiring it, this would be suggestive of the law’s influence even among those who would have said that pulling the switch is morally better even without hearing about this law.

³⁰ For example, if a greater share (of all subjects) were to say they are “very sure” that pulling the switch is morally better, when told of a law prohibiting it, there would seem to be no other explanation than some sort of reactance or backlash. Note that the denominator choice is important, here; if a greater share (of those saying pulling the switch is morally better) were to say that they are “very sure,” that could just be because the less-sure people on that side have switched over to saying that doing nothing is morally better.

condition in Experiment 3 refers to changing the path of a train, making it too easy for subjects to imagine that it is a law concerned with harm or even more specifically with a class of situations like the given dilemma.³¹ (2) The description of liability will be more specific about the legal consequences (“liable for the resulting injuries”).³² At the same time, the phrasing of the non-labile choice will also make it clearer that no other legal liability exists (“won’t face any liability”). (3) The phrasing will be more consistent in its treatment of inaction, in describing both the law conditions and the actor’s choices.

Thus the basic conditions will read:

{Civil harm}

There is a law saying that Michael (who works for the railroad) must not do anything to cause harm to anyone on the tracks. This means that if he pulls the switch, he will be held liable for the resulting injuries. If he doesn’t do anything, he won’t face any liability.

{Bureaucratic}

There is a law saying that Michael (who works for the railroad) must not operate any equipment without authorization. This means that if he pulls the switch, he will be held liable for the resulting injuries. If he doesn’t do anything, he won’t face any liability.

{Civil duty}

There is a law saying that Michael (who works for the railroad) must not try to reduce casualties from accidents. This means that if he doesn’t do anything, he will be held liable for the resulting injuries. If pulls the switch, he won’t face any liability.

³¹ The difficulty of using a more arbitrary law condition, like “must not operate any equipment on a Tuesday,” is that it needs still to be believable to the subject that such a law could exist. It also seems to me that a subject who can imagine a dilemma-relevant basis for the law proposed above could also imagine it for this one.

³² This change replaces the possibly moralizing tone of the phrasing in the earlier experiments (“liable for breaking the law”) with a drier and more precise statement of the degree of liability; the flip side is that it may then draw attention to the financial or collateral consequences for the actor. Downplaying the specific extent of liability in earlier experiments had the advantage of not drawing attention to the asymmetry that not doing anything would result in liability for two persons’ injuries or deaths, whereas pulling the switch would result in liability for only one person’s. But that asymmetry is not a problem as long as one avoids directly comparing the magnitudes of the effects of legal prohibition versus legal requirement (a comparison that should be avoided for other good reasons, too, such as the lack of any reason to assume that, say, a 10 percent move in one direction on a given moral judgment margin is somehow equivalent to a 10 percent move in the other direction).

Candidates for Experimental Conditions

What follows on the next few pages are the leading candidates for possible further conditions to include in Experiment 4. Again, your suggestions about which of them might be most interesting or useful would be much appreciated!

A. “Social Norms” Conditions

This set of conditions is designed to test whether the law’s influence works partly by serving as a proxy for other people’s or for society’s moral judgments (let’s call it “social norms”). In short, the test works by seeing if telling people about the law adds any influence beyond just telling people what the norm is.

The first condition supplies only the norm itself: a roughly true statement that four out of five people would say that pulling the switch is the morally better decision.³³ This tests whether the norm itself, directly told to the subject, has some degree of influence. Note that this is quite a strong norm.

The other conditions add information about laws that oppose or align with this norm. Let’s begin with the aligned law, which would be {Civil duty}. If the influence of this law works primarily as a proxy for information about social norms, then telling people about it should have no further influence beyond directly telling people about the norm. That is, the {Duty norm} and {Duty norm & civil duty} conditions should both show similar effects to the plain {Civil duty} condition. If either {Civil duty} or {Duty norm & civil duty} show more influence than {Duty norm} on its own, then one might infer that the law is doing some work beyond proxying for the norm.³⁴

As for the opposed laws: If proxying for social norms is an important mechanism for the legal prohibitions, then the norms information should more powerfully diminish the influence of an opposed law based on harm (by ruling out that this law is proxying for social norms) more than that of the opposed bureaucratic law (which isn’t proxying for social norms in the first place). The complication, however, is that there is also an overlay of the further possible effects of pitting a strong social norm against the other possible mechanisms at work. Given the messiness of drawing inferences here, these

³³ The reason I propose using only this duty norm, but not a prohibition norm, is that most subjects will probably have a good feel for what others are likely to think. Thus, telling subjects that “most people would say that not doing anything is the morally better decision” just wouldn’t be credible (as it is quite far from being true). In addition, the sample size required for including norms on both sides would be infeasibly large.

³⁴ In theory it is possible that the law information could be signaling an even stronger norm, but that seems hard to imagine given that the baseline is straight-up telling people that 80 percent of people agree on the moral answer.

opposed conditions, {Duty norm & civil harm} and {Duty norm & bureaucratic}, may not be as useful to include as the {Duty norm & civil duty} condition described above.

