2010

Some Realism About Punishment Naturalism

Dan M. Kahan  
Yale Law School

Donald Braman

David A. Hoffman

Follow this and additional works at: https://digitalcommons.law.yale.edu/fss_papers

Part of the Law Commons

Recommended Citation
Kahan, Dan M.; Braman, Donald; and Hoffman, David A., "Some Realism About Punishment Naturalism" (2010). Faculty Scholarship Series. 4690.  
https://digitalcommons.law.yale.edu/fss_papers/4690

This Article is brought to you for free and open access by the Yale Law School Faculty Scholarship at Yale Law School Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship Series by an authorized administrator of Yale Law School Legal Scholarship Repository. For more information, please contact julian.aiken@yale.edu.
In this Article, we critique the increasingly prominent claims of Punishment Naturalism—the notion that highly nuanced intuitions about most forms of crime and punishment are broadly shared, and that this agreement is best explained by a particular form of evolutionary psychology. While the core claims of Punishment Naturalism are deeply attractive and intuitive, they are contradicted by a broad array of studies and depend on a number of logical missteps. The most obvious shortcoming of Punishment Naturalism is that it ignores empirical research demonstrating deep disagreements over what constitutes a wrongful act and just how wrongful a given act should be deemed to be. But an equally serious shortcoming of Punishment Naturalism is that it fails to provide a credible account of the social and cognitive mechanisms by which individuals evaluate both crime and punishment, opting instead for explanations that are either specific and demonstrably wrong or so vague as to be untestable.

By way of contrast, we describe an alternative approach, Punishment Realism, that develops the core insights of legal realism via psychology and anthropology. Punishment Realism, we argue, offers a more complete account of agreement and disagreement over the criminal law and provides a more detailed and credible account of the social and cognitive mechanisms that move people to either agree or disagree with one another on whether a given act should be praised or punished and how much praise or punishment it deserves. The differences between these two empirical accounts also suggest contrasting implications for how those interested in maximizing social welfare and public satisfaction with the law should approach questions of crime and punishment.

† Associate Professor of Law, The George Washington University Law School.
‡‡ Elizabeth K. Dollard Professor of Law, Yale Law School.
‡ Associate Professor of Law, Temple University Beasley School of Law.

We thank Kenworthy Bilz, Bernard Harcourt, Owen Jones, Sarah Lawsky, Richard McAdams, Janice Nadler, Kitty Richards, Stephanie Stern, and Lior Strahilevitz for their comments and advice.

1531
The sure fatality is to imagine variance not there or wait for it to go away.

—Clifford Geertz

INTRODUCTION

You are at the bus station, bringing the rings to your best friend's wedding, when your wallet and ticket are stolen. No one will lend you money to pay the fare. You notice a well-heeled fellow traveler heading to the restroom, leaving jacket and ticket behind. You think about it. He could afford another ticket. There's no chance you'd be caught. But in the end you can't bring yourself to do it. As painful as it will be to miss the wedding, you just know that stealing the ticket would be wrong.

But how do you know that? How do any of us know right from wrong? Is our morality by and large determinate and innate, the product of evolutionary forces acting over millions of years, or do we acquire it within our lifetimes, reading acts in relation to variable social norms that we have assimilated from those around us? How we answer these questions matters. If humans share highly specific intuitions about justice as a consequence of innate moral mechanisms, then it will be quite difficult, perhaps even impossible, to alter those intuitions, and we should be very cautious if we plan to adopt an approach to punishment that deviates from these innate preferences. If, on the other hand, we develop a sense of morality over our lifetimes in relation to varied social norms, then we might learn how our moral intuitions are shaped and develop means of fostering conceptions of justice that are both satisfying to us and compatible with our collective welfare.

This Article argues that although moral judgments depend on numerous cognitive and physiological mechanisms that are presumably the product of evolutionary pressures, they are not innate insofar as they depend crucially on social meaning that varies across cultural groups. In our opening hypothetical, you (or, rather, our hypothetical version of you) refused to steal the ticket. But not everyone would, as evidenced not only by the hypothetical (though perhaps familiar) theft of your wallet, but also by extensive empirical research that we describe below.

In developing this account, this Article critiques the increasingly prominent claims of Punishment Naturalism—the notion that highly nuanced intuitions about most forms of crime and punishment are

---

1 Clifford Geertz, Local Knowledge: Further Essays in Interpretive Anthropology 219 (Basic Books 1983).
broadly shared because they are innate.\textsuperscript{2} To their credit, Punishment Naturalists marshal an impressive array of empirical research into widely shared human intuitions. Many humans do share broad intuitions that provide them with nearly effortless appraisals of wrongdoing. But if the core claims of Punishment Naturalism are deeply attractive and intuitive, they are not unassailable. There are extensive data showing dissensus over punishment for which naturalism cannot account. There is also a troubling void in naturalism where one would expect a credible account of either the social or cognitive mechanisms by which individuals evaluate crime and punishment.

A fuller and more accurate explanation of human intuitions about wrongdoing is offered by what we call Punishment Realism.\textsuperscript{3} Uniting the insights of legal realists with research conducted by anthropologists, social psychologists, and evolutionary biologists, Punishment Realism is based on the premise that while individuals do hold deep and abiding intuitions regarding wrongdoing and responses to it, these intuitions depend on social constructs that are demonstrably plastic. Thus, while there are a number of important (perhaps even universal) features of human cognition that shape our understandings of wrongdoing, they are features that interact with, and enable the construction of, varied social norms rather than produce them in a determinate manner.

How varied are our norms? If you thought you should refrain from taking the ticket because it was wrong, then you agree with most Americans. On a naturalist account, this makes sense: the “taking of property without consent” is a moral violation, part of the “core of wrongdoing” — something on which nearly all humans normally agree.

\textsuperscript{2} Although this Article is limited to the specific application of naturalism to punishment theory, Punishment Naturalism partakes of a broader trend toward legal analyses drawing on research in the area of evolutionary psychology, much of which avoids the pitfalls we describe herein. See, for example, Owen D. Jones and Sarah F. Brosnan, Law, Biology, and Property: A New Theory of the Endowment Effect, 49 Wm & Mary L Rev 1935, 1953–54 (2008); Rose McDermott, James H. Fowler, and Oleg Smirnov, On the Evolutionary Origin of Prospect Theory Preferences, 70 J Polit 335, 337–38 (2008); Herbert Gintis, The Evolution of Private Property, 64 J Econ Behav & Org 1, 2–3 (2007); Jeffrey E. Stake, The Property "Instinct," 359 Phil Transactions Royal Socy B: Bio Sci 1762, 1767 (2004); Paul H. Rubin, Darwinian Politics: The Evolutionary Origin of Freedom 173 (Rutgers 2002). The naturalism that we describe here is distinct from, and should not be confused with, the philosophical use of the term. See, for example, Keith Campbell, Naturalism, in Donald M. Borchert, ed, 6 Encyclopedia of Philosophy 492, 492 (Thomson 2d ed 2006) (defining naturalism in the philosophical context as representing the proposition that “the natural world is the only real one, and that the human race is not separate from it, but belongs to it as a part”).

\textsuperscript{3} See generally Donald Braman and Dan M. Kahan, Legal Realism as Psychological and Cultural (Not Political) Realism, in Austin Sarat, Lawrence Douglas, and Martha Merrill Umphrey, eds, How Law Knows 93, 112–13 (Stanford 2007).

\textsuperscript{4} For further discussion of these terms, see note 35 and accompanying text.
And no one can blame you for missing the wedding—after all, you have a reasonable excuse for missing it. But researchers who posed the same question in India found that the vast majority of participants there thought that you would be justified in taking the ticket. Where American subjects tended to justify missing the wedding in moral terms that centered on individualized justice and personal property, Indian participants tended to justify the theft in moral terms that emphasized the social and relational responsibilities of friends, particularly at such an important event. When researchers asked tribal leaders in Papua New Guinea how they would resolve a similar scenario, the leaders not only thought that stealing would be justified, but blamed the people who failed to be of assistance: “If nobody helped him and he did that I wouldn’t charge him for that because I would say we had caused that problem.” These broad cultural differences reflect variations in underlying norms regarding property, mutual responsibility, and accountability—norms that fundamentally shape the way we evaluate the wrongfulness of specific acts.

Before we turn to the details of this critique, we feel it is vital to disclose our motivations for undertaking it. We apprehend the world of criminal law from the intertwined vantage points of scholars, teachers, and interested citizens. What we see fills us simultaneously with wonder and fear, hope and anxiety.

To us, the most conspicuous feature of the criminal law landscape is political conflict. We observe persistent and intense disagreement on a wide variety of issues, many going to the core of the State’s twin obligations to protect its citizens from harm and to respect their freedom. When a man kills an attacker in a public space despite the opportunity to flee, is that murder or a justified exercise of self-defense? How about when a woman kills a sleeping husband who for years has subjected her to physical torment and emotional degradation? If a

---

5 See Joan G. Miller and David M. Bersoff, Culture and Moral Judgment, 62 J Personality & Soc Psych 541, 547 (1992) (reporting that 45 percent of American adults and 85 percent of Indian adults thought taking the ticket was appropriate, and that 43 percent of American third graders and 98 percent of Indian third graders thought that taking the ticket was appropriate).


8 See Kahan and Nussbaum, 96 Colum L Rev at 332–33 (cited in note 7) (comparing how jurisdictions treat this issue to how they treat duty to flee under self-defense doctrine).
man has sex with a woman who repeatedly says "no," should he be
deemed a rapist—or even punished at all?\(^9\) Should a man who "loses
control" and kills his wife for having sex with another man be treated
as less culpable than a premeditated murderer?\(^9\) How about a man
who kills another man for soliciting sex from him?\(^11\) Should people be
sent to jail for using recreational drugs?\(^12\) Is a corporation’s decision to
promote its stock with boastful speech about its balance sheet a form
of criminal fraud, or merely puffery that is protected by both the
common law and the First Amendment? To us, disputes over issues
like these attest to the remarkable heterogeneity of cultural values
within our society.

The diversity of positions political communities have adopted on
such issues—over place and over time—makes us conscious of the
plasticity of social norms and of the resulting urgency of using law to
promote morally defensible norms. At the same time, our recognition
of the unavoidable connection between the law’s position in such con-
licts and the status of contested visions of the good life makes us anx-
ious when assessing the proper scope for norm shaping in a liberal
society and intent on discovering means for avoiding cultural domina-
and accommodating difference.

This is decidedly not the picture of the criminal-law world painted
by Punishment Naturalists. They perceive not conflict but consensus,
not cultural heterogeneity but biological uniformity. As they read the
evidence (generated by their studies and those of others), “human
intuitions of justice about core wrongdoing ... are deep, predictable,
and widely shared,”\(^13\) the product of “evolved predisposition” and of
“social learning arising only from an aspect of human life experience
... so fundamental as to be essentially universal to all persons without
regard to circumstances or culture.”\(^14\)

---

9 See generally Susan Estrich, Real Rape (Harvard 1987) (discussing the controversy over
this issue).
10 See Kahan and Nussbaum, 96 Colum L Rev at 346–47 (cited in note 7) (noting contested
and changing attitudes on this issue).
(1999) (describing the political conflict over whether such offenses should be graded as the
aggravated “hate crime” form of murder or instead mitigated to voluntary manslaughter).
12 See John Hoeffel, Bid to Legalize Pot Advances: Initiative Backers Gather What Is Likely
to Be Enough Signatures to Put Their Measure on the California Ballot in November, LA Times
A14 (Jan 29, 2010). Consider also Gary Fields, Shorter Sentences Sought for Crack: Administra-
tion Tells Congress It Favors Ending Disparity with Powder Cocaine, Wall St J A3 (Apr 30, 2009)
detailing proposals by the Obama administration to eliminate the sentencing disparity between
reckless and powder cocaine offenses).
13 Paul H. Robinson and Robert Kurzban, Concordance and Conflict in Intuitions of Just-
ice, 91 Minn L Rev 1829, 1892 (2007).
14 Paul H. Robinson, Robert Kurzban, and Owen D. Jones, The Origins of Shared Intuitions
Where we see mutability in norms, and hence the inescapability of collective responsibility for their content, the naturalists apprehend their stability and warn of the futility and even perversity of using criminal law as an instrument of norm reform. "[T]he universal and intuitive nature of core judgments about justice" cautions against being "optimistic that arguments or education necessarily will produce ... change[s] in judgments about justice." And "trying to alter people's intuitions of justice" through law reform—or as the naturalists put it, "criminal law manipulation" by "social engineers" aimed at "get[ting] people to view conduct ... as condemnable or more condemnable"—must be viewed with deep suspicion: "[A] criminal justice system that regularly fails to do justice or that regularly does injustice, as judged by shared intuitions of justice ... will inevitably be seen as failing in a mission" that the community thinks important, thereby vitiating its "moral credibility" and fomenting "generalized contempt for the system in all its aspects, and a generalized suspicion of all of its rules." The dilemma of how to manage the norm-shaping potential of law in a liberal society is thus dispelled by the proclaimed nonexistence of meaningful cultural conflict combined with the prudential necessity of respecting genetically programmed moral instincts.

But the tangled complex of hopes and fears we experience when we survey criminal law is not vanquished by Punishment Naturalism. We think it is important to advise others who share our sensibilities that there is nothing in Punishment Naturalism to make them feel better (or worse). As far as we can tell, there is not a single position of any consequence on any of the contested issues we have already adverted (or on many like ones) that is ruled out or in by Punishment Naturalism.

The Punishment Naturalists might well demur; their target, they might claim, consists of marginal "academics or policy wonks" who believe that law should be structured on the basis of a utilitarian calculus that excludes public sensibilities altogether, or that the institution of criminal law or the practice of "punishment" should simply be abolished. (Ironically, the most serious purveyors of these positions are a pair of Antipunishment Naturalists who draw exactly the opposite

---

16 Id at 52.
17 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1688 (cited in note 14).
18 Robinson and Darley, 81 S Cal L Rev at 24 (cited in note 15). See also id at 23.
19 Id at 54.
20 See Robinson, Kurzban, and Jones, 60 Vand L Rev at 1688 (cited in note 14).
conclusion from the materials upon which the Punishment Naturalists themselves rely.3)

But their sweeping language, as well as some of their own examples of suspect "[norm]-reform programs"—ones aimed at "eradicat[ing] the notion that women often pretend to withhold consent to intercourse to appear more alluring," at "rais[ing] the general level of societal condemnation of the [domestic] abuser," and at "build[ing] public acceptance of both same-sex intercourse and the legal recognition of same-sex unions," for example22—invite a more expansive understanding of the significance of their work. As one thoughtful commentator reviewing Punishment Naturalist writings recently concluded:

Whatever theorists may think people should feel as a normative matter, as an empirical matter, members of the public share surprisingly fixed notions of justice in traditional crimes—and especially the kinds of crimes discussed in a criminal law course. . . . From the standpoint of law reform, then, reformers likely need to accept these shared intuitions as settled. And from the standpoint of teaching criminal law, I would add, professors need to recognize that there are relatively fixed and surprisingly hard-wired judgments widely shared in society that help to generate the legal rules found in criminal law codes and casebooks.23

Accordingly, in this Article, we address Punishment Naturalists' arguments on the assumption that they are intended to have implications for the pressing and conspicuous issues that are the everyday focus of mainstream criminal law scholars and of ordinary citizens interested in criminal law. And we show why it would be a mistake for anyone to accept that what they have to say counsels against arguing for reform of existing law.

In what follows, we argue that variations in cultural norms pervade evaluations of wrongdoing, even within what the Punishment Naturalists describe as the "core of wrongdoing." We explore these issues in four Parts. We begin by making the strongest case we can for Punishment Naturalism—and that case is tantalizing. It gets many things right and taps into deep intuitions that many individuals have about justice and the law. So while naturalism is significantly—even

---

21 See generally Joshua Greene and Jonathan Cohen, For the Law, Neuroscience Changes Nothing and Everything, 359 Phil Transactions Royal Socy B: Bio Sci 1775 (2004) (arguing that advances in neuroscience will create a shift in people's intuitions regarding free will and responsibility, resulting in a turn toward consequentialist punishment models).
22 Robinson and Darley, 81 S Cal L Rev at 52–53 (cited in note 15).
23 Orin Kerr, The Intuition of Retribution (Feb 17, 2010), online at http://crim.jotwell.com/the-intuition-of-retribution (visited May 1, 2010).
fundamentally—flawed, to understand why it is also deeply attractive and surprisingly persistent, it is important to acknowledge what it gets right and, perhaps more importantly, what it gets nearly right.

After laying out the naturalist claims, we then describe some of the problems that emerge from an examination of the available data. A host of studies in the fields that naturalists cite cannot bear the weight that naturalists place on them. For the most part, naturalists make the same kinds of mistakes that those making grand claims about universal human nature have long made: they fail to see the significance of the diversity that exists—diversity present not only in many of the studies they cite, but even in studies they themselves conduct.

We then describe Punishment Realism, an alternative approach that accounts for more of the available data and, we think, offers more practical purchase. Using cross-cultural examples and statistical analyses, we present a series of cases for which realism offers more detailed and parsimonious explanations than naturalism in two ways. First, rather than ignoring or downplaying diversity of intuition about wrongdoing, realism suggests that, to the extent that people value persons, objects, and practices differently, they also evaluate injuries to and interferences with them differently. Second, rather than positing an untestable "moral organ," Punishment Realism explains evaluations of wrongdoing with reference to well-established features of human cognition that are open to empirical evaluation.

Finally, we make a pragmatic pitch for the comparative advantage that Punishment Realism offers in the face of social dissensus. Where Punishment Naturalism suggests that attempting to educate individuals away from their instinctual intuitions regarding wrongdoing will be either fruitless or exceedingly difficult, Punishment Realism points reformers toward the cognitive and social mechanisms of norm formation. Conflict and dissensus based on differing worldviews will always be hard to resolve, but getting the source of the disagreement right, we think, is a step in the right direction.

I. NATURALISM AND WRONGDOING

Punishment Naturalism, which holds that our sense of right and wrong is largely innate, rests on observations of broadly shared sentiments about justice. Generally speaking, when someone commits a wrong—murder, rape, theft, or fraud, say—we share an intuitive sense that the wrongdoer should be punished. Moreover, we are likely to agree that some crimes are far worse than others: all other things

24 For a description of the use of the term "moral organ," see note 28 and accompanying text.
equal, a drawn-out, brutal, and deliberate rape-homicide seems worse than a quick and impulsive homicide. And just as we can distinguish between types of killings quickly and easily, so too can we distinguish among more and less serious forms of aggression from murder to rape to assault to battery, and the same can be said of theft and fraud: within each extremely general category, we can distinguish more serious from less serious cases.

As Punishment Naturalists note, this highly nuanced set of distinctions seems to come effortlessly. Where, Punishment Naturalists ask, do those intuitions come from? And why are they so widely shared? Surely, they answer, it is natural to feel the way we do about crime and punishment.25 A specialized cognitive module devoted to moral evaluations, naturalists argue, would explain both the extent of our shared intuitions and the ease with which we arrive at moral judgments. Marc Hauser, a professor of psychology at Harvard, captures the idea in his book, Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong, writing that humans have “evolved a moral instinct, a capacity that naturally grows within each child, designed to generate rapid judgments about what is morally right or wrong based on an unconscious grammar of action.”26

Naturalists deploy a series of analogies to bring this point home. Marc Hauser, John Mikhail, and others, extending Noam Chomsky’s famous (and famously controversial) analogy of linguistic cognition to the functioning of bodily organs,27 posit a “moral organ” or “module” in the mind that provides every normal human with a universal “grammar of action,” a generic moral code that underlies the apparent

25 While naturalism might seem like a more modest restatement of natural law, it is distinct in a number of ways. First and foremost, Punishment Naturalism makes no claims that these broadly shared sentiments are anything like a law. Nor does it claim that human intuitions are deontologically fair or materially useful; indeed, naturalists acknowledge that intuitions about the law may be unfair or counterproductive. On this account there is nothing “natural” about justice itself, there is only something “natural” about our intuitions about justice. Another way of saying this is that whereas the ambition of a theory of natural law is primarily normative, the ambition of Punishment Naturalism is primarily positive. Punishment Naturalists might derive practical implications from their research, but their goal is to tell us how humans actually do think, not how they should think, about crime and punishment. Punishment Naturalism thus dispenses with the philosophical debates of traditional legal theory by making claims that can be tested empirically and evaluated in terms of their practical value to policymakers and ordinary citizens. See Paul H. Robinson, Empirical Desert, in Paul H. Robinson, Stephen P. Garvey, and Kimberly Kessler Ferzan, eds, Criminal Law Conversations 29, 38 (Oxford 2009).


diversity of values and practices in the world.28 Paul Robinson, Robert Kurzban, and Owen Jones have similarly argued that "intuitions about morality and justice seem to develop" in "the same way that baby teeth grow from gums and adult teeth replace baby teeth."29 There is, on this account, a sense of right and wrong that, while non-obvious, is ever present, governing our evaluations of one another in ways that are surprisingly consistent.