Altogether, these conditions might read:

{Duty norm}

Most people (four out of five) would say that pulling the switch is the morally better decision.

{Duty norm & civil duty}

Most people (four out of five) would say that pulling the switch is the morally better decision.

Also, there is a law saying that Michael (who works for the railroad) must try to reduce casualties from accidents. This means that if he doesn't do anything, he will be held liable for the resulting injuries. If he pulls the switch, he won't face any liability.

{Duty norm & civil harm}

Most people (four out of five) would say that pulling the switch is the morally better decision.

Yet there is a law saying that Michael (who works for the railroad) must not do anything to cause harm to anyone on the tracks. This means that if he pulls the switch, he will be held liable for the resulting injuries. If he doesn't do anything, he won't face any liability.

{Duty norm & bureaucratic}

Most people (four out of five) would say that pulling the switch is the morally better decision.

Yet there is a law saying that Michael (who works for the railroad) must not operate any equipment without authorization. This means that if he pulls the switch, he will be held liable for the resulting injuries. If he doesn't do anything, he won't face any liability.

B. “Laws Elsewhere” Conditions

This set of conditions extends a line of inquiry from Experiments 1 and 2 (but was not included in Experiment 3): Does the law have influence even when there are no legal consequences for the actor? In the earlier experiments, the given reason why there would be no legal consequences was that the law would not be enforced (these were called the {Criminal & not liable} and {Duty & not liable} conditions). One possible complication with the prior formulation is that it may suggest that the law was not a serious law to begin with, or that a legal authority figure was making an active choice not to enforce.

A new law-without-liability formulation might be to say that many other places have such laws, but this particular place does not, which is why there won’t be any liability for this actor. This formulation also goes beyond the earlier one in allowing sharper sorting between the moral-wisdom pathway (which should be active, though possibly weakened) and the better-obey pathway (which should be suppressed in all its forms, including the moral-duty-to-obey-the-law mechanism, as well as any hierarchy/authority considerations that might attend the bureaucratic law condition).³⁵

The conditions might read:

{Harm laws elsewhere }

In many other places, there are laws saying that someone like Michael (who works for the railroad) must not do anything to cause harm to anyone on the tracks. This means that if he pulled the switch, he would be held liable for the resulting injuries. If he didn’t do anything, he wouldn’t face any liability.

But where Michael is, there are no relevant laws, and so he doesn’t need to worry about any liability.

{Bureaucratic laws elsewhere}

In many other places, there are laws saying that someone like Michael (who works for the railroad) must not operate any equipment without prior authorization. This means that if he pulled the switch, he would be held liable for the resulting injuries. If he didn’t do anything, he wouldn’t face any liability.

But where Michael is, there are no relevant laws, and so he doesn’t need to worry about any liability.

³⁵ By contrast, in the earlier non-enforcement conditions, some of the better-obey mechanisms may still have been active, even if the law wasn’t enforced; for example, some subjects may feel a moral duty to obey the law even absent enforcement. The flip side is that the new formulation won’t help to sort among the mechanisms within the better-obey pathway.

{Duty laws elsewhere }

In many other places, there are laws saying that someone like Michael (who works for the railroad) must try to reduce casualties from accidents. This means that if he didn't do anything, he would be held liable for the resulting injuries. If he pulled the switch, he wouldn't face any liability.

But where Michael is, there are no relevant laws, and so he doesn't need to worry about any liability.

C. “Law Next Year” Conditions

This set of conditions offers another model of law-without-liability. Here, a newly passed law does not take effect until next year. Thus it follows the spirit of the earlier non-enforcement conditions in Experiments 1 and 2, while still avoiding the possible complication of active non-enforcement. It does introduce, however, some new problems. First, it may be ambiguous whether subjects view this as a binding law (making it hard to say whether the moral-duty-to-obey-the-law mechanism is at work). Second, the law blurb may not follow as naturally from the basic scenario, and it is so detailed as to be possibly jarring to readers.

The conditions might read:

{Harm law next year}

There is a new law saying that someone like Michael (who works for the railroad) must not cause harm to anyone on the tracks. This means that someone who pulls the switch will be held liable for the resulting injuries, and someone who doesn't do anything won't be held liable.

But because this law only goes into effect next year, Michael himself doesn't need to worry about any liability.

{Bureaucratic law next year}

There is a new law saying that someone like Michael (who works for the railroad) must not operate any equipment without prior authorization. This means that someone who pulls the switch will be held liable for the resulting injuries, and someone who doesn't do anything won't be held liable.

But because this law only goes into effect next year, Michael himself doesn't need to worry about any liability.

{Duty law next year}

There is a new law saying that someone like Michael (who works for the railroad) must try to reduce casualties from accidents. This means that someone who doesn't do anything will be held liable for the resulting injuries, and someone who pulls the switch won't be held liable.

But because this law only goes into effect next year, Michael himself doesn't need to worry about any liability.

* * *