Moreover, Punishment Naturalists argue, the innate nature of our intuitions about wrongdoing is of significant practical consequence. The fact that our intuitions arise from millions of years of shared evolutionary pressure matters because, if true, attempts to educate individuals away from their innate instincts are likely to be controversial, costly, and largely ineffective.30 Whatever pressures may have produced our innate sense of morality over the course of our evolution, they argue, "it is [now] beyond the normal influence of culture or demographic. If it were not so insulated, one would see differences in intuitions of justice among different demographics and cultures."31 Upon reflection, we might not like what our instincts tell us; but we should know what these instincts are. Highlighting the potential clash between socially constructed norms and natural intuitions, they argue that

policy wonks and politicians should listen more closely to ... the moral voice of our species. ... [For] in developing policies that dictate what people ought to do, we are more likely to construct long-lasting and effective policies if we take into account the intuitive biases that guide our initial responses to the imposition of social norms.32

The level of specificity at which the hypothesized innate mechanisms operate is crucial. Punishment Realists and Punishment Naturalists alike accept some form of innate sensitivity to social norms as part of

29 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1664 (cited in note 14).
30 See Robinson and Darley, 81 S Cal L Rev at 18 (cited in note 15) (arguing that "deviations [from humans' innate intuitions about punishment] can have undesirable consequences and unjustified costs that can ultimately hurt rather than help effective crime control").
31 Id at 11.
32 Hauser, Moral Minds at xx (cited in note 26). See also Robinson and Darley, 81 S Cal L Rev at 11 (cited in note 15) ("This insulation [from culture] means that there may be serious limits on whether and how social engineers can manipulate intuitions of justice, at least those intuitions of justice about core wrongdoing upon which there is broad agreement.").
humans’ capacity to learn. But the form of naturalism that we are describing rejects the notion that moral intuitions are principally the result of generic abilities to develop and act in response to socially constructed and shared meanings. The naturalism of the scholars we describe here, while readily admitting that individuals are made sensitive to norms by something like generic cognitive mechanisms, rejects innate-norm-sensitivity accounts as too limited, lacking in specialized cognitive mechanisms that supply much of the content of morality. Instead, naturalists argue, humans have cognitive features that determine the structure of highly predictable and largely invariant intuitions about wrongdoing, intuitions that began as advantageous human variations among our ancestors and, over centuries, were continually selected and refined as collective heritable traits. Although humans are not, on the naturalist account, identical moral machines, our innate moral intuitions are shared at surprisingly fine levels of granularity because of innate cognitive structures rather than socially acquired norms.

In support of these claims, naturalists offer empirical studies documenting the extent to which individuals share intuitions about whether acts are wrongful and how wrongful they are. They also develop an account of the sources of agreement and disagreement. Let us review each claim—the extent of shared intuitions and the source of this agreement and disagreement—in turn.

A. The Extent of Shared Intuitions

Where crime and punishment are concerned, humans certainly appear to disagree quite often; headlines and policy debates are filled, it seems, with clashing moral accounts. Is it possible that there is a deeper order lurking within the variance and dissensus that we observe around us? A growing number of researchers argue that there is a structure to our intuitions that is both nuanced and pervasive. Paul Robinson and Robert Kurzban, two prominent theorists working in the area of criminal law, have developed some of the most striking empirical studies supporting naturalist claims. Reviewing dozens of studies and conducting several themselves, they write:

---

33 There are dozens of norms-based models of cognition. For a recent review and addition, see Chandra Sekhar Sripada and Stephen Stich, A Framework for the Psychology of Norms, in Carruthers, Laurence, and Stich, eds, The Innate Mind 280, 289–90 (cited in note 28) (arguing that the “acquisition mechanism” people use to observe the existence of a norm “is both automatic and involuntary”).

[A]vailable evidence suggests that human intuitions of justice about core wrongdoing ... are deep, predictable, and widely shared. While there are disagreements about the relative blameworthiness of wrongdoing outside the core, the core wrongs themselves—physical aggression, takings without consent, and deception in exchanges—are the subject of nuanced and specific intuitions that cut across demographics.\(^3\)

The studies to which Robinson, Kurzban, and other naturalists cite involve participants who have been asked to rank the seriousness of offenses. The remarkable thing about such ranking exercises, they suggest, is the relatively stable rank order of the offenses evaluated by participants.

Consider the following sample from one recent study conducted by Robinson and Kurzban. Topping out the serious end of the spectrum is kidnapping an eight-year-old girl for ransom, raping her, recording her screams while burning her with a cigarette lighter, and then killing her once a demanded ransom is received. Consistently ranked as less serious than that is keeping pitbulls that escape repeatedly and ultimately kill someone. Less serious than that is slapping (and thus bruising) a man wearing a hat that makes fun of the defendant's favorite band. Less serious still is stealing a drill from a garage. And at the bottom of the culpability spectrum is taking (without eating) two whole pies from an "all you can eat" buffet.\(^4\)

That list is not exhaustive; the study includes over twenty acts that participants rank, and which they rank with a very high degree of consistency and ease. (A full listing of the offenses is provided in the Appendix.) How consistently do members of the public rank these offenses? Participants agreed on 91.8 percent of all pairwise judgments, and the ranking produces a Kendall's W of 0.88.\(^5\) As you might imagine, the most common disagreements were on those acts that were ranked just next to one another. When the researchers discounted the "flipping" of adjacent offenses, the extent of agreement rose to 93.9 percent.\(^6\)

A summary of the scenarios and their rankings is provided in Table 1 below, and full descriptions are provided in the Appendix. Notice that the first four scenarios were generally thought to deserve no

---

\(^{35}\) Robinson and Kurzban, 91 Minn L Rev at 1892 (cited in note 13).

\(^{36}\) Id at 1869 (providing a table ranking behavior according to the amount of punishment the study's subjects believed was warranted).

\(^{37}\) Id at 1877–78.

\(^{38}\) Id at 1878 (noting that one-third of all deviations were "adjacent flip" deviations).
punishment and thus are unranked. The following offenses are listed in order of ranked seriousness.

**TABLE 1. RANKINGS OF RELATIVE WRONGFULNESS BY VARIOUS DEMOGRAPHIC GROUPS**

<table>
<thead>
<tr>
<th>Act</th>
<th>All Subjects</th>
<th>Male</th>
<th>Female</th>
<th>Non-white</th>
<th>White</th>
<th>&lt;$60K Income*</th>
<th>&gt;$60K Income*</th>
<th>&lt;2yr Degree</th>
<th>&gt;2yr Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-defense</td>
<td>0†</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Coerced theft</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Umbrella mistake</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hallucination</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pies</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Short change</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>T-shirt</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Radio theft</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7.8‡</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Drill theft</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Microwave theft</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>TV destruction</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Slap</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Head-butt</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Stitches</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Necklace &amp; stitches</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Robbery</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Clubbing</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Pitbulls</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Infant</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Stabbing</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Ambush</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Abduction</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Burning</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Ransom</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
</tbody>
</table>

N = 246 123 123 53 193 102 103 169 77

* Forty-one subjects did not provide income information.
† "No punishment" as the modal response is shown as 0.
‡ The two ranks were a tie, thus both modes are reported.

Not only are the rankings highly consistent, but the researchers found "little variation in the modes of scenario rankings" across a broad array of demographic variables (arrayed across the top of Table 1 above). In addition to these demographic variables, the researchers also report that an "investigation of ... political party, ideology, marital status, whether they have children, religion, level of religious activity, [and] libertarianism showed a similar lack of any meaningful difference between demographic groups' modal rankings."

The acts ranked here are also, Robinson and Kurzban claim, broadly representative of criminal wrongdoing more generally. While they concede that individuals disagree about the wrongfulness of some acts, these divisive acts are outside of what the researchers consider to

---

40 Id.
41 Id at 1879 n 200.
be the "core" of wrongdoing: "physical aggression, takings without consent, and deceit in exchange." The core, by Robinson and Kurzban's estimation, comprises the vast majority of criminal conduct: "94.9% of the offenses committed in the United States." Nor are these findings specific to a single time or place. While Robinson and Kurzban's is the most striking of recent studies conducted in the United States, ranking studies of this sort have been conducted for over forty years and in a number of countries with similar results. Thorsten Sellin and Marvin Wolfgang's classic 1964 study demonstrated that Americans consistently ranked many crimes in the same order, and that these could be reliably reported as an index of crime seriousness. Dogan Akman and Andre Normandeau's 1967 study reported similar findings across a dozen samples taken from various Canadian locales, concluding that rankings of many offenses were stable and reliable enough to construct a crime index for Canada. In 1980 Sandra Evans and Joseph Scott reported that American and Kuwaiti students ranked many offenses and punishments similarly. And in 2006, Sergio Herzog reported remarkable similarities in the rankings of offense seriousness across cultural groups in his study of Israeli Arabs and Israeli Jews.

Robinson and Kurzban are careful to note that agreement on the relative seriousness of various forms of wrongdoing is not equivalent to agreement on how to punish in absolute terms. Some individuals may be more punitive than others overall, generating disagreement that has long masked the pervasive structure of our punishment intuitions about

42 Id at 1892.
43 Robinson and Kurzban, 91 Minn L Rev at 1867 (cited in note 13).
44 See generally Thorsten Sellin and Marvin E. Wolfgang, The Measurement of Delinquency 263–65 (Wiley 1964) (finding similar rankings among judges, police officers, and two groups of students, despite their positions in very "different sociocultural groups"). Several studies have since replicated and extended these findings in the United States. See, for example, Don C. Gibbons, Crime and Punishment: A Study in Social Attitudes, 47 Soc Forces 391, 395 (1969) (studying proposed sentences among Californians and finding general agreement on which crimes should get severe versus more moderate punishment); Peter H. Rossi, et al, The Seriousness of Crimes: Normative Structure and Individual Differences, 39 Am Soc Rev 224, 230–31 (1974) (finding broad agreement on assessments of crime seriousness among residents of Baltimore regardless of race, sex, or educational attainment); Stephen D. Gottfredson, Kathy L. Young, and William S. Laufer, Additivity and Interactions in Offense Seriousness Scales, 17 J Resch Crime & Deling 26, 29 (1980) (studying the relative-seriousness rankings of several criminal offenses by undergraduate and graduate students at Johns Hopkins University). See also Monica A. Walker, Measuring the Seriousness of Crimes, 18 Brit J Criminol 348, 348–51 (1978). But see Parts II and III.
wrongfulness. To use the analogy to Chomskian "universal grammar" that many naturalists employ, the relative order in which wrongdoings are ranked is a "principle" determined by a moral mechanism or module in the brain, and the amount of punishment to be meted out—the "end point[s] of the punishment continuum," as Robinson and Kurzban say—is a "parameter" that may be set by culture, experience, and other non-innate influences on preference. 49

On the naturalist account, then, while we may disagree over some things, this superficial dissensus masks the deeper structure of our shared intuitions. Looking for the kinds of serious disagreements that are often described as refuting naturalist accounts, Robinson and Kurzban report having "failed to find the limits of shared intuitions for core wrongdoing." 50 The levels of agreement in rank ordering are, they argue, "astonishingly high." 51

B. The Source of Shared Intuitions

Researchers involved in these studies describe the degree of shared intuitions about wrongdoing and punishment as "stunning[]," "remarkable," "striking," and even "shock[ing]." 52 To explain the extraordinary concordance they see, they develop an evolutionary theory of human psychology. By and large, they propose one or more specialized cognitive mechanisms developed in response to evolutionary pressures. As Marc Hauser writes: "Part of [our natural sense of justice] was designed by the blind hand of Darwinian selection millions of years before our species evolved; other parts were added or upgraded over the evolutionary history of our species, and are unique both to humans and to our moral psychology." 53 For naturalists, developing a moral sensibility over the course of a lifetime is "like growing a limb"—a highly specialized form that normally develops in a predictable manner.

---

48 Robinson and Kurzban, 91 Minn L Rev at 1855 (cited in note 13).
49 Hauser develops this concept extensively. See Hauser, Moral Minds at 419–20 (cited in note 26).
50 Robinson and Kurzban, 91 Minn L Rev at 1867 (cited in note 13).
51 Id.
52 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1654 (cited in note 14).
53 Evans and Scott, 22 Criminol at 53 (cited in note 46).
55 Joss Whedon, dir, Dr. Horrible's Sing-Along Blog Act 2, 00:11:59 (2008), online at http://dhorrible.com (visited May 1, 2010) ("All the time that you beat me unconscious I forgive / ... It's a brand new me / I got no remorse / Now the water's rising / ... I'm gonna shock the world / Gonna show Bad Horse.").
56 Hauser, Moral Minds at xvii (cited in note 26).
57 Id at xviii (clarifying that the acquisition of moral norms does not occur through formal education).
The advantages that these intuitions provide are often not specific to an individual, they argue, but rather accrue to kin groups, to local populations, or to the species as a whole. Thus, while it may be costly for an individual to demand or inflict punishment on another for a core wrongdoing, overall the group to which that individual belongs (and, presumably, within which she has many genetic relatives) will thrive if she does. Robinson, Kurzban, and Jones develop a specific conception of the mechanism by which our shared intuitions have evolved, one that rests on the conditions of mutual interdependence and social interaction:

We argue . . . that human sociality has laid the foundation for an evolved predisposition to acquire shared intuitions of justice and that such intuitions benefit the individuals bearing them. . . . [E]volution has in particular contributed to intuitions that physical harm, the taking of property, and cheating in exchanges are matters for particular attention and condemnation.  

Another way to put this is that specific forms of antisocial behavior are evolutionarily counterproductive, so groups (and individuals in groups) that have innate rules that foster cooperation—including cooperation around punishment—are more likely to thrive.

In support of this theory, researchers commonly cite two pools of evidence: experimental games of cooperation and punishment among humans, and studies that turn on distinctions between moral and conventional wrongdoing. We review each of these in turn.

1. Fairness games.

Consider, first, one of the most common tools for assessing how individuals assess fairness and how much they are willing to sacrifice to punish someone who is behaving unfairly: the “Ultimatum Game.”  
The structure of the game is simple. Two people—a “Proposer” and a

---

58 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1646 (cited in note 14).
59 Naturalists provide other evidence, though in general it tends to be at a greater inferential remove. Brain imaging studies, for example, are cited quite often. While these studies do provide insight into which regions of the brain (it is usually multiple regions) are most active when individuals are attempting to resolve various problems, they do not provide much in the way of evidence about whether the content is innate or learned. There is, so far as we can discern, no evidence that moral decisions are made exclusively or even predominantly by regions of the brain that are responsive to only “innate,” as opposed to “learned,” intuitions. See id at 1659–64 (arguing, based on neurological studies, that “basic moral sentiments humans share are products of evolutionary processes”). See also id at 1655–59 (providing evidence from animal studies); id at 1664–74 (providing evidence from studies of child development).
"Responder"—are anonymously paired. The Proposer is offered a modest or significant amount of money. She then proposes a split of the money with the Responder. The Responder decides whether to accept the split or reject it. If she accepts, both will get the amount allocated by the proposed split. If she rejects, neither gets anything. One common interpretation of such a rejection is that it is a form of punishment that the Responder visits on the Proposer for being unfair, a punishment that costs whatever the Responder would have received had she accepted the Proposer's suggested split.

If humans are selfish actors in the way neoclassical economics posits, then the Responder would never reject an offer greater than zero, no matter how small. Why? A rational Responder should accept an offer of any size because some money, no matter how little, is better than no money. As such, every neoclassically rational and selfish Proposer would offer as little as possible, keeping the lion's share for herself.

Yet in most studies of industrialized societies, the mean offer is between 40 percent and 50 percent. Moreover, if Proposers offer significantly less than this, Responders tend to reject the offers in proportion to their divergence from the norm. But few people make such low offers. Indeed, experimenters often had to add in random offers to test the lower bounds of what a Responder would accept because Proposers deviated from the mode so rarely and, when they did, by very little. The consistency of these findings across a range of settings led many researchers to posit a "taste" for fairness at approximately these levels.

Marc Hauser has conducted a number of studies of the Ultimatum Game. In a recent version broadcast on national television, for example, he gave half of a group of students some Skittles candies and had them determine whether and how many they wanted to share with those who were given none. To a person, they all gave half. The reason?

---

61 Id at 801.
62 Id at 797 (characterizing the infrequent nature of low offers as demonstrative of fairness and concerns of reciprocity).
63 See, for example, James Konow, A Positive Theory of Economic Fairness, 31 J Econ Behav & Org 13, 32-33 (1996) (identifying accountability, altruism, and efficiency as important motivators); Gary E. Bolton and Rami Zwick, Anonymity versus Punishment in Ultimatum Bargaining, 10 Games & Econ Behav 95, 113 (1995) (arguing that "punishment for unfair treatment" accounted for most of the variation from perfect equilibrium play in the study's results); Colin Camerer and Richard H. Thaler, Anomalies: Ultimatums, Dictators and Manners, 9 J Econ Persp 209, 216 (1995) (explaining that etiquette and perceived norms of fairness often overcome income maximization as motives in the Ultimatum Game); Alvin E. Roth, Bargaining Experiments, in John H. Kagel and Alvin E. Roth, eds, The Handbook of Experimental Economics 253, 264-65 (Princeton 1995) (detailing types of bargaining experiments and noting that some theorists have rallied around the explanatory power of fairness considerations).
When pressed, the students gave explanations that seem either intuitive or like a confabulation to most people: “Because then . . . you all get the same amount.” To them, it was just natural to share equally.

Like Hauser, most naturalists describe the typical deviation from the rational-actor model in evolutionary terms. The argument for naturalism based on these games is that natural selection has crafted a sense of fairness in humans that might be modified modestly by experience but is essentially innate. Fairness, on this account, is no more cultural than any other human organ; instead, it is an organic “moral faculty—an organ of the mind that carries a universal grammar of action.” On this account, the limited range of fairness is part of human adaptive fitness, and the “architecture of our mind, leftover circuitry from the cavemen.”

Hauser’s theory also comes with at least one prediction. Because these constraints on fairness have long been essential to our survival, he argues that “no culture will ever [accept] offers under 15 percent, and no culture will ever offer more than 50 percent. If they do, such patterns will exist for the blink of an eye in human history.” We return to Hauser’s claim later, as it shares a form of logical error common to many naturalist claims about punishment and human intuition.

2. Conventional and moral wrongdoing.

A second body of work cited as supporting the naturalist account derives from empirical studies distinguishing “moral” from “conventional” wrongdoing, a distinction that roughly tracks the legal distinction between acts that are traditionally described in legal parlance as mala in se and mala prohibita. Following Elliot Turiel, Judith Smetana, and

65 Id.
66 Hauser, Moral Minds at 11 (cited in note 26).
67 Id at 85–86.
68 We have corrected a typo here; Hauser writes: “no culture will ever reject offers under 15 percent,” id at 85, but that cannot be what he means because most do. But see notes 119–29 and accompanying text (critiquing this form of hypothesizing in general and describing experiments among the Sukuma in Mahenge, Tanzania that exceed Hauser’s hypothesized bounds on fairness).
70 Many legal scholars have argued that a context-independent distinction between the two is impossible. See, for example, Peter Aldridge, Making Criminal Law Known, in Stephen C. Shute and A.P. Simester, eds, Criminal Law Theory: Doctrines of the General Part 103, 106–10 (Oxford 2002) (discussing the difficulties of distinguishing these categories, particularly when
Larry Nucci’s early empirical work in the late 1970s and early 1980s, these studies suggest that humans, in the course of normal development, learn to distinguish moral wrongs, which implicate “justice, rights, or welfare” (hitting, stealing, or refusing to share an abundant good, for example), from conventional wrongs, which merely violate a local convention (wearing pajamas to school or work, swearing, or eating lunch while standing up, for example). Conventional transgressions are thought to be less serious, and assessments of their seriousness are dependent on context and rules set by authorities; moral transgressions, on the other hand, are thought to be more serious, typically involving clear harm to a victim, and the seriousness of the transgression is thought to be “authority independent”—that is, it does not depend on what any authority says is acceptable.

Naturalists often cite studies of moral and conventional wrongs showing that children appear to learn that hurting others is wrong before they learn other norms. For example, the finding that “the first moral concept to appear in children is the concept that physical aggression is wrong” is relevant, Robinson, Kurzban, and Jones argue, because it seems “likely more than coincidence that this is also the first step in [a naturalist] account of the evolutionary origins of intuitive justice.” That is, the development of morality in childhood parallels

applied to market situations where it is unclear whether the legal rule is simply “declaring the pre-legal public wrongfulness of the actions in question”); Nancy Travis Wolfe, Mala In Se: A Disappearing Doctrine?, 19 Criminol 131, 139–40 (1981) (arguing that mala prohibita offenses are different from mala in se offenses only in that the legislatures have labeled them as such).

71 See Larry P. Nucci, Conceptual Development in the Moral and Conventional Domains: Implications for Values Education, 52 Rev Educ Rsch 93, 100 (1982) (describing early studies by Turiel, Nucci, and Smetana in which subjects evaluated less wrongful actions “in terms of their relation to the social order, social expectations, social institutions, and contextual or culturally specific regulations and standards of behavior” and more wrongful actions “in terms of the effects the actions had on the rights or well-being of others”).

72 There are several good reviews of the literature. See, for example, Judith G. Smetana, Understanding of Social Rules, in Mark Bennett, ed, The Development of Social Cognition: The Child as Psychologist 111, 112–14 (Guilford 1993); Marie S. Tisak, Domains of Social Reasoning and Beyond, in Ross Vasta, ed, 11 Annals of Child Development 95, 100–01 (Jessica Kingsley 1995); Larry P. Nucci, Education in the Moral Domain 7–9 (Cambridge 2001). See generally Larry P. Nucci, Elliot Turiel, and Gloria Encarnacion-Gawrych, Children's Social Interactions and Social Concepts: Analyses of Morality and Convention in the Virgin Islands, 14 J Cross-Cult Psych 469 (1983) (finding that adults and preschoolers from the Virgin Islands responded to “moral transgressions” by pointing out the hurtful or unjust consequences of the actions upon victims, but reacted to “conventional” transgressions by referring back to aspects of the social order); Marida Hollos, Philip E. Leis, and Elliot Turiel, Social Reasoning in Ijo Children and Adolescents in Nigerian Communities, 17 J Cross-Cult Psych 352 (1986) (finding that Nigerian children’s “moral” and “conventional” judgments can be distinguished along similar axes); Yau and Smetana, 74 Child Dev 647 (cited in note 69) (finding that Chinese preschool children also treated “personal,” “moral,” and “conventional” events differently).

73 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1670 (cited in note 14).
the evolutionary development of the moral organ. Lack of cooperation around physical aggression would, on this account, pose especially grave risks to survival, so intuitions governing its regulation would be the most fundamental and earliest to develop.

Over the last quarter century, several researchers have reported similar patterns across a diverse set of subjects ranging in age from toddlers as young as three-and-a-half years to adults, with a substantial array of different nationalities and religions. Reading these studies, naturalists have drawn the inference that these distinctions between moral and conventional wrongs are “universally recognized, similar among boys and girls, and even consistent in cultures with seemingly different parental styles—in China and the United States.”

Moreover, they argue, because these studies tend to show that “moral rules are inviolable and universally applicable,” they can be taken as evidence of an “evolutionary explanation for the origins of intuitions of justice.”

In light of the universal nature of these intuitions, naturalists suggest that the alternative is simply implausible:

If there were no specific developmental system for the acquisition of moral intuitions, if intuitions of justice were simply a matter of general social learning, then the developmental route of the acquisition of intuitions of justice would depend on the environment in which the child developed. The things that the child learned were wrong would include acts the child witnessed, ideas communicated through language, pedagogy from various sources, and so forth. Because all of these elements are likely to differ widely across cultures, and even across family and peer groups

---

74 This argument echoes Ernst Haeckel’s fascinating recapitulation theory, which (incorrectly) held that “ontogeny recapitulates phylogeny”—that is, that the physical development of each human over the course of its lifetime parallels the evolutionary development of the species. See Stephen Jay Gould, Ontogeny and Phylogeny 7–9 (Belknap 1977) (discussing the origins of the recapitulation theory and arguing that it collapsed once “Mendelian genetics repudiated the generality of its two necessary principles—terminal addition and condensation”).

75 See Nucci, Turiel, and Encarnacion-Gawrych, 14 J Cross-Cult Psych at 469 (cited in note 72) (studying Virgin Islands children and adults); Hollos, Leis, and Turiel, 17 J Cross-Cult Psych at 352 (cited in note 72) (studying Nigerian children); Yau and Smetana, 74 Child Dev at 647 (cited in note 69) (studying Hong Kong preschoolers). For reviews, see Smetana, Understanding of Social Rules at 126–33 (cited in note 72) (discussing research in various domestic contexts as well as in such countries as Japan and Zambia); Tisak, Domains of Social Reasoning at 103 (cited in note 72) (discussing results across age ranges); Nucci, Education in the Moral Domain at 20–51, 94–106 (cited in note 72) (discussing results across a variety of religions and cultures).

76 Hauser, Moral Minds at 291 (cited in note 26).

77 Id at 292.

78 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1666 (cited in note 14) (finding collateral support for this explanation in evidence that “children everywhere progress through similar stages of moral reasoning about justice at roughly the same ages”).
within cultures, such a general learning system would yield very
different paths and timing in the acquisition of intuitions of jus-
tice for different individuals.\textsuperscript{79}

In short, they argue, there is profound agreement on the core moral
wrongs that we confront, and that agreement appears to be intuitive,
nuanced, and organic—in all probability the product of a specialized
and innate cognitive moral organ that has developed over millions of
years through natural selection.

II. PROBLEMS WITH NATURALISM

Naturalists have assembled an impressive collection of studies in
support of their claims, and the literature on precisely which aspects of
morality are innate has become a booming cottage industry.\textsuperscript{80} More-
over, they do so by referencing empirical data, which is surely an ad-
vance on many earlier anecdotal studies.\textsuperscript{81} But they face a host of
problems. Some stem from simple logical missteps that underlie their
most strident claims. A more serious problem is posed by empirical
evidence contradicting the central claim that evaluations of serious
wrongfulness do not vary across social conditions or individuals. Set-
ting up the discussion of Punishment Realism in Part III, this Part
starts with a few basic examples that give a sense of the research that
naturalists have overlooked or failed to incorporate, then describes
some of their broader logical errors.\textsuperscript{82}

\textsuperscript{79} Id.

\textsuperscript{80} See generally Erica Roedder and Gilbert Harman, Grammar, in John M. Doris, Shaun
Nichols, and Stephen Stich, eds, Empirical Moral Psychology (Oxford forthcoming 2010); Ron
Mallon, Reviving Rawls’s Linguistic Analogy Inside and Out; in Walter Sinnott-Armstrong, ed, 2
Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity 145 (MIT 2008);
Jesse Prinz, Resisting the Linguistic Analogy: A Commentary on Hauser, Young, and Cushman, in
Sinnott-Armstrong, ed, The Cognitive Science of Morality 157 (rejecting the analogy to Noam
Chomsky’s “universal grammar” by suggesting that “general-purpose emotion systems and so-
cially transmitted rules” could be an alternate explanation for the empirical findings of Hauser
and others); Frances Myrna Kamm, Intricate Ethics: Rights, Responsibilities, and Permissible
Harm (Oxford 2007); John Mikhail, Universal Moral Grammar: Theory, Evidence, and the Future,

\textsuperscript{81} See generally Robinson, Empirical Desert (cited in note 25) (describing the relative
virtues of empirical conceptions of desert).

\textsuperscript{82} We do not mean to select Paul Robinson, Robert Kurzban, John Darley, and Owen
Jones for special scrutiny; in fact, they are to be applauded for presenting claims in a manner that
is amenable to critical examination and testing. Other naturalists who cite many of the same
studies theorize a “universal moral grammar” that has, so far as we can tell, no rules that can
actually be tested. As Michael Waldmann has noted, the notion of a universal moral grammar is
developed without an explanation of what, exactly, distinguishes the rules from parameters in the
proposed universal moral grammar:

Findings that show that different cultures generate similar intuitions . . . are viewed as evidence
for universal rules, whereas other studies showing huge cultural differences are interpreted as
A. Some Skepticism about Scope

One of the first major hurdles that naturalism faces is the tremendous scope of disagreement over what constitutes wrongdoing. Naturalists’ principle strategy is to suggest that the disagreements are relatively minor and cloud our view of the inner workings of moral intuition. There are disagreements about crimes, they concede, but these are marginal crimes, relatively infrequent when compared to the “core” crimes on which there is significant agreement. As noted above, Robinson and Kurzban, drawing on data from the National Criminal Victimization Survey conducted by the Bureau of Justice Statistics, claim that the “kinds of offenses in the scenarios” they study “represent 94.9% of the offenses committed in the United States.” As such, the kinds of offenses on which people disagree are necessarily less common.

But is the public really in agreement about the relative seriousness of the vast majority of bad acts committed in the United States? Anyone familiar with the source of these data will immediately recognize one problem with the claim: a survey of criminal victimization does not include any so-called “victimless” or “vice” crimes—crimes over which there is tremendous public disagreement. As indicated in Tables 2 and 3 below, the incidence of these crimes greatly outnumbers the incidence of criminal victimizations. Indeed, the number of people estimated to be using marijuana in the last year alone exceeded the number of all those estimated to have suffered criminal victimization of any kind. Add prostitution (recent studies find that more than one in six adult males has paid for sex) and you begin to see just how common controversial crimes are. Also excluded from the list are a number of regulatory crimes. While far harder to estimate, surveys suggest that rates of willful tax evasion—the seriousness of which is also disputed—run as high as 25 percent of the population.

evidence for the role of parameters. This flexibility of the theory makes it hard to envision what could constitute a strict empirical test of the theory.

Indeed, many of the empirical studies [cited] could even be taken as evidence against the moral grammar view.

83 See note 43 and accompanying text.
84 Whether or not they are actually “victimless” is one of the points of contention.
85 See sources cited in note 91.
TABLE 2. INCIDENCE OF CRIMINAL VICTIMIZATION: 200687

<table>
<thead>
<tr>
<th>Criminal Victimization</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murder</td>
<td>17,034</td>
</tr>
<tr>
<td>Rape/Sexual assault</td>
<td>260,940</td>
</tr>
<tr>
<td>Robbery</td>
<td>712,610</td>
</tr>
<tr>
<td>Assault</td>
<td>5,120,840</td>
</tr>
<tr>
<td>Household burglary</td>
<td>3,560,920</td>
</tr>
<tr>
<td>Motor vehicle theft</td>
<td>992,260</td>
</tr>
<tr>
<td>Theft</td>
<td>14,362,570</td>
</tr>
<tr>
<td>Any criminal victimization</td>
<td>25,200,384</td>
</tr>
</tbody>
</table>

TABLE 3. INCIDENCE OF VICE CRIMES

<table>
<thead>
<tr>
<th>Crime</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marijuana use88</td>
<td>16,700,000/mo</td>
</tr>
<tr>
<td>Underage drinking89</td>
<td>27.2% of 12–20 year-olds</td>
</tr>
<tr>
<td>Tax evasion80</td>
<td>24% of adults</td>
</tr>
<tr>
<td>Paying for sex91</td>
<td>15% of adult men</td>
</tr>
</tbody>
</table>


88 US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, 1 Results from the 2009 National Survey on Drug Use and Health: Summary of National Findings 13 (Sept 2010), online at http://www.oas.samhsa.gov/NSDUH/2k9NSDUH/2k9ResultsPpdf (visited Sept 24, 2010).

89 Id at 35.

90 Mason and Calvin, 13 L & Soc Rev at 80–81 (cited in note 86) (reporting a tax evasion rate of 24.2 percent).

These estimates are, of course, quite rough. But the point stands: even if we were to add only those crimes listed in Table 3—and there are many more that could be added92—the crimes of the sort that naturalists believe form stable rankings turn out to be in the minority rather than the majority.

Another reason to be skeptical of the suggestion that we share intuitions about most classes of wrongful acts is that the classes of acts listed also exclude acts that a substantial number of Americans believe should be crimes, but which are not—a few of which are summarized in Table 4 below. For example, while hard to estimate with a high degree of accuracy, most researchers estimate that there are more than one million abortions performed each year in the United States.93 Sodomy, which was illegal until quite recently in many jurisdictions, is estimated to be more common among men and women, both straight and gay, than all violent crime, property crime, and illegal drug use combined.94 Similarly, more people possess and view pornography than are listed as victims of all the crimes in the statistics that Robinson and Kurzban cite.95 Nor does the “core” cover the failure to assist others who are in need—for example, in instances where a person could help a small child who is being abused but does not.96 These noncriminal acts should also be considered when estimating the extent of agreement because the theoretical question being addressed is not

---

92 Child pornography, narcotics use and distribution, public urination, and indecent exposure, to name just a few.


94 ABC News Primetime Live, The American Sex Survey at 2 (cited in note 91) (indicating that one in five respondents reported having looked at pornography on the Internet).

95 See Stacy Finz, Killing of Girl, 7, in Casino Spurs Good Samaritan Bills, SF Chron A21 (Dec 9, 1998) (observing that the inability to charge a college student for his failure to either prevent or report the murder of a young girl prompted the legislative introduction of reporting requirements in California).
about what happens to be listed in a victimization survey, but about what kind of wrongdoing is considered serious.

**TABLE 4. ESTIMATED INCIDENCE OF CONTROVERSIAL NONCRIMINAL ACTS**

<table>
<thead>
<tr>
<th>Act</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodomy(^{97})</td>
<td>80,000,000/yr</td>
</tr>
<tr>
<td>Abortion(^{98})</td>
<td>1,000,000/yr</td>
</tr>
<tr>
<td>Internet pornography(^{99})</td>
<td>21,000,000/mo</td>
</tr>
<tr>
<td>Gambling(^{100})</td>
<td>&gt;60% of adults/yr</td>
</tr>
<tr>
<td>Nude performances</td>
<td>Unknown</td>
</tr>
<tr>
<td>Failure to assist in an emergency</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

How do we know that people disagree about the seriousness of these criminal and noncriminal acts? For starters, different communities regulate these activities in a wide variety of ways. Prostitution is legal in Nevada, but not in New York;\(^{101}\) Internet gambling is legal in New York, but not in Louisiana;\(^{102}\) nude performances are illegal in Iowa,\(^{103}\) but not in California;\(^{104}\) failing to help someone who is in grave danger when you can do so without much trouble is not punished in California, but it is in Vermont.\(^{105}\)

---

\(^{97}\) People engaging in sodomy per year (presumably some engage in it more than once a year). See Michael, et al, *Sex in America* at 140 (cited in note 91) (observing that 10 percent of men and 9 percent of women have engaged in anal sex within the past twelve months).

\(^{98}\) Jones, et al, 40 Persp Sex & Repro Health at 9 (cited in note 93).


\(^{100}\) National Opinion Research Center, *Report to the National Gambling Impact Study Commission* 8 (Apr 1, 1999), online at http://www2.norc.org/new/gamb-fin.htm (visited May 1, 2010) (showing that 60 percent of women and 67 percent of men had gambled in the previous year).

\(^{101}\) Compare Nev Rev Stat § 244.345 with NY Penal Law § 230.00 (McKinney). Rhode Island only recently barred citizens from paying money for sex, but street solicitation and the operation of brothels were already prohibited. See Associated Press, *Rhode Island: New Prostitution Law*, NY Times A17 (Nov 4, 2009).

\(^{102}\) Compare NY Penal Law §§ 225.00, 225.05, 225.10 (McKinney) (regulating only certain gambling activities) with La Rev Stat Ann § 14:90.3 (West) (prohibiting Internet gambling). See also Julia Kollewe, *Former Gambling Chief Dicks Is Freed in US*, Independent (Sept 30, 2006), online at http://www.independent.co.uk/news/business/news/former-gambling-chief-dicks-is-freed-in-us-418184.html (visited May 1, 2010) (noting that the governor of New York refused to sign an order to extradite an alleged Internet gambler to Louisiana).

\(^{103}\) Iowa Code Ann § 728.5 (West).

\(^{104}\) See *Nunez v Holder*, 594 F3d 1124, 1144–45 (9th Cir 2010) (Bybee dissenting) (discussing the application of Cal Penal Code § 314, which prohibits indecent exposure, to nude dancing at clubs).

\(^{105}\) Compare Cal Penal Code § 152.3 (West) (imposing the duty to report only in certain situations involving children) with 12 Vt Stat Ann § 519(a) (Equity) (mandating that a person who knows that another person is "exposed to grave physical harm" must, under certain circumstances,
But even more convincing is evidence from experiments conducted by Robinson and Kurzban themselves, reported in the same article arguing that there was broad agreement on intuitions regarding wrongdoing, summarized in Table 5 below. They found, for example, that a third of participants thought that smoking marijuana should bring no penalty at all. A similarly large percentage of the population felt the same way about prostitution. And while many of the acts that Robinson and Kurzban included are less controversial (most would agree that an abortion in the seventh month is wrong, but what about in the third or fourth month?), they provide enough evidence of public dissensus on these issues to make one wonder how they can be so confident in their claims that our understandings of wrong acts are so broadly shared and deeply nuanced.

### TABLE 5. SUMMARY OF RANKINGS OF CONTROVERSIAL ACTS SHOWING SIGNIFICANT DISAGREEMENT OVER RELATIVE WRONGFULNESS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean Rank</th>
<th>Modal Rank</th>
<th>Percent Assigning “No Liability”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marijuana</td>
<td>2.2</td>
<td>t</td>
<td>33</td>
</tr>
<tr>
<td>Prostitution</td>
<td>2.4</td>
<td>t</td>
<td>30</td>
</tr>
<tr>
<td>Cocaine</td>
<td>4.0</td>
<td>t</td>
<td>19</td>
</tr>
<tr>
<td>Bestiality</td>
<td>4.2</td>
<td>t</td>
<td>16</td>
</tr>
<tr>
<td>Teen alcohol</td>
<td>4.8</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Drunk crash</td>
<td>6.2</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Third theft</td>
<td>7.1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Late abortion</td>
<td>7.5</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Cocaine dealer</td>
<td>7.9</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Unwanted sex</td>
<td>8.7</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Cocaine importer</td>
<td>8.9</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Rape</td>
<td>11.1</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td><strong>N = 246</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† These scenarios had a modal rank of “no punishment.”

"give reasonable assistance to the exposed person unless that assistance or care is being provided by others"). Minnesota also has a statutory duty to assist. See Minn Stat Ann § 604A.01 (West).

106 Robinson and Kurzban, 91 Minn L. Rev at 1883 (cited in note 13) (“[T]here are punishment-assignment issues on which people do indeed disagree.”).

107 Id at 1887 table 8.
B. Core Meltdown

How do naturalists explain this disagreement? These controversial acts, they argue, should not be evaluated alongside the others because they are outside of the "core" of wrongdoing: "physical aggression, takings without consent, and deception in exchanges." On this account, there are some acts that are so important to our individual and collective welfare that we have evolved a shared intuition that they are wrong; others are less important, so there is more room for diverse intuitions.

One reason to be unsatisfied with the core–periphery distinction is that it fails to tell us what, exactly, distinguishes the important core from the unimportant periphery of crimes. Are we agreed that controversial acts (incest, abortion, prostitution, mistakes about sexual consent, failing to help a child in need, drug use, whippings, cannibalism, just to name a few) are unimportant? Sexual misconduct, for example, might reasonably be included in the "core" on evolutionary grounds, as sexual activity (so far as we can tell, anyway) is central to continued survival; and yet there is dramatic cross-cultural disagreement over the enforcement of sexual mores and the punishment of sexual misconduct.

Many of the cross-cultural ranking studies that naturalists cite actually support the conclusion that there is no reliable core–periphery or moral–conventional distinction. Consider, for example, the rankings reported by Evans and Scott in their cross-cultural comparison of crime seriousness among US and Kuwaiti students, excerpts of which are reported in Table 6 below.
<table>
<thead>
<tr>
<th>Description of Act</th>
<th>Kuwaiti Ranking</th>
<th>US Ranking</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A married woman committed adultery</td>
<td>1</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>A married man committed adultery</td>
<td>5</td>
<td>31</td>
<td>26</td>
</tr>
<tr>
<td>A man killed his wife during an argument</td>
<td>6</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>A woman engaged in prostitution</td>
<td>7</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>A man stabbed his wife with a knife during an argument</td>
<td>8</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>A male engaged in homosexuality</td>
<td>9</td>
<td>34</td>
<td>25</td>
</tr>
<tr>
<td>An individual abandoned religion and espoused atheism</td>
<td>13</td>
<td>36</td>
<td>23</td>
</tr>
<tr>
<td>An individual threw burning liquid in someone's face, which caused scars</td>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>An individual accused a woman of adultery without adequate proof</td>
<td>17</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>A single man committed fornication</td>
<td>18</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>A woman had an illegal abortion</td>
<td>22</td>
<td>35</td>
<td>13</td>
</tr>
<tr>
<td>An individual intending only to injure someone by throwing a stone accidentally killed him</td>
<td>25</td>
<td>6</td>
<td>19</td>
</tr>
</tbody>
</table>

The disagreements are stark. Notice, for example, that Kuwaitis rank a woman committing adultery as more serious than a man killing his wife, and they rank a male engaging in homosexuality as more serious than an individual who throws burning liquid in someone's face, causing scars. Americans, in contrast, seem relatively unconcerned about adultery and homosexuality, and relatively distressed about the killing of adulterous wives and acid attacks—although, again, there is significant disagreement across subcommunities in the United States on the former two.\(^{109}\)

---

\(^{108}\) Evans and Scott, 22 Criminol at 48-49 table 3 (cited in note 46). Seriousness is ranked on a scale from one to thirty-seven.

\(^{109}\) While Maryland, for example, has explicitly excluded spousal infidelity as "adequate provocation" and potential grounds for mitigation of murder to manslaughter, most states have not. Compare, for example, Md Crim Code Ann § 2-207(b) with Commonwealth v Schnopp, 417 NE2d 1213, 1215–16 (Mass 1981) (holding that the killing of a spouse can be voluntary manslaughter when it immediately follows the victim's oral admission of adultery).
C. The Morality Convention

What of the studies distinguishing moral and conventional wrongdoing? A series of critiques has been leveled against the moral–conventional studies, noting that the findings reported depend on carefully selecting the questions asked. As one group of researchers recently noted, “the range of transgressions involving harm that has been included in these studies is remarkably narrow,” typically involving “behaviors that would be familiar to youngsters, such as pulling hair or pushing someone off a swing.”\(^{110}\) This poses a problem because teachers across cultures discourage hitting, pulling hair, and so on, making it difficult to disentangle what is innate from what is learned.

Studies that varied the cultural frame, however, generated substantially different results. Many children, for example, hold clear and authority-independent intuitions about the wrongness of acts that do not fit the pattern of moral (rather than conventional) transgressions. Across many countries, for example, children were found to consistently rank a broad array of transgressions as serious independent of authority, including “privately washing the toilet bowl with the national flag,” “mixed-sex bathing,” “addressing a teacher by his first name,” and violating a number of religious rules.\(^{111}\)

Another set of studies has challenged the distinctness of moral harms. In one study of adults, for example, Daniel Kelly and his fellow researchers asked participants about a series of paired harms. Here is one:

(1A) Mr. Williams was an officer on a cargo ship 300 years ago. One night, while at sea, he found a sailor drunk at a time when the sailor should have been on watch. After the sailor sobered up, Williams punished the sailor by giving him 5 lashes with a whip.

Is it OK for Mr. Williams to whip the sailor? [Yes/No] On a scale from 0 (not at all bad) to 9 (very bad), how would you rate Mr. Williams’ behavior?

(1B) Mr. Adams is an officer on a large modern American cargo ship in 2004. One night, while at sea, he finds a sailor drunk at a time when the sailor should have been monitoring the radar screen. After the sailor sobers up, Adams punishes the sailor by giving him 5 lashes with a whip.


\(^{111}\) Id at 120. See also Nucci, Education in the Moral Domain at 52–75 (cited in note 72) (discussing these studies in detail).
Is it OK for Mr. Adams to whip the sailor? [Yes/No]
On a scale from 0 (not at all bad) to 9 (very bad), how would you rate Mr. Adams' behavior?\textsuperscript{112}

Based on the notion that moral harms are intuitive and generalizable, one would expect a consistent answer to both questions. As indicated in Figure 1 below, however, this was not the case. Rather, most participants in the study indicated that it was “OK” to whip the sailor three hundred years ago, while only one in ten thought it was “OK” today. Participants also considered the two acts to be significantly different when evaluating the wrongfulness of the act (“how bad” the whipping was).

**FIGURE 1. JUDGMENTS ABOUT THE ACCEPTABILITY OF WHIPPING A DERELICT SAILOR\textsuperscript{113}**

![Bar graph showing acceptability and severity of whipping 300 years ago versus today.]

Similar variations were observed across several other scenarios, including the acceptability of abusing military trainees when prohibited and not prohibited by authority, the acceptability of eating the flesh of a dead person at a funeral when customary and when not customary, a teacher spanking students when prohibited and not prohibited, and practicing slavery in ancient Rome and in the United States.\textsuperscript{114}

The point here is not that we cannot or should not distinguish between more or less wrongful acts. Rather, it is to say that if the main


\textsuperscript{113} Id at 127–28 (presenting the results of an online survey conducted by the authors). The bar graph on the left shows the percent of “yes” responses to the binary “Is it OK?” question ($\chi^2 = 79.01; p = 0.000$). The bar graph on the right represents responses to the question: “How would you rate Mr. X’s behavior?” ($t(198) = 13.55; p = 0.000$).

\textsuperscript{114} See id at 126–28.
distinction that can be made between important and unimportant criminal offenses is the degree to which individuals agree on them, then the statement that people tend to agree about core offenses amounts to saying that people tend to agree about offenses about which they tend to agree. It is no help to say that what distinguishes "core" from "periphery" is the importance of the act to our collective survival if the only manner of discerning the importance of behavior to our survival is our degree of agreement about the behavior. It is, at its core (so to speak), a contentless distinction.

D. Playing Fair with Fairness Games

Similar problems have arisen with the studies of ultimatum games cited by Punishment Naturalists. Many anthropologists and economists were not satisfied with early research in the field; they noticed that the studies, while conducted across many countries, all focused on educated students in highly industrialized societies. They decided to take the same game further afield to see if the results in other cultures resembled those reported by researchers studying individuals who were well integrated into Western capitalist culture.115

What they found was revealing. Although something does move Responders to sacrifice what they might have gained from an unfair offer to punish the Proposer of the unfair offer, precisely what is considered fair and unfair varies significantly.116 For example, in societies where norms regarding equal distribution are strong, the Proposer is far more likely to propose something close to an even split than in societies where egalitarian distribution is not the norm; and if the Proposer in an egalitarian culture offers a lopsided split benefitting herself, the Responder is highly likely to reject the proposal, sacrificing her own share to punish the Proposer.117 But norms regarding fair distribution are far from universal; many societies demand egalitarian sharing while others feature intricately delineated social hierarchies. Consider Figure 2 below, which graphs the wide variation in offers made by people occupying the role of the Proposer across fifteen societies with


116 This does not mean that people necessarily conform their behavior to what is considered fair because they are intrinsically motivated to be fair. It might be the case that individuals adjust their behavior strategically in order not to be punished for what they believe others will perceive as unfair behavior.

differing cultural attitudes and a variety of levels of integration into world markets.

**Figure 2. Modified Box Plot of Ultimatum Game Offers Across Fifteen Societies**

One parsimonious explanation for these differences features variations in social meaning. Where there is an expectation of egalitarian sharing, a proposal that disproportionately rewards one individual at the expense of another will seem untoward and worthy of punishment. (Why should one person expect to gain more than another from this arrangement?) In others, it will seem quite reasonable and sensible (after all, no matter what the offer, the Responder will be getting

---

118 Joseph Henrich, et al, “Economic Man” in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies *54 figure 2 (working paper), online at http://www.som.yale.edu/Faculty/keith.chen/negot.%20papers/CamererEtAll_CrossCultUltimatum01.pdf (visited May 1, 2010). For the published version of this paper, see note 60. The box gives the interquartile range for offers in each society. The vertical line within each box, except for the Machiguenga, is the mean offer, not the median as in a standard box plot. The mean offer for the Machiguenga lies outside of the interquartile range and is represented by the vertical line just to the right of the box.
something for nothing). In both cases, the sense of fairness (or lack thereof) and the desire to accede (or to punish) will turn on local customs that prescribe what form fairness takes.

The Lamalera, for example, were among the most generous ultimate players—with Proposers typically offering up half or more than half of the money.119 They are subsistence whale hunters who share their bounty communally, often dividing a single whale among a hundred or more individuals—many of whom help maintain the boats, dry and cook the meat, or conduct other important non-hunting-related tasks in the village. Subsequent to a whale hunt, then, the dozen men who ventured out into the sea to catch the whale will typically take only a small portion of the whale they have caught for themselves. For the Lamalera, the most important part of their subsistence economy requires regular partitioning of goods in ways that might seem foolish to Westerners, but which make sense when considering the other benefits that accrue to whale hunters as a result.120

The Machiguenga, by contrast, hunt and gather small amounts of food, largely for themselves and their immediate families. Hunters (typically men) eat first and typically take the most, followed by women and children who eat whatever remains.121 Observers speculate that this is because men expend a significantly larger amount of energy hunting and gathering than the women, but, whatever the reason, the practice appears to instill a very different norm regarding the fair division of goods. The Machiguenga Proposers appear, by Western standards, exceptionally selfish, typically offering only a quarter of the money. Ethnographers, however, describe them as kind, decent, and thoughtful; they simply have, the researchers suggest, a different understanding of what fairness and generosity entail.122

Among the Gnau and Au of Papua New Guinea, another set of norms prevails. With extensive reciprocal demands made of one another (individuals are often expected to give away or share common possessions and goods on demand), receiving a gift is seen as incurring a kind of burden or debt. Because individuals are expected to reciprocate gift-giving or incur significant social costs, they are reluctant to accept offers of gifts that they feel will place a serious potential burden on

120 See Alvard and Nolin, 43 Curr Anthro at 540 (cited in note 119).
them. So while an individual may achieve social status through generous giving, others may reject gifts in order to resist indebtedness and its attendant lower status. Among the Gnau and Au, the Ultimatum Game generated an exceptionally high number of “fair” (50 percent) and “hyper-fair” (more than 50 percent) offers. But hyper-fair offers were often rejected. The suggestion by researchers is that this reflected the common aversion to accepting overly generous gifts.

Hyper-fair offers are not rejected in all societies, though. Among the Sukuma of southwestern Tanzania (who are not shown on the chart above), the most common offer was 90 percent, the mean was 61 percent, and no offer was rejected. Moreover, participants were willing to accept offers of as little as 10 percent, even though no person actually made an offer that low. Again, the researchers explain the results as reflecting local norms. Sukuma socialize their children to be extremely generous, requiring them to give away much of their food. They also have strong ingroup identifications and exceptionally generous responses to poverty, which anthropologists attribute to the stochastic nature of their agricultural economy and the necessity of pooling resources to survive.

If these studies suggest variation among social groups, another set of studies suggests that conceptions of fairness are socially contingent even within Westernized societies that are well integrated into the capitalist market system. In a series of studies conducted by Swee-Hoon Chuah, Robert Hoffman, Martin Jones, and Geoffrey Williams in the United Kingdom and Malaysia, what individuals thought was fair and the amount that individuals were willing to sacrifice to punish offers

\[123\] Id at 811. Those who repeatedly fail to reciprocate are shunned and disparaged. Consider id at 812.

\[124\] Id at 811. See also Herbert Gintis, et al, Explaining Altruistic Behavior in Humans, 24 Evol & Hum Behav 153, 159 (2003) ("[T]his reflects Melanesian culture of status-seeking through gift giving. Making a large gift is a bid for social dominance in everyday life in these societies, and rejecting the gift is a rejection of being subordinate."). The hypothesis offered by Gintis and his coauthors is consistent with the extensive anthropological literature on reciprocal exchange in many societies, with "gifts" being thought of as conferring status on the giver and a burden on the recipient. See, for example, Marcel Maus, The Gift: The Form and Reason for Exchange in Archaic Societies 65 (Routledge 1990) (W.D. Halls, trans) (originally published 1950) ("The unreciprocated gift still makes the person who has accepted it inferior, particularly when it has been accepted with no thought of returning it.").


\[126\] Id at 431.

they thought of as unfair varied significantly with the values they espoused. Individualism, desire for gender equality, and a host of other values related to market participation strongly influenced the behavior of participants, suggesting that cultural mores within each society were important determinants of fairness as well.128

How do naturalists account for this kind of variation in fairness games? For the most part, they either ignore or gloss over the data. One strategy, employed by Hauser, is to offer a substantially diminished version of naturalism. On this "softer" naturalism account, we have not evolved any specific intuitions; rather, we have moral "principles" like fairness, and culture sets the "parameters" that tell us what is fair and what is not. The "principles and parameters" approach to a universal moral grammar is employed as an analog to Chomsky's "principles and parameters" approach to constructing a universal grammar of human language. The problem with this tack is that on this softer account, nature asks culture to do all the work. If fairness can be whatever culture supplies, then it is not clear what work the hypothesized moral organ is doing.

The alternative approach (also employed by Hauser at times) is little better. Arguing that nature sets specific limits on our conception of fairness requires a specification of what those limits are. Hauser, as we noted above,129 very conveniently chooses 15 percent and 50 percent as the lower and upper bounds—precisely the limits observed in the studies he had read at the time! If that is how one determines the limits set by our common moral organ, then it is certainly true that it will (as a matter of logic) always accurately reflect observed data; but it loses any explanatory or predictive force. It also cordons off as "parameters" the richness of the social meanings and practices that give rise to norms governing fairness, sharing, reciprocity, and punishment. Naturalism can tell us nothing about why we have different intuitions from the Quichua, or why the Quichua have different intuitions from the Sakuma. For that, we need a theory that incorporates variations in social norms.

Our point here is not that people's reactions are random or without structure—quite the reverse. There is a deep but highly generalized structure—individuals are willing to make significant sacrifices to punish those they believe are being unfair—but that structure relies upon socially constructed norms to give it content. Without recognizing the way social meaning provides for the specific articulation of that structure, it is impossible to give a coherent interpretation of the

129 See note 68 and accompanying text.
data. As the researchers involved in conducting these cross-cultural
studies concluded, "[f]ailure to recognize the extent of human diversi-
ity and the range of processes that have generated the human mosaic[]
may doom large sections of social science to an empirically false and
culturally limited construction of human nature."130

III. REALISM VERSUS NATURALISM AND "CORE" OFFENSES

We doubt that naturalists will discover some independent way to
distinguish the core of harms from the periphery, moral transgressions
from conventional, or principles from parameters. But even if there is
some yet-to-be-discovered distinction, naturalists would still face a more
serious problem: there is substantial disagreement about what constitutes
wrongdoing and how serious given offenses are within the so-called
"core" of wrongdoing. As such, the claim that core offenses are noncon-
troversial requires not only that we ignore disagreement over what con-
stitutes core and noncore offenses, but also that we ignore significant con-
traversies within the three categories of core offenses: "physical aggres-
sion, takings without consent, and deception in exchanges."131

Comparing the abilities of realist and naturalist accounts to manage
both agreement and disagreement over the wrongfulness of physical aggres-
sion, takings without consent, and deception in exchanges, however,
requires at least a preliminary account of the realist perspective.

A. Punishment Realism

Punishment Realism, in our account, applies to the study of pun-
ishment the insights of classical legal realism and contemporary em-
pirical research into human judgment.132 Legal realism observes that
abstract concepts, doctrines, and rules of law do not provide unique,
determinate resolutions to most difficult cases, and that in deciding
such cases, legal actors—consciously or not—are necessarily moved by
extralegal influences that shape their choice of one or another of the
various possible justifications and outcomes. For the most part, these
extralegal influences will move legal actors to agree, but sometimes
they will move them to disagree.

Realists just want to know what those extralegal influences are
and how they manifest themselves so that they can better predict legal
outcomes and manipulate policy to enhance whatever social welfare,

131 Robinson and Kurzban, 91 Minn L. Rev at 1892 (cited in note 13).
132 See generally Braman and Kahan, Legal Realism as Psychological and Cultural (Not
Political) Realism (cited in note 3).
fairness, or expressive concern they favor. As such, realists want to understand the cognitive biases and heuristics that move individuals to interpret law and facts in particular ways. But where the naturalist account describes highly specific moral intuitions, the realist account emphasizes the interplay between relatively generic cognitive mechanisms and varied social meanings. On the realist account, cognition is, to be sure, shaped by a host of demonstrable and perhaps nearly universal cognitive biases and heuristics, many or all of which are the product of evolutionary pressures or accidents. But Punishment Realism, at least as we conceive of it, views these innate cognitive traits as interacting with and generating a variety of social meanings that ultimately determine our understanding of and reaction to wrongdoing.

Punishment Realism recognizes that intuitions about wrongdoing and punishment like these will often seem natural and universal even when they are, in fact, socially contingent. Perhaps the most obvious way that individuals come to see their own parochial conceptions of justice as natural and universal can be described in terms of explicit value preferences: individuals simply prefer their own value hierarchies over those of others. Classic cultural clashes over sodomy, abortion, slavery, and many other issues are often described in these terms: participants may recognize that the moral hierarchies of others vary, but they are unlikely to prize other people’s mores and commitments more highly than their own; at best they may view other value structures as strange or foreign, at worst as false and debased. And while those involved in such moral disputes may understand that their preferred outcomes derive from their values, they will often have trouble articulating the source of their values. Their values will seem, at least to them, to be natural.

---

133 There are a number of accounts that might fit this description. For a comparison of the two main accounts, see Dan M. Kahan, “Ideology in” or “Cultural Cognition of” Judging: What Difference Does It Make?, 92 Marq L Rev 413, 422 (2009).

134 Consider Mary Douglas and Aaron Wildavsky, Risk and Culture 72-73 (California 1980) (making the analogous point that private individuals “choose not to be aware of every danger,” and that when choosing between risks, “subjective values must take priority”).

135 That our values are not universal or transcendental, but historically specific intuitions of our collective making, is a perspective well described by Stanley Fish:

I intend [the title of the book Doing What Comes Naturally] to refer to the unreflective actions that follow from being embedded in a context of practice. This kind of action . . . is anything but natural in the sense of proceeding independently of historical and social formations; but once those formations are in place (and they always are), what you think to do will not be calculated in relation to a higher law or an overarching theory but will issue from you as naturally as breathing.

But the overt privileging of one’s values over those of others often masks a subtler and even more pervasive way that individuals come to see their own intuitive sense of justice as natural and universal: cultural cognition. Cultural cognition refers to the tendency of individuals to conform their perceptions of risks and their factual beliefs to their core cultural commitments. It is cognitively easier to believe factual assertions that comport with our norm-pervaded moral evaluations and cognitively harder to believe those that conflict with or threaten them.

Numerous studies have shown that culture implicitly shapes factual perceptions in this way, shaping our beliefs without our noticing that it is doing so. Culture constructs our understandings of fact both through cognitive mechanisms (such as avoiding cognitive dissonance, the tendency of individuals to discount information that conflicts with their existing beliefs and values) and social practices (such as selecting information sources like favored news outlets and friends who share our values). Individuals with varied and durable conceptions of what is noble and what is base thus form equally varied and durable conceptions about what is true and what is false. As a result, even where individuals are willing to agree to a single legal standard that requires specific factual findings (as jurors must), a host of cognitive biases and heuristics can move them to conform their understanding of relevant facts so that they arrive at varied appraisals of wrongfulness.

These two forms of cultural influence—one explicit and one implicit—are often mutually reinforcing. Because individuals tend to credit factual claims that are consistent with their normative visions of a just social order, when they reflect on their cultural commitments they have plenty of facts to suggest that their worldview is naturally preferable to others. And, because their cultural commitments will seem naturally preferable to them, they are less likely to question these commitments or their influence on their factual perceptions.

We have more to say below about Punishment Realism and the various social and cognitive mechanisms that sustain it, but with that brief summary in hand, we turn to crimes within the so-called “core of wrongdoing.”

B. The Core Offenses

There are recurrent themes in the kinds of acts that are prohibited in many cultures. Robinson and Kurzban have helpfully collected them under the rubrics of “aggression, takings without consent, and deception in exchange,” and argue that acts falling into these categories constitute “the core of wrongdoing.” As we argue below, none of these categories is composed of acts free from dissensus, and the nature of the systematic dissensus that pervades each of these categories is
at least as interesting and informative as any agreement that can be found. We start with “takings without consent” and “deception in exchange,” two easy cases. Then we take on rape, followed by the hardest case—the core of the core of wrongdoing, so to speak—murder, around which we develop our argument in greater detail.

1. Takings without consent.

The anecdote at the beginning of this Article provides some measure of the problems faced by naturalist claims about a universal or normal intuition regarding takings without consent. A substantial part of the dissensus over takings relates to varied conceptions of property. As the anthropologists Franz von Benda-Beckmann, Keebet von Benda-Beckmann, and Melanie Wiber describe, the concept of property can be thought of as depending on three variable concepts: “first, the social units (individuals, groups, lineages, corporations, states) that can hold property rights and obligations; second, the construction of valuables as property objects; and third, the different sets of rights and obligations social units can have with respect to such objects.”

As we noted in the Introduction, even where everyone is in agreement on the idea that someone owns something of value—a train ticket, for example—that individual’s rights and obligations can vary dramatically from context to context. In the United States, we typically have highly individualistic conceptions of rights and obligations—at least relative to those living in India and Papua New Guinea. As a result, an act that would be considered an invasion of some property-like right in one time or place can seem perfectly normal in another because the norms governing who has access to what and under which conditions vary so dramatically across time and place.

The anthropological literature on non-Western cultures provides ample illustration of this, but we need look no further than recent

---


137 Anthropologists have come to view the notion of property as often contested:

[P]eople have at any given moment a number of “languages” available to them for characterizing objects in circulation as commodified, gift-like, inalienable, and so on. These languages are often in tension; actors also have differential access to them. And they use these languages within a context that may constrain the use of some idioms and support the use of others. This perspective . . . helps us understand how multiple or hybrid forms of value occur simultaneously.

Elizabeth Emma Ferry, Not Ours Alone: Patrimony, Value, and Collectivity in Contemporary Mexico 18 (Columbia 2005). David Graeber and Maurice Godelier provide two recent and influential general accounts of the way value and property vary across time and place. See generally David
domestic debates (over, for example, the law of taxation, in which the ownership of a good that is highly valued—money—is continually contested; adverse possession, nuisance, or intellectual property, on which the members of the public, the academy, and the bench regularly disagree) to get a flavor of the sticky disensus over "takings without consent" closer to home.

It is no help to modify the naturalist account by suggesting that, although the idea of property varies, the notion does not vary that transgressions of those rights are intuitively wrong. Without providing content to the rights themselves, this simply passes along the cognitive puzzle of what is wrongful to local norms governing what exactly it is that comprises a property right. This is not to say that there may not be some very general traits that humans share with respect to affection for various possessions. The question is whether we have universal intuitions about when an act is theft and, if so, how wrongful it is relative to other acts. And that is something that simply cannot be resolved without reference to variable social norms.

This is a modified version of Jerry Fodor’s "input problem" for evolutionary theories that rely on multiple cognitive modules of this sort. If the argument is that we have a module that helps us quickly compute a judgment such as "theft is wrong," we need to have some sense of when something qualifies as theft. But the definition of theft (or fraud or murder) is fairly complex and socially contingent in evolutionary contexts, depending on social groups, status, and a host of other concerns. Because you need complex social information to assess whether something is theft, you have not really bought any cognitive efficiency with a module that tells you that theft is wrong, because


139 As Janice Nadler and Shari Diamond have found in their research, variable concepts like “subjective attachment to property” are paramount in shaping “the perceived justice of a taking.” Janice Nadler and Shari Seidman Diamond, Eminent Domain and the Psychology of Property Rights, Proposed Use, Subjective Attachment, and Taker Identity, 5 J Empirical Legal Stud 713, 713 (2009). See also, for example, Kelo v City of New London, 545 US 469 (2005), which provoked public debate over property rights and takings without consent.

140 Special thanks to Stephanie Stern for making this point to us.

141 See, for example, Richard C. Stedman, Toward a Social Psychology of Place, 34 Envir & Behav 561, 563 (2002).

the same complex social information that tells you that something is theft can also tell you that it is wrong and how wrong it is.

It is, of course, possible to move away from specifics and develop general rules on which people agree. But such definitions invariably depend on some variable notion of what makes something wrongful. Taking without consent and deception in exchange are thus not always wrong; rather, they are wrong when social customs tell us so.

2. Deception in exchange.

If ever there was a messy and discordant conception in law and morality, it is that governing “deception in exchange.” Punishment Naturalists assert that deception in exchange is one of the core areas of agreement in our moral development. They suggest that the moral norm against deception arises because of the “analogical closeness to inflicting direct personal harm on another” or because it is an “extremely useful mechanism for a society to develop”:143 intuitions like these enable cooperation because they punish defectors and cheats.144

We agree that across societies, individuals exhibit a general (and widely shared) dislike of shirking and fraud. The positive version of this dislike is instantiated in the norm of reciprocity.145 But that general principle falls apart at the level of specificity at which the law typically operates. Individuals, it turns out, have quite divergent views about whether specific kinds of lying are wrongful, and how far the law ought to go to protect buyers in commercial exchanges from their own bad judgment in relying on a seller’s speech. Because these are the live issues in the criminal regulation of deception, we briefly explore such divergent views here.

Consider first the definitional problem. For example, judges commonly take from the jury actions for civil or criminal fraud in sale of goods cases where the seller has “puffed” her goods.146 This alone suggests that the general principle “do not lie in commercial exchange” has

---

143 Robinson and Darley, 81 S Cal L Rev at 58–59 (cited in note 15).
144 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1647–49 (cited in note 14).
146 Many of the issues in puffery, such as questions of falsifiability in false advertising cases, are often resolved as matters of law rather than fact. See Jean W. Burns, Confused Jurisprudence: False Advertising under the Lanham Act, 79 BU L Rev 807, 867–71 (1999); Ivan L. Preston, The Definition of Deceptiveness in Advertising and Other Commercial Speech, 39 Cath U L Rev 1035, 1040–41 (1990).
a variety of concrete meanings. The Supreme Court itself has advanced distinct—and often competing—definitions of what the term "misleading" means. Both jurists and lay people simply do not have stable preferences about what constitutes "deception." Instead, their views about whether speech in fact deceives turn on their implicit views of whether it should.

One might be tempted to believe that these various rules on the meaning of deception turn on a general empirical finding that individuals do not believe sales talk. But, as a number of studies have found, "puffery is believed by large numbers of consumers," though not all. In this way, the law of fraud is full of conflicts over values—do we want individuals to bear the responsibility for their own choices, or do we want individuals to recognize context and market power as important—masquerading as disputes about fact. That dissensus in turn produces the hotly disputed political fights regarding the law of deception in exchange we see all around us, including the scope of the securities laws and the appropriateness of most forms of consumer protection regulation.

147 See David A. Hoffman, The Best Puffery Article Ever, 91 Iowa L Rev 1395, 1400–16 (2006) (defining the puffery defense in false advertising, securities, UCC warranty, and promissory estoppel cases). See also Ivan L. Preston, Puffery and Other "Loophole" Claims: How the Law's "Don't Ask, Don't Tell" Policy Condomes Fraudulent Falsity in Advertising, 18 J L & Commerce 49, 54–55 (1998) (noting that puffery may take a variety of forms, including claims that a product is the "best," "best possible," "better," and "specially good"). Hoffman provides a host of instances where intuitions vary on seemingly similar cases:

Advil's claim that it, "like Tylenol," "doesn't upset the stomach" was found not to be immune puffery because a court believed that consumers would have viewed the statements to be a factual comparison with other brands. Similarly, a motor-oil company's claim to provide "longer engine life and better engine protection" was not held to be puffery. By contrast, a puffery defense succeeded with respect to Bayer's statement that it made the "the world's best aspirin" that "works wonders." And a videogame manufacturer escaped liability, despite claiming to have made "The Most Advanced Home Gaming System in the Universe." ... The claim that yogurt is "nature's perfect food" apparently may be falsified and is not puffery. But, to enthusiasts' chagrin, Nestlé's boast that it sells the "very best chocolate" is a meaningless puff. If, upon eating too much chocolate yogurt, one needed a diet, the makers of topical gel could be liable for claiming to "dramatically interfere with the process of converting calories to fat" and "inhibit the creation of new fat cells." But, the makers of a weight-loss pill trumpeting the drug's ability to cause you to "Lose Weight Fast" would be protected.

Hoffman, 91 Iowa L Rev at 1404 (alterations omitted).


150 See Hoffman, 91 Iowa L Rev at 1442 (cited in note 147).
Thus, while it may be "intuitively easy to make the connection between physically taking property and physically harming another,"\textsuperscript{151} the category of "deception in exchange" lacks legal coherence. Rather, some kinds of deception, in certain circumstances, are actionable and morally wrongful. Sometimes, violators of the "norm" of reciprocity are held to be legally responsible by most of the population, sometimes by only part of the population, and sometimes by only a small minority.\textsuperscript{152} The contingency of the finding turns on individuals' views of what we owe to one another as citizens, the degree to which individuals should be responsible for their own flourishing or should turn to social systems for protection, and the amount of freedom we ought to permit speakers to falsely extol or mislead by omission.

3. Rape.

The law of rape has been a site of intense legal and political conflict for over thirty years, and "date" or "acquaintance rape" has been at the center of that dispute.\textsuperscript{153} In particular, those involved in the debate disagree over how the law should deal with cases in which a woman engages in "verbal resistance"—that is, says "no"—but does not display the form or quantum of "physical resistance" demanded by the traditional, common law definition of rape.\textsuperscript{154} Arguing that the law's resistance to convicting in such cases leaves women unprotected from one especially common form of coerced sex, feminist and other reformers have successfully attained a variety of reforms. All jurisdictions have now adopted evidentiary rules that prohibit proof of a complainant's "sexual history" designed to show a propensity to consent.\textsuperscript{155} Some, but not others, have modified elements of the traditional common law definition of rape, such as elimination of the "force or threat of force" element or the reasonable mistake of fact defense.

\textsuperscript{151} Robinson and Darley, 81 S Cal L Rev at 58–59 (cited in note 15) (arguing for the "analogical closeness" of these categories).
\textsuperscript{153} See generally Susan Caringella, Addressing Rape Reform in Law and Practice 12–17 (Columbia 2009); Stephen J. Schulhofer, Unwanted Sex: The Culture of Intimidation and the Failure of Law 29–46 (Harvard 1998).
\textsuperscript{154} Compare, for example, Estrich, Real Rape at 102–03 (cited in note 9) (arguing for a "no means no" standard in the law) with D.N. Husak and G.C. Thomas, Date Rape, Social Convention, and Reasonable Mistakes, 11 L & Phil 95, 122–25 (1992) (arguing against a standard that treats a verbal "no" as sufficient).
\textsuperscript{155} See, for example, FRE 412.
relating to consent, aimed at forcing judges and jurors to treat "no" as "no" for purposes of rape law. These reforms, however, seem to have had little impact in practice and continue to generate scholarly and political debate.

Punishment Naturalists have voiced skepticism, if not hostility, toward such reform efforts. They identify the imposition of "strict liability in cases where culpability may be difficult to prove, but is likely to exist" as an ill-considered departure from what are asserted to be shared intuitions. Among "the reform programs" that they identify as involving "criminal law manipulation . . . to alter people's intuitions of justice" is the "attempt[] to eradicate the notion that women often pretend to withhold consent to intercourse to appear more alluring or simply to avoid appearing 'promiscuous,' rather than as a genuine indication of not wanting to engage in sexual activity." But nothing in their carefully conducted empirical studies of shared intuitions in fact supports the sort of conservative stance toward reform efforts that these comments imply.

The Punishment Naturalists conclude that rape is among the "core" forms of "wrongdoing" that are "the subject of nuanced and specific intuitions that cut across demographics." The evidence consists of multiple studies showing that demographically diverse individuals are highly likely to agree that "rape" should be punished and is a more "serious" form of wrongdoing than various other offenses. The Punishment Naturalists have themselves found that subjects tend to regard "rape" as more serious than imposition of mere "unwanted sex."

It is simply not possible to derive from this evidence any reason to be skeptical, much less any reason to oppose, date-rape reform efforts. There might well be "consensus" that rape should be punished and is "worse" than inducing another to engage in "unwanted sex." But there most manifestly is not consensus in American society on how "rape" should be defined, and in particular whether a man who engages in sex with a woman who repeatedly tells him "no" before and during intercourse has committed "rape" or merely succeeded in achieving "unwanted sex," or over how severely, if at all, to punish such

156 See, for example, Schulhofer, Unwanted Sex at 30–33 (cited in note 153).
159 Robinson and Darley, 81 S Cal L Rev at 46–47 (cited in note 15).
160 Id at 52.
161 Robinson and Kurzban, 91 Minn L Rev at 1892 (cited in note 13).
162 See id at 1853 & n 100, 1856 & n 123, 1859 & nn 135, 138.
163 See id at 1885–88.
a man. Indeed, the Punishment Naturalists themselves have reported finding less consensus surrounding whether and how much to punish "unwanted sex," presumably because people do disagree about whether it should be regarded as a crime, and, if so, as serious a crime as rape.  

The source of that disagreement is cultural. Psychologists and sociologists specializing in women's studies have shown that disagreements over a host of beliefs and attitudes toward rape correlate with competing sets of moral norms—one that is "hierarchical" in nature and prescribes highly stratified gender roles, and another that is more "egalitarian" and rejects the proposition of separate male and female spheres in society. A recent mock-juror study conducted by the Cultural Cognition Project at Yale Law School found that the outlooks individuals subscribed to predicted high levels of disagreement over whether a man should be found guilty of rape in a case patterned on Commonwealth v Berkowitz, a "no means yes" acquaintance rape case that provoked intense political controversy in the 1990s and that continues to be featured in scholarly commentary. The disagreement among ordinary citizens over such cases can be linked to a form of cultural status competition insofar as both hierarchical and egalitarian individuals perceive that the stance the law adopts on this issue will align it with the norms of one or the other cultural group. Indeed, the group most resistant to and resentful of reform of the common law of rape, the study found, consisted of hierarchical women (particularly older ones), whose high social status is most conspicuously tied to continued public endorsement of the traditional, but not bitterly contested, norms of sexuality. This controversy is fraught with difficult issues. Should the law weigh in on the side of those who want to make "no" mean "no" for purposes of rape law as a means of promoting egalitarian norms? Or would that be an inappropriately partisan and illiberal application of law to promote a moral and cultural orthodoxy? Alternatively, if the law resists demands for change, is it not siding with the hierarchical position, effectively endorsing that position's understanding of idealized gender norms? If the law is to be made to take a side in this debate, how can it do so effectively? If it wants to be genuinely neutral, what stance would effectively communicate that intention? These questions cannot even be framed intelligibly, much less answered satisfactorily, by any

164 See id at 1890–91 & n 230.
165 See, for example, Michael W. Wiederman, The Gendered Nature of Sexual Scripts, 13 Fam J 496, 499 (2005); Martha R. Burt, Cultural Myths and Supports for Rape, 38 J Personality & Soc Psych 217, 225 (1980).
166 641 A2d 1161 (Pa 1994).
167 See Kahan, 158 U Pa L Rev at 734 (cited in note 158).
theoretical framework of criminal law that insists on confining its attention to issues on which there is "consensus"—issues that are peripheral to urgent, pressing debate whether or not they might be said to be at the "core" of a set of "shared intuitions."

Indeed, a framework that fails to acknowledge or recognize that it is addressing issues at the practical and political periphery of criminal law can easily generate unreliable explanations and prescriptions. One Punishment Naturalist, Owen Jones, posits (on the basis of extrapolation from sociobiological theory) differences in "male and female brains" that cause them to "process rape victimization differently," with the latter predisposed to take it much more seriously because of the impact it had in disrupting "female mate choice in ancestral environments." Jones surmises (on the basis of further conjecture) that these ingrained biological differences are the likely source of the inefficacy of rape law reforms and identifies (without necessarily endorsing) various reforms aimed at making the biological foundations for male–female disagreements manifest, thus promoting greater resolve on the part of the legal system to convict rapists and punish them more severely.

As fascinating and insightful as it is, this account will not be of much use to anyone earnestly engaged in trying to understand and promote morally appropriate solutions to the existing debate over rape law reform. The one feature of this account that admits of empirical examination—its assertion that the inefficacy of rape-reform laws stems from male and female differences over the seriousness of rape victimization—is contrary to all the available evidence. Indeed, without (as far as we know) following any of Jones’s strategies for remediating a deficiency in how seriously men take the harm of rape, the law has made progress in reducing the incidence of violent stranger rape comparable in degree to the progress it has made in reducing many other forms of common crimes, including homicide, in recent decades. The form of rape that apparently has evaded reduction is exactly the type—date or acquaintance rape—at which the "no means no" reforms have been directed. As explained, the force that has limited

169 See id at 917–20.
170 See Bureau of Justice Statistics, Key Facts at a Glance: National Crime Victimization Survey Violent Crime Trends, 1973–2008, online at http://bjs.ojp.usdoj.gov/content/glance/tables/viostrtatdb.cfm (visited Oct 22, 2010) (reporting a drop in the incidence of rape from 2.8 per 1,000 to 0.4 per 1,000 between 1979 and 2004).
171 Clay-Warner and Burt, 11 Violence Against Women at 167 (cited in note 157) (concluding that the incidence of reporting of "simple" or acquaintance rape, as opposed to "aggravated" or stranger rape, has not changed since the 1970s).
those reforms is not biological but cultural. The conflict here is not one originating (in brains, genes, or anything else) between men and women; it is between men and women of one cultural outlook and men and women of another. Indeed, the group with the narrowest apprehension of what “rape” is—the one most likely to see “no” as meaning “yes”—consists of hierarchical women. What reason is there for supposing that “contextualizing women’s emotional reactions to rape within the evolutionary processes” would help make them more likely as citizens to support rape law reform or to vote as jurors to convict under reform statutes once they are enacted?

Indeed, far from helping to advance the cause of those who want to reduce the incidence of acquaintance rape, Jones’s attempt to derive guidance from the (conjectured) sociobiological differences in men’s and women’s apprehensions of the harm of rape are more likely to obstruct it. Jones, for example, argues against the enactment of “sexual assault” statutes, apparently unaware of the role that such statutes are intended to play in norm reform: calling nonconsensual sex that is unaccompanied by force or threat of force “sexual assault” is less likely to trigger resistance to punishing forms of “unwanted sex” that some men and women condemn but do not regard as “rape”; and by assuring at least some degree of punishment for such behavior now, such statutes make it more likely that in the future more men and women will join the ranks of those who already regard such behavior as “rape” and who see it as meriting designation and punishment as such. How successful this strategy has been, and whether it is otherwise morally appropriate, are matters of reasonable debate. But the only arguments that will contribute meaningfully to that discussion are the ones that come to grips, in an empirically informed way, with the real cultural differences in individuals’ understandings of what “rape” is.

Perhaps naturalists’ conceptions of takings without consent, deception in exchange, and rape are complicated in these ways because these offenses are less central to the naturalist conception of “the core of wrongdoing” than offenses involving the most serious form of physical aggression: killing. In what follows, we focus on the most serious of wrongdoings in this class: murder. If naturalism is to prevail anywhere, surely it should be with the most serious crimes in our legal repertoire.  

172 See Jones, 87 Cal L Rev at 918 (cited in note 168).
173 See Schulhofer, Unwanted Sex at 104–05 (cited in note 153); Kahan, 158 U Pa L Rev at 752 (cited in note 158).
175 Naturalists themselves agree with this prioritization of violence over other wrongs. See, for example, Robinson, Kurzban, and Jones, 60 Vand L Rev at 1635 n 5 (cited in note 14).
4. Murder.

Punishment Naturalism holds that people agree both on what constitutes murder and on how serious a given murder is relative to other potentially bad acts. Punishment Realism, in contrast, holds that while people agree on many cases (for example, the heinous kidnapping-torture-murder case described above175), they also frequently disagree about both whether an act is so wrong as to be criminal and, if it is, how serious the criminal offense is.

a) Ambiguous agreement. Take, as an example, Robinson and Kurzban’s study showing that people generally agree that the killer in the first of the following vignettes is guilty and should be punished while the killer in the second is innocent and should not:

SCENARIO A: John knows the address of a woman who has highly offended him. As he had planned the day before, he waits there for the woman to return from work and, when she appears, John shoots her to death.177

SCENARIO B: John is knocked down from behind by a man with a knife who moves to stab him. As the man lunges for him, John stabs him with a piece of glass he finds on the ground, which is the only thing he can do to save himself from being killed. The man later dies of his injuries.178

We do not doubt that there is little disagreement over either claim in cases like these in the contemporary United States. Most people will define the former as a crime and the latter as not a crime; and even where they do believe the latter to be a crime, they consistently rank it as less serious than the former. But what underlies this consistency?

Naturalism and realism both furnish explanations for this agreement. The naturalist explanation features evolutionary pressures: if we did not agree on the wrongfulness of taking human life, our existence would be—at least relatively speaking—nasty, brutish, and short. Collectively, then, humans who intuitively viewed this kind of aggression as wrong and deserving of punishment were more likely to survive as a group; those who did not were less likely to survive. The result was a gradual growth of human sociality;179 but this sociality should be thought of not as an agreement or an implicit norm to which people

176 See text accompanying note 36.
177 Robinson and Kurzban, 91 Minn L Rev at 1897 (cited in note 13).
178 Id at 1894.
179 See Robinson, Kurzban, and Jones, 60 Vand L Rev at 1646–54 (cited in note 14) (arguing that human social impulses laid the foundation for a predisposition to acquire shared intuitions of justice, which provided evolutionary benefits to the individuals bearing them).
are acculturated, but as a modest but ever increasing biological disposition to view these acts as wrong and deserving punishment.

Realism, by contrast, explains this agreement as the product of shared social meaning, an agreement deriving from cultural norms that are widely shared in our society. While both scenarios are quite vague, they provide enough for readers to form a picture of the defendant in each case and to evaluate, relative to socially constructed norms, the moral quality of each act. On this account, the impulse to view the first act as wrong and the second as innocent stems not from an innate moral organ, but rather from shared socialization. This does not mean that there are not innate cognitive mechanisms at work, but rather that they are quite general and allow for the construction of social meanings that may vary substantially. Nor does it mean that humans will necessarily disagree: if they are similarly socialized, then they will in all probability evaluate the social meaning of these acts in similar ways.

Both accounts fully explain the lack of variation found by Robinson and Kurzban on this item. And, if humans always agreed on what distinguishes a good from a bad killing, it would be impossible to figure out which of the two accounts furnishes a better explanation of the available data. To distinguish between the two accounts, then, we have to alter the scenarios such that the social meaning of an act is in dispute. We could then see if variations in cultural outlooks explained variations in appraisals of guilt.

In what follows, we do that. We look first at examples in which there are explicit disagreements over which standards should govern what constitutes a serious wrong. We then look at instances in which, even when individuals accept a single standard, they disagree over which acts meet the standard.

b) Disagreement over standards. One way to evaluate these two accounts is to ask whether there have been cultural regimes in which the meaning of these acts varied. History, as it happens, furnishes many such examples; we describe just a few.180

While the contemporary formulation of self-defense doctrine addresses persons in universal terms, supplying a unitary standard that makes no reference to the social identities of the persons entitled to use deadly force or those against whom they are entitled to use it,181 this was

---

180 These are drawn from Kahan and Braman, 45 Am Crim L Rev at 3 n 2 (cited in note 34).
181 The traditional standard is couched in terms that are reflected in nearly every jurisdiction in the United States today: a person who has not otherwise provoked aggression is entitled to resort to deadly force against another (and hence is protected from criminal liability for doing so) when she honestly and reasonably believes that deadly force is necessary to prevent an imminent threat of death or great bodily harm to herself. See Wayne R. LaFave, 2 Substantive Criminal
not always the case. Historically, many societies conditioned the use of deadly force in self-defense on membership in a privileged class. 182

One need not leave the United States to find such contentious social meanings: the law in the antebellum American South also made such distinctions, denying blacks the authority to use deadly force to protect themselves from deadly assaults by whites and affording whites greater authority to use deadly force against blacks than against fellow whites. 183 As Justice William Brennan noted in his famous dissent in McCleskey v Kemp 184 during the colonial period, “black slaves who killed whites in Georgia, regardless of whether in self-defense or in defense of another, were automatically executed,” but “a person who willfully murdered a slave was not punished until the second offense, and then was responsible simply for restitution to the slave owner.” 185

What would members of that historical moment have made of the following vignettes?

MODIFIED SCENARIO A: John owns a slave who has highly offended him. As he had planned the day before, he waits for his slave to return from work and, when he appears, John shoots him to death.

MODIFIED SCENARIO B: Joe, a slave, is knocked down from behind by his owner, John, who moves to stab him. As John lunges for him, Joe stabs him with a piece of glass he finds on the ground, which is the only thing he can do to save himself from being killed. John later dies of his injuries.

While it would be hard for naturalists to account for the distinctive understandings of these vignettes in colonial and contemporary American communities, realism offers a straightforward explanation for the observed variation. A realist account would describe the

---

Law § 10.4(b) at 145–47 (West 2d ed 2003). If the threat is of some lesser magnitude, a person may repel it only with nonlethal force. See id.


183 See, for example, A. Leon Higginbotham, Jr and Anne F. Jacobs, The “Law Only as an Enemy”: The Legitimization of Racial Powerlessness through the Colonial and Antebellum Criminal Laws of Virginia, 70 N C L Rev 969, 1042 (1992):

Despite the unrelenting punishments and beatings that a slave might receive at the hands of an overseer, an owner, or another white, there were only rare instances in which a slave might claim self-defense in the killing of a white person. Such cases generally involved whites of low socioeconomic background.


changing importance of social status within colonial and contemporary American society and the various rights and duties (or lack thereof) that accompanied status. Within a society that embraces such status distinctions, the social meaning of aggression by someone of stature and rank against a subaltern was distinct from the meaning of aggression by the subaltern against a person of status.  

Again, even if we limit our scope to the United States, it is well known that tolerance of the use of deadly force to protect various nonvital interests has varied significantly across time and place. Marks of social status, such as displays of deference in public space and male dominion over the sexual lives of wives and daughters in particular, are a conspicuous characteristic of communities guided by honor norms. For example, in many Southern jurisdictions in the United States, it was once the case that the paramour could not “lawfully defend himself against the husband’s violence, and stand his ground and shoot or cut in order to repel the husband’s attack upon him.” Again, we can modify the vignettes slightly to alter their social meaning in this historical context:

**MODIFIED SCENARIO A:** John knows the address of a man, Tom, who has offended him by implying he had sex with his wife. As he had planned the day before, he waits there for Tom to return from work and, when he appears, John confronts him and shoots him to death.

**MODIFIED SCENARIO B:** After implying he had sex with John's wife, Tom is knocked down from behind by John who moves to stab him with a knife. As John lunges for him, Tom stabs him with a piece of glass he finds on the floor, which is the only thing he can do to save himself from being killed. John later dies of his injuries.

In considering how varied the evaluations of these vignettes might be, consider the experiments conducted by Richard Nisbett and Dov  

---

186 Similar status hierarchies can be found in many societies. Consider, for example, the Tokugawa administrative code in feudal Japan that granted Samurai the privilege of “*kiri-sute-gomen,* that is the privilege of a *samurai* to cut down a commoner with impunity.” E. Herbert Norman, *Japan’s Emergence as a Modern State: Political and Economic Problems of the Meiji Period* 18 (Institute of Pacific Relations 1940). See also David B. Kopel, *The Samurai, the Mountie, and the Cowboy: Should America Adopt the Gun Controls of Other Democracies?* 30 (Prometheus 1992) (“Any disrespectful member of the lower class could be executed by a Samurai’s sword.”).

187 See Richard E. Nisbett and Dov Cohen, *Culture of Honor: The Psychology of Violence in the South* 32 (Westview 1996) (arguing that the culture of the South included a tolerance for the use of violence for the protection of vital as well as what would now be considered nonvital interests).

Cohen discussing honor norms that govern interactions like these in the contemporary United States.\textsuperscript{189}

In one particularly revealing study, Nisbett and Cohen told participants about a man named Fred and asked how justified Fred would be in fighting an acquaintance who had affronted him in some way. The first set of questions asked respondents how justified physical aggression would be if it were in response to another man who “looks over Fred’s girlfriend and starts talking to her in a suggestive way,” “insults Fred’s wife, implying that she has loose morals,” or “tells others behind Fred’s back that Fred is a liar and a cheat.” In a second set of questions, they asked whether Fred would be justified in \textit{shooting} the person who had committed certain “more serious affronts.”\textsuperscript{190}

A summary of the results are provided below in Figure 3. Across all the questions, Southerners were more likely to suggest that a violent response was “extremely justified” and that Fred would not be “much of a man” if he did not respond violently.\textsuperscript{191}

\textbf{Figure 3. Percentage of Southerners and Midwesterners Approving of a Violent Response to Various Scenarios}\textsuperscript{192}

\begin{center}
\includegraphics[width=0.5\textwidth]{figure.png}
\end{center}

\textsuperscript{189} Nisbett and Cohen, \textit{Culture of Honor} at 31 (cited in note 187).

\textsuperscript{190} Id.

\textsuperscript{191} Id.

\textsuperscript{192} Id at 32.
Nisbett and Cohen reason that the regional variation they see in the responses reflects differences in local norms regarding what constitutes appropriately masculine behavior. "Fighting to answer an affront is part of the masculine ideal for southerners in a way it is not for midwesterners," but this difference "was not due simply to midwesterners' being more nonviolent generally. When questions were asked about men who fight when there has been no affront, midwesterners and southerners gave the same assessment of the men" as fitting "poorly" with their "definition of manhood."193

Dov Cohen, Richard Nisbett, Brian Bowdle, and Norbert Schwartz expanded on these studies in a series of ingenious "ethnographic experiments."194 Participants were selected for having grown up in either the North or South. All participants had to pass a (large, 6'3", 250 lbs) confederate of the researchers in a narrow hallway where there was only room for one person to walk comfortably. The new confederate walked down the center of the hall on a collision course with the participant and did not move (except at the last second to avoid bumping into the participants).195

Members of one group, after making their way past this single confederate, were exposed to a battery of tests, including tests for cortisol and testosterone levels, and were asked to self-assess their masculinity. Members of the other group were exposed to the same stimuli with one addition: prior to passing the large confederate in the hallway, they were insulted by a different confederate who bumped into each participant and called him an "asshole."196 While the differences between Northern and Southern groups who were not bumped were insignificant, the differences between regional groups who were bumped were remarkable.

As the researchers described their findings (and as displayed in Figures 4 through 7 below), compared to Northerners, insulted Southerners were "more likely to think their masculine reputation was threatened," "more upset (as shown by ... cortisol levels)," "more physiologically primed for aggression (as shown by ... testosterone levels)," "more cognitively primed for aggression," and "more likely to engage in aggressive and dominant behavior" (as indicated by their unwillingness to back down when encountering the second confederate).197

193 Nisbett and Cohen, Culture of Honor at 31 (cited in note 187).
195 Id at 948, 950, 953.
196 Id at 948.
197 Id at 945.
FIGURES 4 AND 5. CHANGES IN CORTISOL AND TESTOSTERONE LEVELS FOR INSULTED AND NON-INSULTED SOUTHERNERS AND NORTHERNERS

The researchers conclude that Southern culture supplies a social meaning to physical aggression and status that is distinct from that...
supplied by Northern culture. Southerners are more likely to view insults as diminishing a man’s status, and Southern males are more likely to attempt to restore lost status through aggressive or violent behavior.200

Clearly, our argument here is not against biology or natural selection for cognitive mechanisms that underwrite intuitions about the wrongfulness of violent acts of aggression. Individuals, on this account, do have rapid, intuitive, emotional responses fed by cognitive and biological mechanisms that have emerged over the course of human evolution; but those responses, while partially driven by physiological and biochemical responses,201 are dependent on the social meaning of the acts that precede them rather than a discrete moral module within the brain. Social norms, as these researchers describe, shape what individuals view as untoward behavior and what individuals consider appropriate responses to that behavior.

c) Disagreements over which acts meet a given standard. The examples above illustrate divergent standards governing behavior and appropriate responses to perceived wrongs, and one can easily imagine those evaluating the acts describing their disagreements in terms of explicit value differences and self-consciously norm-inflected morality. But as we mentioned above, cultural cognition will often produce subtler forms of dissensus that reflect the implicit influence of our diverse cultural commitments in the face of a single standard. Thus, even when individuals agree on a legal or moral standard to be employed, they may disagree vehemently over whether those standards have been met. That is, they may disagree about the facts as much as—or more than—they disagree about the law.

By way of illustration, we describe two examples from a series of large-scale experiments that we conducted and that are reported in greater detail elsewhere.202 In each, we asked members of the public to serve as mock jurors on a case, and in each case participants were asked to make factual findings and determine guilt.

200 Id at 956–57. There are evolutionary explanations that we do not evaluate here. See, for example, Todd K. Shackelford, An Evolutionary Psychological Perspective on Cultures of Honor, 3 Evol Psych 381, 389 (2005) (arguing that these results can be explained via evolutionary psychology).

201 Pain often generates responses that are similar to those stemming from a perceived social threat, even when the pain is internal to, and completely independent of, the social circumstances. See Raymond W. Novaco, Anger, in Alan E. Kazdin, ed., Encyclopedia of Psychology 170, 171 (Oxford 2000).

The first, modeled on the facts of the Bernard Goetz case, featured a slight white man who shot a larger black youth after the youth demanded, "Give me five dollars." The defendant had been mugged twice before and claimed that this time, based on past experience, he knew that his victim was about to seriously hurt him. He also claimed, and an expert witness avowed, that as a result of his prior muggings he suffered from posttraumatic stress syndrome. Participants were asked to read the following summary of the facts before making any factual findings or rendering a verdict:

George is charged with murdering Alvin.

George (a 48-year-old white male; 5' 7", 142 lbs.) fatally shot Alvin (a 17-year-old African American male; 6' 2", 215 lbs.) after Alvin stated "give me some money, man." The shooting occurred on a city subway platform at 5:30 p.m. on a weekday evening. After shooting Alvin, George fled but turned himself in to police three hours later.

George had been mugged on three previous occasions. On one of these, he had been beaten and required fifteen stitches under his eye. George had reported the robberies, each of which had been committed by persons George described as "teen aged, African American males," but police failed to make any arrests. George bought the handgun used in the shooting after the third mugging.

Testifying in his own defense, George told the jury that, although he'd never seen Alvin before, George "could tell from his body language and the aggressive tone of his voice" that Alvin was "going to mess with me." "It was exactly like the other time I had been attacked," George stated. "I felt I had no choice but to shoot him," George said, "because I knew if I didn't he was going to hurt me real bad." Alvin had a pocket knife on his person, but had not displayed it before being shot.

The defense also called an expert witness: Dr. Leonard Wallace, a Ph.D. psychiatrist on the faculty of a major university. Based on a [thorough] psychiatric examination of George, Wallace offered his opinion that George was suffering from "post-traumatic stress syndrome." "Like many victims of repeated violent beatings," Wallace testified, "George lived in constant fear of additional attacks." "In my opinion, George honestly perceived that Alvin would attack him if he didn't kill him first; that belief was quite reasonable,

---

given the muggings George had previously suffered, and the effect of those muggings on his psyche," Wallace concluded.

The second, based on the trial of Judy Norman, featured a wife who, after years of severe physical abuse, shot her husband in his sleep. She too claimed that, based on past experience, she sensed that her husband would seriously hurt or kill her when he awoke. She also claimed, and an expert witness avowed, that as a result of her prior abuse she suffered from battered-spouse syndrome. Participants were asked to read the following summary of the facts before making any factual findings or rendering a verdict:

Julie is charged with murdering her husband, William, whom she shot in the head as he slept.

William had persistently abused Julie during their ten-year marriage. This mistreatment included physical beatings, some of which resulted in injuries (facial cuts; broken ribs; twice a broken nose) requiring emergency medical treatment. Three times the police arrested William for assaulting Julie, but released him from custody each time after Julie declined to press charges.

Testifying in her own defense, Julie told the jury that William had beaten her on the morning of the shooting after returning home from a night of hard drinking and then fallen asleep in the bedroom. Julie testified that she then went to her mother's nearby home and obtained the hand gun used in the shooting. "I felt I had no choice except to shoot him," she stated, "because I knew when he woke up this time he was going to hurt me really bad."

The defense also called an expert witness: Dr. Leonard Wallace, a Ph.D. psychiatrist on the faculty of a major university. Based on a thorough psychiatric examination of Julie, Wallace offered his opinion that Julie was suffering from "battered woman syndrome." "Like other victims of chronic domestic violence," Wallace testified, "Julie believed that she was powerless to leave and that no one could or would help her." "In my opinion, Julie honestly perceived that her husband would attack her if she didn't kill him first; that belief was quite reasonable, given the beatings

---

204 See Kahan and Braman, 45 Am Crim L Rev at 26, 65 (cited in note 34).
205 See State v Norman, 378 SE2d 8, 13 (NC 1989) (affirming the conviction of Judy Norman for voluntary manslaughter because there was no evidence that she "reasonably believed that she was confronted by a threat of imminent death or great bodily harm").
she had previously suffered, and the effect of those beatings on her psyche," he concluded. 206

The stimuli to which participants were exposed, it should be noted, are distinct in several ways from the stimuli in ranking studies described in Parts I and II. The factual summaries read by respondents in our studies were far more detailed. Participants who read these more detailed scenarios were also provided with jury instructions summarizing the doctrinal standard and specifying the relevant facts they needed to find in order to convict or acquit. Participants were then asked to answer a series of questions regarding legally relevant facts and, once they made those findings, to render a verdict.

Thus, whereas participants in ranking studies are asked whether they think an act described in highly simplified terms is wrong, participants in our studies were given highly detailed fact patterns and a specific standard under which to evaluate the wrongfulness of the act in question. Given the naturalist assertion that the "potential for exaggerating the extent of disagreement becomes greater as the crime descriptions become more skeletal, and is at its worst when researchers use crime labels rather than factual descriptions," 207 we would expect, on the naturalist account, to find far less disagreement here than in the ranking studies.

How did the participants react to these stimuli? To begin with, there was significant variation across several dimensions. Blacks were more likely to convict George than they were to convict Julie, while whites were more likely to convict Julie than George. Similar patterns emerged for women and men, Democrats and Republicans, liberals and conservatives, and egalitarians and hierarchs, communitarians and individualists. In each case, the former were more likely than the latter to see George as more deserving of punishment than Julie. The results are provided in Table 7 below.

206 Kahan and Braman, 45 Am Crim L Rev at 26, 79 (cited in note 34).
207 Robinson and Kurzban, 91 Minn L Rev at 1860 (cited in note 13).
TABLE 7. FREQUENCY OF GUILTY VERDICTS ACROSS RACE, GENDER, PARTY, IDEOLOGY, AND CULTURAL ORIENTATION

<table>
<thead>
<tr>
<th></th>
<th>George</th>
<th>Julie</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>772</td>
<td>838</td>
</tr>
<tr>
<td>black : white</td>
<td>56% : 29%</td>
<td>41% : 48%</td>
</tr>
<tr>
<td>male : female</td>
<td>34% : 32%</td>
<td>50% : 44%</td>
</tr>
<tr>
<td>Republican : Democrat</td>
<td>24% : 39%</td>
<td>51% : 43%</td>
</tr>
<tr>
<td>liberal : conservative</td>
<td>43% : 23%</td>
<td>42% : 56%</td>
</tr>
<tr>
<td>egalitarian : hierarch</td>
<td>44% : 23%</td>
<td>42% : 51%</td>
</tr>
<tr>
<td>individualist : communitarian</td>
<td>26% : 40%</td>
<td>51% : 43%</td>
</tr>
</tbody>
</table>

These cross-tabulations begin to suggest what the differences across the population are like. Every demographic group listed above showed significant differences (at $p \leq 0.10$) in determinations of guilt with one exception: men and women did not significantly differ over George’s case.208 These findings, of course, stand in stark contrast to the ranking studies described above,209 which found no differences in relative seriousness.

But this kind of simple comparison is far from an ideal evaluation of differences of opinion across the population. People are not generically black or white, male or female, Republican or Democrat, liberal or conservative, egalitarian or individualist; these characteristics and values tend to come in packages. How would more fleshed-out types of people react to each of the cases?

Imagine two Americans.210 Ron, a white male who lives in Arizona, overcame his modest upbringing to become a self-made millionaire businessperson. He deeply resents government interference with markets but is otherwise highly respectful of authority, which he believes should be clearly delineated in all spheres of life. Politically, he identifies himself as a conservative Republican. Linda is an African-American woman employed as a social worker in Philadelphia,

---

208 Kahan and Braman, 45 Am Crim L Rev at 34 table 1 (cited in note 34).
209 See Part I.A.
210 Yes, these are the same folks made famous in a recent and brilliant article assessing the Supreme Court’s decision in Scott v Harris, 550 US 372, 386 (2007) (finding that a law enforcement officer acted reasonably in terminating a car chase by taking an action that caused substantial injuries to the driver). See Dan M. Kahan, David A. Hoffman, and Donald Braman, Whose Eyes Are You Going to Believe? Scott v. Harris and the Perils of Cognitive Illiberalism, 122 Harv L Rev 837, 895–99 (2009) (arguing that the Supreme Court’s reasoning failed to connect perceptions of societal risk and contested visions of the ideal society, and invested the law with culturally partisan overtones that detract from the law’s legitimacy).
Pennsylvania. She is a staunch Democrat and unembarrassed to be characterized as a "liberal."

Zelig, a statistical application designed by Kosuke Imai, Gary King, and Olivia Lau, facilitates simulating such complex profiles, furnishing a perfect fit for comparing responses across more detailed types of people. It allows for reasonable statistical predictions of the perceptions of fairly specific types of people by setting pertinent characteristics—cultural values, gender, race, region of residence, political ideology, and party affiliation—to appropriate values in Zelig simulations.

Individually and collectively, these analyses present a test for the naturalist and realist perspectives. If, as naturalists assert, humans evaluate cases involving human aggression with high degrees of consistency, then we would expect similar assessments of the wrongfulness of each act across the population. Recall that naturalists hold that while individuals may disagree about how much to punish bad acts, they agree on what constitutes a bad act. As such, on the whole, the population should agree that, in each case, the defendant is either guilty or innocent. On the other hand, if the realist position is right, then the social meaning of the acts will move them to evaluate the cases differently, either increasing or decreasing the likelihood of conviction or acquittal.

Relatedly, on the naturalist account, the cases should be viewed as consistently more or less bad relative to each other. That is, the perceived wrongfulness of acts might not be absolute across the population, but it should be consistently ranked across the population. On the realist account, by contrast, the contingency of the social meaning of the acts should cause egalitarians and communitarians like Linda to view George's shooting of strangers (with racial overtones) as worse than the act of Julie, the battered woman shooting her husband; and it should cause those who favor individualism and traditional social hierarchies

---

211 See Kosuke Imai, Gary King and Olivia Lau, Toward a Common Framework for Statistical Analysis and Development, 17 J Computational & Graph Stats 892, 894 (2008). In conventional regression analysis, the influence of some set of explanatory variables on a dependent variable is expressed in a mathematical equation, the elements of which (regression coefficients, standard errors, p-values, and so forth) are reported in a table. Zelig is intended to generate data analyses that simultaneously extract more information and present it more intelligibly. Using Zelig, an analyst specifies values for the independent variables that form a regression model. The application then generates a predicted value for the dependent variable through a statistical simulation that takes account of the model's key parameters (including the standard errors for the regression coefficients). It then repeats that process. Then it repeats it again. Then it repeats it again and again and again—as many times as directed by the analyst (typically ten thousand times, or enough to give a reasonable approximation of the probability distribution for the dependent variable). The resulting array of values for that dependent variable can then be analyzed with techniques that are statistically equivalent to those used in survey sampling to determine an average predicted value, plus a precisely calculated margin of error. See id at 895–96.
to view Julie’s shooting of her husband as worse than George’s shooting, which they would view as a legitimate act of self-defense.

So how would these two distinct members of the American venire evaluate these cases? As indicated in Figures 8 and 9 below, Zelig reveals the kind of demographic and values-based variation that naturalism theorists suggest does not exist.

**FIGURES 8 AND 9. RON AND LINDA’S WILLINGNESS TO CONVICT OR ACQUIT**

![Graph](image)

Notice that it is not the case, as naturalists argue, that disagreement is generally about the “endpoints” of punishment. What we see here are *different rates of conviction and acquittal*. Moreover, and even more strikingly, we see that people with a cultural profile like Ron are inclined to acquit George but convict Julie, whereas those with cultural profiles like Linda are inclined to do just the reverse.

---

212 With 95 percent confidence intervals.
Because it assumes a lack of diversity in the core of wrongdoing, naturalism cannot account for the variation we see in the data. If individuals have an intuitive sense of the relative wrongfulness of acts, then we would expect people with cultural profiles like Ron and Linda to agree—perhaps not on precisely how much punishment a person deserves, but at the very least on the relative culpability of the two defendants. For naturalism, dissensus in the core of wrongdoing remains a puzzle.

IV. PRACTICAL IMPLICATIONS

What should those who care about these issues take away from all of this? Does either theory provide useful guidance with respect to the practical questions that those involved in live debates over criminal law face?

A. (Anti-)Punishment Naturalism?

To begin with, it is hard to see how legal actors can draw any normative conclusion from the naturalist literature. We have pointed out how evidence amassed by social scientists in various disciplines furnishes ample reason to doubt that universally shared "core intuitions of justice" dispel dissensus, across space and time, about the sorts

---

213 With 95 percent confidence intervals.
of conduct that should be forbidden and punished. But however one characterizes the extent or importance of innate apprehensions of wrongfulness, the issue of the moral significance of such sensibilities is an entirely different matter. Since at least Hume, it has been well known that facts—about anything—do not entail moral "oughts." Accordingly, showing that intuitions of justice are shared does not mean they have to be respected.

The Punishment Naturalists, consistent with their characteristic care and thoughtfulness, of course never suggest otherwise. Their injunction that lawmakers and advocates of reform respect laws consistent with "shared intuitions" reflects a judgment about the enormous effort that would be required to talk humans out of predispositions that reflect "600 million years" of biological programming. "Evidence suggests that it takes a dramatic, concerted effort to alter fundamentally a person's intuitive notions of justice. Such changes in core judgments have been notably observed in cases of coerced indoctrination, often referred to as 'brainwashing.'" This understanding of the intractability of core judgments of punishment transforms into a conservative admonition to be wary of even trying, but only after Punishment Naturalists take stock of the potentially disastrous consequences of failed attempts to do so. "The criminal law can most effectively maximize its moral credibility and thereby minimize resistance and subversion by adopting criminal rules that track shared community intuitions of justice," Paul Robinson and John Darley observe. 217 "The danger of failing to harmonize criminal codes with intuitions of justice is that the code may lose credibility on a wide array of prohibitions if too many are perceived to be against notions of what is just." 218 Surely, no one would be in favor of any reform program that can depend for its success only on the sorts of mental reprogramming strategies used by the "Chinese military on American soldiers captured during the Korean War," and that would likely, in any case, culminate in "a generalized contempt for the system in all its aspects, and a generalized suspicion of all of its rules" and ultimately in the emergence of "active forces of subversion and resistance" within the general population. 220

216 Robinson and Darley, 81 S Cal L Rev at 51 (cited in note 15).
217 Id at 28.
218 Id.
219 Id at 54.
220 Robinson and Darley, 81 S Cal L Rev at 24 (cited in note 15).
But are these admittedly terrifying prospects really the likely outcome of efforts to use law to try to change norms on, say, “date rape,” “drunk driving,” “drug use,” or “same-sex intercourse and . . . same-sex unions”221—examples that the Punishment Naturalists cite as involving “criminal law manipulation” by “social engineers”?222 Where is the empirical evidence of that? Indeed, where is the empirical evidence that even the much more wide-ranging reform effort that Punishment Naturalists oppose—one evincing uniform hostility to popular retributivist sensibilities generally and their replacement nonjudgmental utilitarian schemes of treatment and control—would result in widespread social tumult?

Perhaps the most obvious clue that the conservative posture associated with Punishment Naturalism does not follow in any straightforward or obvious way from the evidence of the origins of punitive sensibilities on which it is based can be found in the work of another group of highly accomplished scholars who draw exactly the opposite conclusions from that same evidence. We call these scholars, who include Joshua Greene and Jonathan Cohen, the Antipunishment Naturalists.

In work that Punishment Naturalists actually cite for support,223 Greene and Cohen (in work to which Darley also contributed) present evidence of the biological origins of widespread moral sensibilities. The evidence includes fMRI studies that show that “deontological” moral judgments, of which retributive intuitions are a conspicuous component, originate in parts of the brain that are associated with fast-acting, unconscious, and automatic affective processes. Accordingly, they are much more likely to influence action than are “consequentialist” or utilitarian moral judgments, which these same studies show originate in a more slow-acting, reflective part of the brain, whose thought processes can override those of the faster-acting, reactive part only with the exertion of considerable, time-consuming effort.224

Like the Punishment Naturalists, Greene and Cohen identify an evolutionary or genetic origin for retributive and like judgments, which

---

221 Id at 52–53.
222 Id at 52.
they claim were well adapted to social conditions distinctive of the sorts of ties and transactions that were characteristic of our remote tribal past. They agree too that it is difficult, if not impossible, to “educate” or talk people out of the punitive positions that they are impelled to by their genetic-neurological hardwiring, in part because integral to the same circuitry is a disposition to “confabulate”—that is, to seize on post hoc rationalizations that occur to us after our unconscious affective sensibilities have committed us to a moral position and that stubbornly resist the battering of conscious, reasoned examination.

But from this foundation—as close as it is to that of the Punishment Naturalists in various critical particulars—the Antipunishment Naturalists derive a very different set of normative conclusions. For them, the evolutionary origins of our widespread punitive intuitions, far from enhancing the moral authority of retributive and like sensibilities, strips them of any pretense of being moral at all: “as an evolutionary matter of fact, we have a taste for retribution, not because wrongdoers truly deserve to be punished regardless of the costs and benefits, but because retributive dispositions are”—or at least were at one point—“an efficient way of inducing behavior that allows individuals living in social groups to more effectively spread their genes.” For Greene and Cohen, the truly moral judgments are the ones that can be defended on the basis of (nonconfabulatory) reflection on what conduces to the best state of affairs in our current situation.

Of course, if, as the Punishment Naturalists warn, it were futile or even self-defeating to oppose retributive sensibilities (however outmoded and insusceptible of reasoned defense they are), it would also be foolish to try to supplant them with intuitions that reflect a consequentialist orientation. But the Antipunishment Naturalists have a different account of how such a reform program would fare. As they see it, the discoveries of sociobiology and neuroscience on which they and the Punishment Naturalists both rely will themselves transform our culture:

The net effect of this influx of scientific information will be a rejection of free will as it is ordinarily conceived, with important ramifications for the law. As noted above, our criminal justice system is largely retributivist. . . . [R]etributivism . . . ultimately

---


227 Id at 71.
depends on an intuitive, libertarian notion of free will that is undermined by science. Therefore, with the rejection of commonsense conceptions of free will comes the rejection of retributivism and an ensuing shift towards a consequentialist approach to punishment, i.e. one aimed at promoting future welfare rather than meting out just deserts.28

The Antipunishment Naturalists offer no empirical evidence (as opposed to conjectural storytelling) to back up this account of how a program to abolish a retributive system of criminal law will be received. But that just means that they present no less empirical evidence than do the Punishment Naturalists in support of their grim account of the consequences that meaningful pursuit of that vision would entail.

If one conclusion can confidently be drawn from this disagreement, then, it is that none of the materials on which both the Punishment and the Antipunishment Naturalists rely has any obvious moral upshot. The privilege of thinking about what to do, and the empirical work necessary to determine whether and how it can be done, survive naturalism of any variety.

B. Relevant to What?

We take it, though, that Punishment Naturalists believe that the evidence they furnish is of significant practical import to debates over the criminal law. On their account, the most serious problem posed by our intuitions about wrongdoing is the attempted imposition of social norms that are at odds with human nature. As Robinson and Darley recently argued:

[T]hese findings regarding the nature of intuitions of justice have serious implications for a variety of criminal justice debates that focus on substantial alterations of criminal justice systems, including the abolition of punishment, the distribution of punishment according to principles that conflict with shared intuitions of justice, and programs to change people’s intuitions about what constitutes serious wrongdoing and about how much it should be punished.29

Marc Hauser is similarly worried about “policy wonks and politicians” who attempt to develop laws that are out of step with our natural intuitions.30 This might be generalized to something along the following

29 Robinson and Darley, 81 S Cal L Rev at 11 (cited in note 15).
30 Hauser, Moral Minds at xx (cited in note 26).
lines: the state should pay close attention to widely shared intuitions about justice; it may like them or not, but it ignores them at its peril.

Do realists object to that? No. But to the realist it is puzzling advice, for it does not address any live policy debate. What political candidate is running on a platform of "Abolish Punishment Now!"? What legislator is attempting to implement a program punishing crimes that most people think are very serious less harshly than those that most people think are not very serious? Even if we were to accept the naturalist account of moral intuition, the added value of assuming that our intuitions about punishment are natural rather than social seems negligible; it just does not address the practical problems that we face related to punishment.

But Punishment Naturalists do offer advice to people involved in contemporary debates over the criminal law. And, from criminal reform efforts on everything from date rape to drug use, the advice that they have to offer is pretty discouraging. On their account, issues that fall within the core of wrongdoing—and recall that, on their account, this comprises the vast majority of criminal acts—there will be little chance of making a lasting impact. As Robinson and Darley put it: "Because of the universal and intuitional nature of core judgments about justice . . . these judgments cannot easily be changed." And, for those few wrongful acts that remain outside of the core, they offer advice for those interested in reform. The activity in question has to be plausibly viewed as intentionally inflicting harm on others in ways that can be viewed as similar to some wrong within the "core" of wrongdoing, and the most effective mode of argument is to analogize to that core wrong.

What could possibly be wrong with this advice? In fact, picking reasonable targets for reform and then hammering home the message that the targeted activity is similar to other stigmatized and punished acts certainly seems like common sense.

But is it? The advice on offer strikes us as simply inapposite. Individuals disagree about whether a car salesman who successfully convinces his target to buy a lemon has committed fraud—at the core of deception in exchange—or has simply displayed admirable American

231 Robinson and Darley, 81 S Cal L Rev at 53 (cited in note 15). To drive home the point of how difficult any reform efforts (even those outside of the core of wrongdoing) will be, they describe the failures of various indoctrination campaigns, including the immense effort required for, and relatively short-term effects of, the brainwashing of POWs during the Korean war, the impossibly strained conditions that produce Stockholm Syndrome, and the failure of Prohibition to reform intuitions about alcohol consumption—and these involved efforts aimed at offenses outside of the "core!"

232 See Kerr, The Intuition of Retribution (cited in note 23).
salesmanship. If fraud strikes you as not quite "core" enough, imagine, for example, that you were concerned with reducing acquaintance rape, including the imposition of unwanted sex on women whose verbal resistance is ignored. The idea that such behavior is "analogous" to—or just is—rape has been exactly what reformers have been arguing, and their opponents resisting, for decades. The argument is not about what is analogous to the core; it is about what the core is. No position that abstracts away from the cultural dispute over the definition of rape can possibly generate advice to reformers about what they should do.

Or imagine that you were concerned with reducing violence against women. You see that women who are victims of homicide are often killed by husbands or boyfriends who discover or suspect infidelity. The law, you come to believe, encourages this behavior by allowing those who kill an unfaithful partner to be convicted not of murder, but of the lesser crime of manslaughter. Infidelity, the law tells the public, is "adequate provocation" for such mitigation. Judges, moreover, say that the law reflects a perfectly natural sentiment. The advice to "argue from analogy" to agreed-upon "core" offenses is unhelpful because it simply ignores that the core is itself a site of intense cultural dispute: there is a serious dispute about whether the act is—or is not—murder.

Similarly, people of varying cultural outlooks disagree about whether a woman who kills her chronically abusive husband has committed a core crime—murder—or no crime at all, not about how to deal with some peripheral offense "analogous to" murder. And the list goes on.

C. Some Realist Advice

Realists, like naturalists, are circumspect about the prospects for resolving many disputes over the law, though for different reasons and

---

233 Analogously, individuals disagree about whether and when omitting information in an exchange is a form of lying or no wrong at all. See Richard Craswell, Taking Information Seriously: Misrepresentation and Nondisclosure in Contract Law and Elsewhere, 92 Va L Rev 565, 574–75 (2006).


235 One judge, Robert E. Cahill, famously quoted by Cynthia Lee in her book, Murder and the Reasonable Man, lamented his duty to impose any sentence at all on a man who shot and killed his wife after discovering her infidelity, saying "I seriously wonder how many men [on discovering spousal infidelity] would have the strength to walk away without inflicting some corporal punishment." Cynthia Lee, Murder and the Reasonable Man 41 (NYU 2003).

236 See Kahan, 158 U Pa L Rev at 805 (cited in note 158) (discussing the problem of disagreement over the core offense in the acquaintance rape context).
with different consequences. Realist circumspection derives not from
the view that reform within the core pits impotent cultural forces
against powerful innate intuitions, but rather from the view that
reform of most of the salient legal standards pits powerful cultural
forces against one another. The realist reasons that many debates over
the law are so fraught because they are about whose values the law
will privilege. This can lead to fierce social conflict with the status of
the law marking which social group has prevailed.237

But even if the structure of some cultural conflicts requires the law
to choose a winner, in at least some other instances, diverse citizens are
willing to focus on shared concerns. Talk about deterrence, utility, and
social welfare often signals that parties are attempting to resolve their
disputes without resorting to culturally sectarian forms of argument.238
This is, in essence, the basis of liberal democratic deliberation.

But even those committed to a liberal ideal of deliberation over
the law can polarize on issues along cultural lines. A growing literature
suggests that this can be explained by the phenomenon of cultural
cognition, which causes individuals to conform their factual beliefs to
their cultural priors, preventing them from reaching agreement de-
spite their commitment to social welfare maximization or some other
nonsectarian ground for deliberation.239

Here, we think, a little realism may be of assistance. Where par-
ties have agreed to resolve a dispute on nonsectarian terms, but are
hampered by cultural biases that cause them to come to conflicting
conceptions of the facts, it may be possible to help parties attend to
factual data in a less biased manner. While research in this area is on-
going, understanding the nature of the conflict as cultural is crucial to
developing effective strategies for mitigating polarization of factual
beliefs along cultural lines.

D. Doing What Comes Naturally

Ironically, although Punishment Naturalism is focused on shared
intuitions about wrongfulness, taking it to heart seems likely to esca-
late social conflict over the criminal law. To understand why, though,
one has to understand how Punishment Naturalism leverages
well-established psychological phenomena involved in evaluating wrongful
acts and actors.

237 See generally Dan M. Kahan and Donald Braman, Cultural Cognition and Public Policy,
238 See generally Kahan, 113 Harv L Rev 413 (cited in note 11).
The first phenomenon is naïve realism. Naïve realism suggests that people are quite good at spotting bias in others, but not very good at spotting it in themselves. (The realism part of naïve realism refers to the ability of individuals to perceive biasing influences on other people; the naïve part of naïve realism refers to the belief that such biases do not obtain in the self.) As a result of this widely studied mechanism, we are likely to view other people as having biased and abnormal conceptions of the world—at least relative to ourselves. As a result, when it comes to moral disputes, we are likely to view others (rather than ourselves) as suffering from some form of moral bias.

The self-serving nature of naïve realism echoes another phenomenon that is one of the most famous in all of social psychology: fundamental attribution error. When attempting to attribute our own acts to either some fundamental attribute or situational influence, we tend to view our socially desirable acts as stemming from fundamental attributes and undesirable acts as stemming from situational influences. When evaluating the acts of others, though; the reverse is true: actors tend to attribute the undesirable acts of others to fundamental attributes and desirable acts to situational influences. Thus, if we think of our own moral acts and expressions as desirable, we will tend to think of them as reflecting a fundamental moral character rather than some contingent or situational valuation; and the same will be true for our attributions of the undesirable acts—acts we perceive as immoral—of others. In moral disputes, then, we tend to view our different behaviors as reflecting relatively fundamental attributes in both ourselves and those with whom we disagree.

What does this have to do with Punishment Naturalism? Think, for a moment, about the way naturalist explanations orient individuals with respect to their disagreements with one another. Recall that naturalism posits that normally developed humans will share naturally occurring intuitions about the vast majority of wrongful acts. And recall that because of the phenomenon of naïve realism and fundamental attribution bias, individuals are more likely to attribute such biases to others than to themselves and to think of them as reflecting fundamental differences. Then ask yourself this: does thinking that someone


who disagrees with you is innately abnormal in a fundamental moral capacity increase or decrease the likelihood of your understanding their concerns and working toward viable reform? As it happens, what research has been done in this area suggests that it is quite easy to think of outgroup members as fundamentally different from oneself, and that this decreases the likelihood of cooperation.  

When evaluating one another’s claims, realists caution themselves against viewing those with whom they disagree as having an innately or fundamentally abnormal moral instinct. This does not by any means guarantee that they will be able to come to an agreement or effect reform—explicit value differences have underwritten and continue to generate serious social conflict. But it does, we hope, help guard against the all-too-easy jump to thinking that those who disagree with our moral intuitions do so because they suffer from some fundamental moral abnormality. Indeed, is this not already a problem in conflict around the world: that we come to think of the people with whom we disagree as intrinsically less moral rather than contingently different along cultural lines?

Fundamental attribution error and naïve realism can also help us understand why it is that Punishment Naturalism, despite all of the contrary evidence, feels so natural to so many people. These (perhaps even universal!) cognitive mechanisms cause us to favor arguments like Punishment Naturalism. And it is these cognitive mechanisms that Punishment Naturalists reinforce when they argue that we should “listen more closely to the moral voice of our species” and avoid the call of “policy wonks” who tell us that we should adjust our intuitions to fit their reasoned arguments for improving social welfare. Our moral intuitions, our cognitive biases persuade us, both are unbiased and reflect a fundamentally positive aspect of our nature.

---

243 What research does exist in this area suggests that those who view human nature as largely set and inflexible are also more likely to look upon outgroup members as less likely to be cooperative, and are thus less likely to actually cooperate with them; the inverse is true for those who view human nature to be more malleable. See generally, for example, Sheri R. Levy, Chi-yue Chiu and Ying-ji Hong, Lay Theories and Intergroup Relations, 9 Group Processes Intergroup Rel 5 (2006) (describing how various lay theories about human attributes and cognition affect intergroup relations); Nick Haslam, et al, Psychological Essentialism, Implicit Theories, and Intergroup Relations, 9 Group Processes Intergroup Rel 63 (2006) (finding that the belief that human attributes are malleable increases intergroup cooperation); Giulio Boccato, et al, The Automaticity of Infra-humanization, 37 Eur J Soc Psych 987 (2007) (finding support for the infra-humanization hypothesis that uniquely human emotions are automatically more linked in memory with the ingroup than with the outgroup).

244 See Nick Haslam, Dehumanization: An Integrative Review, 10 Personality & Soc Psych Rev 252, 252 (2006) (noting the tendency to treat outgroup members “as animal-like” and to represent them as “objects or automata”).
We think Punishment Realism gets it right—that is, it furnishes a more accurate depiction of human moral intuitions than Punishment Naturalism—and for most people, getting it right will be enough. But we also think it can help solve—or at least not exacerbate—collective action problems. Realism points citizens in a productive direction, focusing attention on issues around which there is genuine dispute and lack of coordination. It also highlights the social and cognitive mechanisms that generate that unnecessary conflict, allowing legal actors to develop tools with which to understand the source of dissensus. That certainly does not guarantee a quick or easy resolution, but it does, we think, provide a reasonable start toward solving the difficult problems involved in such disputes.

CONCLUSION

We hope that the reader has come to this point with an appreciation for the way widely shared—perhaps even universal—cognitive mechanisms can give rise to diverse and often conflicting intuitions about justice. Many of these mechanisms may be the products of natural selection, but their flexibility lends our intuitions tremendous range and scope. Knowledge about our cognitive building blocks, on this realist account, can help us understand why our intuitions about what is just seem so natural, even when they are so clearly subject to cultural variation.

We have the utmost respect for Punishment Naturalism, which we recognize as embodying a rich and growing stock of insights informed by highly rigorous and sophisticated methods. There is a growing trend toward the integration of empirical insights from a variety of disciplines into legal scholarship. The originality of argument and the scope of the research that characterize Punishment Naturalism are a testament to the value of this trend.

We do feel deep concern, however, over what we take to be the politically conservative resonances with which the Punishment Naturalist has been needlessly infused. It is, simply put, extremely difficult to take in the corpus of work that the Punishment Naturalists have amassed without sensing a deep commitment on their part to the status quo—to popular retributive sensibilities as they are (or are depicted with a high degree of uniformity to be), and to laws that conform (or are depicted as conforming) to them. Popular understandings of wrongfulness, we are repeatedly told, are “deep, predictable, and widely shared.”245 They are the product of inexorable biological

245 Robinson and Kurzban, 91 Minn L Rev at 1892 (cited in note 13).
forces—"an evolved predisposition" combined with "social learning" of the sort that "arise[s] only from . . . human life experience[s] so fundamental as to be essentially universal to all persons."246 As such, it is naïve to expect "arguments or education" to change them;247 something much more fundamental, and much more odious, would be necessary "to fundamentally alter . . . intuitive notions of justice," something akin to the "coercive indoctrination" that is characteristic of totalitarian states and that would never "be tolerated in a modern liberal democracy."248 Less extreme "social engineering programs aimed at changing"249 norms through laws are not only likely to fail, but also to blow up in the engineers', and everyone else's, faces: "when the criminal justice system is seen as out of tune with community sentiments," the law suffers a "loss of moral credibility" that can grow into "a generalized contempt for the system in all its aspects, [ ] a generalized suspicion of all of its rules," and ultimately the destruction of its "relevance as a guide to good conduct."250

It is possible that the Punishment Naturalists mean to direct their cautionary, "hands off" admonition only to academic theorists who call for replacing "punishment" informed by retributive sensibilities of any sort with a humanistic—or perhaps simply technocratic—utilitarian regime animated by goals of incapacitation and therapy.251 As we have pointed out, the most intriguing theoretical architects of such a system build their regime on the same psychobiological foundation on which the Punishment Naturalists rest their own populist retributivism.252 But the generality with which the Punishment Naturalists couch the lessons they draw from their work, and the nature of the concrete examples they give of "recent reform programs" that embody "criminal law manipulation . . . by social engineers"253—programs aimed at reforming rape law, at reducing smoking, at increasing punishment of domestic violence, at discouraging recreational drug use, at focusing attention on drunk driving, at combating workplace sexual harassment, and at "build[ing] public acceptance of both same-sex

246 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1646, 1687 (cited in note 14).
247 Robinson and Darley, 81 S Cal L Rev at 52 (cited in note 15).
248 Id at 54–55.
249 Id at 51.
250 Id at 24.
251 See, for example, Robinson and Kurzban, 91 Minn L Rev at 1892 (cited in note 13) ("[I]t may be unrealistic to expect the population to all 'rise above' its desire to punish wrongdoers, or to expect the government to 'reeducate' people away from their interest in punishing wrongdoers, as is urged by some reformers."). See also Robinson, Kurzban, and Jones, 60 Vand L Rev at 1688 (cited in note 14).
252 See Part IV.A.
253 Robinson and Darley, 81 S Cal L Rev at 52–53 (cited in note 15).
intercourse . . . and same-sex unions—invite readers (who have indeed accepted the invitation) to see their resistance to using law to change norms as having much broader normative significance.

Our goal has been to show that Punishment Naturalism does not supply a basis for any particular position on any live and disputed issue in the criminal law informed by clashing cultural values. We conclude with a statement of three interrelated propositions that we think survive close engagement with Punishment Naturalism and that defeat any attempt to derive a generic, conservative suspicion to norm reform from it.

First, as a matter of politically consequential fact, intuitions of justice are characterized by immense cultural heterogeneity. It might be the case that, for the most part, human beings everywhere and at all times have been opposed to murder, rape, and misappropriation. But over space and over time, what counts as murder, rape, and misappropriation have varied tremendously. The reason is that cultural norms define key elements of those wrongs—who counts as a person, for example, what sorts of behavior interfere with a person’s vital interests, what kinds of behavior surrender rightful control of one’s body, what sorts of personal and communal claims constrain individual entitlements to property, and the like. Opposing understandings like these persist across identifiable cultural groups in contemporary American society. They are what animate debates about rape reform, gradations of homicide, abortion, the scope of antifraud provisions, and myriad other issues.

Second, intuitions of justice are plastic. By this we mean that such understandings do in fact change in one place over time. Often change is slow and gradual; but sometimes it is quite sudden and dramatic.

Third, intuitions of justice and law are endogenous. This is the simple point that understandings of wrongdoing and law are reciprocally related: what is considered “wrongful” influences law, and what the law prohibits influences understandings of what is wrongful, and also how wrongful it is. Accordingly, law reform often can be a catalyst for norm change—indeed, for norm change that itself feeds back on law and thus back on itself. Examples of such interplay are legion, including the prominent example of homicide law in the United States, where the changing factual circumstances that the law (formally through doctrine, and practically through jury verdicts) recognizes as

254 Id.
255 See Kerr, The Intuition of Retribution (cited in note 23).
full or partial defenses show how highly interactive culture, law, and
intuitions relating to core wrongdoing can be.256

We want to close with an admission and some appreciation. It is
quite possible that we are wrong. The beauty of the best naturalist
work—and here we are thinking of the work of those we have criticized
above—is that it makes clear claims based on readily discernible data.
We think that the naturalists have missed data and that their claims are
too broad, but they have moved the ball forward significantly by articu-
lating claims that previously had been made without data and which
were thus nearly impossible to engage on the merits. And perhaps we
are too critical. Perhaps we are succumbing to cognitive biases that
cause us to favor evidence supporting our own parochial perspective.
Perhaps. Happily, if you think that, you are already a realist.

256 See Kahan and Nussbaum, 96 Colum L Rev at 346–50 (cited in note 7).
APPENDIX

Items from the ranking studies conducted by Robinson and Kurzban.257

The following four scenarios were typically assigned no punishment and were thus unranked.

DEFENDING ATTACK. John is knocked down from behind by a man with a knife who moves to stab him. As the man lunges for him, John stabs him with a piece of glass he finds on the ground, which is the only thing he can do to save himself from being killed. The man later dies of his injuries.

COERCIVE THREAT TO CHILD. A man grabs John’s child and puts a sharp knife to her throat. He tells John that he will kill the child if John does not steal an expensive digital camera from a nearby shop or he attempts to contact police. Because the man can see everything he does, John does as he is told in order to save his child.

UMBRELLA MISTAKE. John takes another person’s umbrella assuming it to be his own because it is has the same unusual color pattern as his own, a fact that the police confirm.

HALLUCINATION. Another person slips a drug into John’s food, which causes him to hallucinate that he is being attacked by a wolf. When John strikes out in defense, he does not realize that he is in fact striking a person, a fact confirmed by all of the psychiatrists appointed by the state, who confirm that John had no ability to prevent the hallucination.

The following twenty-four were typically ranked as increasingly serious offenses and as deserving increasing quanta of punishment.

WHOLE PIES FROM BUFFET. The owner has posted rules at his all-you-can-eat buffet that expressly prohibit taking food away; patrons can only take what they eat at the buffet. The owner has set the price of the buffet accordingly. John purchases dinner at the buffet, but when he leaves he takes with him two whole pies to give to a friend.

LOGO T-SHIRT FROM STORE. John notices in a small family-owned music store a T-shirt with the logo of his favorite band. While the store clerk is preoccupied with inventory, John places the $15 T-shirt in his coat and walks out, with no intention of paying for it.

SHORT CHANGE CHEAT. John is a cab driver who picks up a high school student. Because the customer seems confused about the money transaction, John decides he can trick her and gives her $20 less change than he knows she is owed.

CLOCK RADIO FROM CAR. As he is walking to a party in a friend’s neighborhood, John sees a clock radio on the back seat of a car parked on the street. Later that night, on his return from the party, he checks the car and finds it unlocked, so he takes the clock radio from the back seat.

ELECTRIC DRILL FROM GARAGE. John does not have all the tools he needs for his workshop but knows of a family two streets over who sometimes leave unlocked the door to the detached garage next to their house. When he next sees his chance, he enters the detached garage through the unlocked door and takes a medium-sized electric drill, intending to keep it forever.

MICROWAVE FROM HOUSE. While a family is on vacation, John jim mies the back door to their house and steps into their kitchen. On the counter, he sees their microwave, which he carries away.

SMASHING TV. While a family is away for the day, John breaks in through a bedroom widow and rummages through the house looking for valuables. He can only find an eighteen-inch television, which angers him. When he gets it outside, he realizes that it is an older model than he wants, so he smashes it onto the driveway, breaking it into pieces.

SLAP & BRUISING AT RECORD STORE. A record store patron is wearing a cap that mocks John’s favorite band. John follows him from the store, confronts him, then slaps him in the face hard, causing him to stumble. The man’s face develops a harsh black and yellow bruise that does not go away for some time.

HEAD-BUTT AT STADIUM. While attending a football game, John becomes angry as he overhears an opposing fan’s disparaging remarks about John’s team. At the end of the game, John sticks his face in the man’s face and head-butts him, causing a black eye and a gash that requires two stitches to close.

STITCHES AFTER SOCCER GAME. Angry after overhearing another parent’s remarks during a soccer match in which John’s son is playing, John approaches the man after the game, grabs his coffee mug, knocks him down, then kicks him several times while he is on the ground, knocking him out for several minutes and causing cuts that require five stitches.
NECKLACE SNATCH AT MALL. As a woman searches her purse for car keys in a mall parking lot, John runs up and grabs her gold necklace but it does not break. He yanks the woman to the ground by her necklace, where she gashes her head, requiring stitches. John runs off without the necklace.

ATTEMPTED ROBBERY AT GAS STATION. John demands money from a man buying gas at a gas station. When the man refuses, John punches the man several times in the face, breaking his jaw and causing several cuts that each require stitches. He then runs off without getting any money.

CLUBBING DURING ROBBERY. To force a man to give up his wallet during a robbery attempt, John beats the man with a club until he relinquishes his wallet, which contains $350. The man must be hospitalized for two days.

MAULING BY PIT BULLS. Two vicious pit bulls that John keeps for illegal dog fighting have just learned to escape and have attacked a person who came to John’s house. The police tell John he must destroy the dogs, which he agrees to do but does not intend to do. The next day, the dogs escape again and maul to death a man delivering a package.

INFANT DEATH IN CAR. John is driving to see a man about buying an illegal gun but must baby-sit his friend’s toddler son. It occurs to him that it is too hot to safely leave the toddler in the car but he decides to leave him anyway and to return soon. He gets talking with the seller, however, and forgets about the toddler, who passes out and dies.

STABBING. John is offended by a woman’s mocking remark and decides to hurt her badly. At work the next day, when no one else is around, he picks up a letter opener from his desk and stabs her. She later dies from the wound.

AMBUSH SHOOTING. John knows the address of a woman who has highly offended him. As he had planned the day before, he waits there for the woman to return from work and, when she appears, John shoots her to death.

ABDUCTION SHOOTING. A woman at work reveals John’s misdeeds to his employer, thereby getting him fired. John devises a plan to get even with her. The next week he forces the woman into his car at knife point and drives her to a secluded area where he shoots her to death.

BURNING MOTHER FOR INHERITANCE. John works out a plan to kill his sixty-year-old invalid mother for the inheritance. He drags [her] to her bed, puts her in, and lights her oxygen mask with a cigarette, hoping
to make it look like an accident. The elderly woman screams as her clothes catch fire and she burns to death. John just watches her burn.

RANSOM, RAPE, TORTURE & STRANGLING. John kidnap's an eight-year-old girl for ransom, rapes her, then records the child’s screams as he burns her with a cigarette lighter, sending the recording to her parents to induce them to pay his ransom demand. Even though they pay as directed, John strangulates the child to death to avoid leaving a witness.
RESPONSES

Realism, Punishment, and Reform

Paul H. Robinson,† Owen D. Jones,‡‡ & Robert Kurzban‡

INTRODUCTION

The discussion here concerns the ideas set out in three articles, each with a different set of coauthors: Concordance and Conflict in Intuitions of Justice1 ("C&C"), The Origins of Shared Intuitions of Justice2 ("Origins"), and Intuitions of Justice: Implications for Criminal Law and Justice Policy3 ("Implications"). Those pieces were an attempt to change the way legal scholars think about intuitions of justice. Professors Donald Braman, Dan Kahan, and David Hoffman ("BKH") offer some criticisms. Some we do not disagree with. Others we do.4

We concede at the start that our past discussions must have been insufficiently careful in their language, as evidenced by the fact that BKH have misread us as they have. We are in BKH’s debt for having revealed the problem. (We also thank them for their true generosity in supporting us in our discussions with the Law Review about writing this Response, and thereby giving us the opportunity to make our positions clear.)

The most important exercise here may be to segregate our false disagreements with BKH from our real disagreements. We suspect that we do have some important disagreements. Part I quickly sketches out our line of analysis in the original articles. Part II examines claims that

† Colin S. Diver Professor of Law, University of Pennsylvania Law School.
‡‡ New York Alumni Chancellor’s Chair in Law and Professor of Biological Sciences, Vanderbilt University; Director, MacArthur Foundation Law and Neuroscience Project.
‡ Associate Professor of Psychology, University of Pennsylvania.
We thank Stephanie Wehrenberg for her excellent research assistance.

1 Paul H. Robinson and Robert Kurzban, Concordance and Conflict in Intuitions of Justice, 91 Minn L Rev 1829 (2007).
4 Two of us (Jones and Kurzban) respond separately to BKH’s analysis of the issues in Origins. See Owen D. Jones and Robert Kurzban, Intuitions of Punishment, 77 U Chi L Rev 1633 (2010).
BKH attribute to us that are not our views. Part III considers possible points of real disagreement with BKH.

I. A BRIEF OVERVIEW

The common wisdom, at least before the publication of C&C, was that there was nothing on which people agreed or ever could agree regarding what was just punishment in specific cases. This would seem to follow from the fact that such judgments seem so subjective, so complex, and so value dependent.

Yet C&C shows the common wisdom to be false. There are some points of agreement—indeed agreement that exists across demographics and at a high level. It is not agreement on the absolute amount of punishment that a particular offender deserves, but rather agreement on the relative blameworthiness among different cases. And it is not agreement on cases involving all offenses and all factors, but only a select few—what we label the “core” of wrongdoing because they represent the point of high agreement. As the second half of C&C—the disagreement study—shows, as one moves out from this core of agreement by adding other factors or offenses, the extent of agreement among people breaks down. The agreement study in C&C finds high agreement only in offenses of theft and violence; the disagreement study shows disagreement in a wide variety of offenses like drunk driving, drug offenses, date rape, prostitution, alcohol use, abortion, and bestiality. The larger point here is that there is not a sea of disagreement on everything, as was thought, but rather a continuum from high agreement to high disagreement, from a small core of issues on which there is almost near unanimity to increasing disagreement as one moves out to the periphery.

The single most interesting, indeed perplexing, finding of the C&C research was that a core of agreement exists at all. How could this be so? How could people’s views about relative blameworthiness—which seem so highly subjective and complex—ever produce such high agreement on any issues in any context? This was indeed a puzzling development, which we sought to explore in Origins.

How do humans come to their judgments about relative blameworthiness? Are those judgments in some part the product of a specialized learning system, or just the product of the standard general

---

5 See, for example, authorities cited in Robinson and Kurzban, 91 Minn L Rev at 1846–48 (cited in note 1).
6 Id at 1891 (explaining that “physical injury, taking without consent, and deception in exchanges” fall within the core of wrongdoing).
7 Id at 1883–87.
8 Id at 1890–92 & n 230.
learning mechanism? We conclude in Origins that, while there are reasons to prefer the former explanation, there is insufficient evidence yet to settle the issue.9

For our larger project regarding the relevance of intuitions of justice to law, however, it is irrelevant which of these two mechanisms produces the shared intuitions of justice. As we note at the end of Origins and reemphasize in Implications, both of the alternative mechanisms lead to the same conclusion on the point that is relevant to criminal law: whatever the source of people’s shared intuitions of justice, those shared intuitions are something to which system designers and social reformers would be wise to give special attention.10 Reformers ought not assume that they can simply educate people out of a core intuition of justice the way they would persuade people to change their views on purely reasoned matters. This is true if core judgments about justice are the result of a specialized mechanism, but it is also true even if they are formed through general social learning. The key fact is not the source of the agreement but rather the high agreement across demographics, for it suggests that whatever the source of the judgments of justice, they are deeply embedded and not easily modified. (The reader may see why we think it somewhat awkward to label us “Punishment Naturalists,”11 when the source of relative blameworthiness judgments is irrelevant to the main argument.)

Imagine the variety of factors that can influence human judgments about relative blameworthiness. Different demographics—income, education, race, political orientation, marital status, religion, gender—can create profound differences in life experience. The fact that any core of high agreement exists itself suggests that such intuitions are held in such a way as to be insulated from the standard influences of everyday life. If that is true, how are social reformers to significantly change those views? Intuitions at the core are perceived by people not as reasoned conclusions but as facts—and what seem to them to be quite obvious facts. People generally have little access to

---

9 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1682–87 (cited in note 2) (noting that several factors, such as the similarity of intuitions across cultures and the way in which children appear to develop moral intuitions, make it appear more likely that the judgments are part of a specialized learning system).

10 See id at 1687–88 (rearking that it may be “unrealistic to expect the population to ‘rise above’ its desire to punish wrongdoers”); Robinson and Darley, 81 S Cal L Rev at 38 (cited in note 3) (arguing that our shared intuitions of justice are not easily changed and can influence the community’s judgments of the law’s credibility).

11 See Donald Braman, Dan M. Kahan, and David A. Hoffman, Some Realism about Punishment Naturalism, 77 U Chi L Rev 1531, 1532–33 (2010) (labeling as “Punishment Naturalists” those who believe that “highly nuanced intuitions about most forms of crime and punishment are broadly shared because they are innate”).
why they feel the way they do about these matters.\textsuperscript{12} While it is certainly possible for rational argument to override even a deeply felt core intuition of justice, we suspect that liberal democracies are not likely to permit the kind of coercive intrusion into the lives of citizens that may be required to change these shared core intuitions—intuitions that have already shown themselves to be immune from even the powerful influence of demographic differences.\textsuperscript{13}

As we explain in \textit{Implications}, the reduced malleability of the high-agreement issues at the core has implications for social reformers.\textsuperscript{14} Reformers should understand that some views will be easier to change than others and should be smart in the battles they choose to fight and how they fight them. For example, they ought to think carefully before they invest their limited time and energies in a program to persuade the community not to want to punish serious wrongdoing, as some reformers would do,\textsuperscript{15} or to persuade the community that serious resulting harm, such as causing death, ought to be insignificant in assessing punishment, as others would do.\textsuperscript{16} \textit{Implications} also gives reformers insights on how best to change people’s judgments about justice.\textsuperscript{17} The greater the level of existing disagreement, the greater the likelihood that people’s views on relative blameworthiness can be successfully modified. \textit{Implications} contains a section for reformers, showing how they can use the high agreement on the core issues to help shift people’s views on issues out from the core.\textsuperscript{18} (In other words, we are not “Punishment Naturalists” but rather “Reform Realists.”)

\begin{flushleft}
\begin{footnotesize}
\textsuperscript{12} Robinson and Darley, 81 S Cal L Rev at 4–8 (cited in note 3).
\textsuperscript{13} See Robinson and Kuzban, 91 Minn L Rev at 1855–61 (cited in note 1).
\textsuperscript{14} Robinson and Darley, 81 S Cal L Rev at 58–59 (cited in note 3).
\textsuperscript{15} See authorities collected in id at 11–12 & n 34.
\textsuperscript{16} See, for example, Larry Alexander, Kimberly Kessler Ferzan, and Stephen Morse, \textit{Crime and Culpability} 172–96 (Cambridge 2009) (arguing that culpability, not the results of an action, should be considered in determining blameworthiness and appropriate punishment); Stephen J. Schulhofer, \textit{Harm and Punishment}, 122 U Pa L Rev 1497, 1498–1503 (1974) (suggesting that the emphasis on harm in the criminal system can “be understood as a vestige of the criminal law’s early role as an instrument of official vengeance” and advocating “a full-scale rethinking of this aspect of the criminal law”). Indeed, the Model Penal Code drafters would take this approach: they seek to make resulting harm insignificant by grading attempts the same as the completed offense except in cases of murder. See MPC § 5.05(1) (ALI 1962) (stating that, with some exceptions, “attempt, solicitation, and conspiracy are crimes of the same grade and degree as the most serious offense that is attempted or solicited or is an object of the conspiracy”). But most states, even those following the Model Penal Code, reject this approach and grade completed offenses higher than attempt. See authorities collected in Paul H. Robinson, \textit{The Role of Harm and Evil in Criminal Law: A Study in Legislative Deception}, 5 J Contemp Legal Issues 299, 305 n 18 (1994) (listing thirty-seven codes that authorized lower sentences for attempted crimes).
\textsuperscript{17} See Robinson and Darley, 81 S Cal L Rev at 51–65 (cited in note 3).
\textsuperscript{18} See id at 60 (suggesting that some attempts to change intuitions about wrongdoing—such as antismoking campaigns and Mothers Against Drunk Driving—have been successful
\end{footnotesize}
\end{flushleft}
One may ask whether there is any reason why one should care about people’s judgments of justice, or about changing them. Certainly the history of modern crime control, with its focus on general deterrence and incapacitation, shows considerable indifference to whether the distribution of punishment provided by those programs conflicts with people’s shared intuitions of justice. 19 We argue in Implications, and studies have since empirically supported us, 20 that it may be very costly for the criminal law to adopt principles for assessing criminal liability and punishment that conflict with the shared intuitions of justice of the community it governs. Gaining a reputation for “getting it wrong”—for regularly and intentionally relying upon rules that do injustice—can promote subversion and resistance to the system, undermine the effective (yet cheap) normative influence of stigmatization, reduce people’s willingness to defer to the law in cases of normative ambiguity, and subvert the criminal law’s ability to shape community norms and to induce people to internalize the norms expressed in the criminal law. 21

II. WHAT IS INCLUDED IN THE “CORE”?

BKH represent us as claiming that:

(1) the vast majority of wrongful acts are part of the core of agreement,” 22

because they “demonstrat[e] that the conduct at issue really does have the condemnable character or effect that people’s intuitions abhor”).


20 See Robinson, Goodwin, and Reisig, 85 NYU L Rev at *41–61 (cited in note 19) (explaining that “doing injustice and failing to do justice can undermine the criminal justice system’s moral credibility” and listing other studies that have reached similar results).

21 See Robinson and Darley, 81 S Cal L Rev at 18–31 (cited in note 3) (noting that “the criminal justice system’s power to stigmatize depends on the legal codes having moral credibility in the community,” and that laws have less moral credibility when they do not match the moral intuitions of individuals).

22 BKH represent us as claiming that people have broadly shared intuitions about “most forms of crime and punishment,” Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1532–33 (cited in note 11), and “the vast majority of wrongful acts.” Id at 1600. As support for their characterization of our view, BKH cite Robinson and Kurzban, 91 Minn L Rev at 1867 (cited in note 1)—the methodology discussion of the study, not the analysis or implications section. See Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1543–44 (cited in note 11). That passage simply points out that we picked common offenses for the study. The passage makes no claim that there is high agreement regarding all offenses, as BKH represent. It also makes no hint that people agree on all aspects or instances of these offenses, as BKH also represent.
(2) people agree as to all aspects of all offenses that are part of the core, and

(3) people also agree as to what conduct should be criminal, and what conduct is justified.

Unfortunately, or fortunately, these are not claims that we make. In fact, we do not know any respectable scholar who makes such claims, but if one can be found we would be happy to join BKH in a battle against these claims and to join BKH in the predictable joint victory.

A. Regarding (1) and (2)

What is the “core”? BKH suggest that its contours are quite vague and difficult to identify. But what constitutes the “core” is not a matter of speculation or theory, or even of interpretation. It is a matter of empirics. The “core” is, by definition, that on which there is high agreement across demographics, like that demonstrated in the C&C agreement study.

What cases are included in the core? Those cases on which there is high agreement across demographics. What kinds of cases are those? The C&C agreement study showed that one could find this level of agreement in cases involving offenses of violence and theft.

23 BKH explain, as if it were in contradiction of our view, that there are “significant controversies within the three categories of core offenses” and “there is substantial disagreement about what constitutes wrongdoing” as related to the core offenses. Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1566 (cited in note 11).

24 BKH say, for example: “Another reason to be skeptical of the suggestion that we share intuitions about most classes of wrongful acts is that the classes of acts listed also exclude acts that a substantial number of Americans believe should be crimes, but which are not.” Id at 1554. “[N]aturalists hold that while individuals may disagree about how much to punish bad acts, they agree on what constitutes a bad act. As such, on the whole, the population should agree that, in each case, the defendant is either guilty or innocent.” Id at 1590.

25 For example, BKH begin their article with a justification case on which they note there is disagreement, id at 1532 (describing a moral dilemma over whether it is appropriate to steal a bus ticket from a wealthy passenger to avoid missing a best friend’s wedding), as if this were in contradiction to our view. For our discussion of their hypothetical, see text accompanying notes 41-45. BKH also repeatedly discuss the justification of self-defense. See Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1578–88 (cited in note 11).

26 BKH say, for example: “We doubt that naturalists will discover some independent way to distinguish the core of harms from the periphery.” Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1566 (cited in note 11), and “One reason to be unsatisfied with the core–periphery distinction is that it fails to tell us what, exactly, distinguishes the important core from the unimportant periphery of crimes.” Id at 1557.

27 See Robinson and Kurzban, 91 Minn L Rev at 1874–80 (cited in note 1). In Origins, we certainly give reasons why we think there is high agreement on the cases at the core, but these theories are not the definition of the core and rather are offered as possible explanations of the core. See generally Robinson, Kurzban, and Jones, 60 Vand L Rev 1633 (cited in note 2).

28 Robinson and Kurzban, 91 Minn L Rev at 1867 (cited in note 1) (finding a high level of agreement among participants’ intuitions about short scenarios involving theft and violence).
Could there be other offenses where one might find cases of such high agreement? Possibly. That will require further research. As we explain in the methodology section to the C&C agreement study, we focused on the most common offenses. If there are other offenses, they are not likely to be common offenses.

What aspects of these offenses are included in the core? BKH seem to assume that we claim that all aspects, all cases, involving any of these offenses are part of the core, but this could hardly be the situation. Our research used factors upon which we judged there was high agreement. To the extent that one substitutes a factor on which there is disagreement, obviously the level of agreement on the relative seriousness of the case would have to decline. If, instead of stealing a clock radio from the car, as appears in one scenario, the offender steals the ashes of the car-owner’s father, obviously the prior high agreement on the relative blameworthiness of the offender will diminish in proportion to the extent of the disagreement on the relative value of a father’s ashes, which might vary widely across cultures and even across individuals within a culture.

What cases beyond those in the C&C agreement study are included in the “core”? We would feel quite comfortable extrapolating well beyond the specific scenarios used in the study. Most objects have an agreed-upon value, like the clock radio; most do not have the disputed value of a father’s ashes. But until the research is done, of course, we cannot know for sure what level of agreement attaches to what facts.

The point of C&C’s Appendix B is to show the reader just how we were able to construct the twenty-four scenarios on which our subjects had such high agreement: by relying upon, and only upon, principles that we knew were deeply embedded intuitions of near unanimity. Specifically, there exist a number of general principles of liability and punishment that are widely accepted. For example, damage to person is more blameworthy than damage to property; purposely

---

29 Id at 1867 n 172.
30 BKH say:

If individuals have an intuitive sense of the relative wrongfulness of acts, then we would expect people with [different] cultural profiles . . . to agree—perhaps not on precisely how much punishment a person deserves, but at the very least on the relative culpability of [] two defendants. For naturalism, dissensus in the core of wrongdoing remains a puzzle.

Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1592 (cited in note 11).
32 We are not entirely ignorant about such matters. There is existing research that gives some hints on a wide range of criminal law issues. See, for example, Paul H. Robinson and John M. Darley, Justice, Liability and Blame 13–51 (Westview 1995) (exploring factors that contribute to blameworthiness).
causing a harm is more blameworthy than doing so recklessly, which in turn is more blameworthy than doing so negligently; the greater the extent of a personal injury, the greater the blameworthiness; the greater the extent of an expectation of privacy, the greater the blameworthiness of an intrusion; and so on. Our suspicion was that the high agreement on these general principles had practical consequences for the level of agreement in individual cases as well, suggesting that one could create high agreement in judging the relative blameworthiness of individual cases, in contradiction of the common wisdom. The C&C agreement study confirmed that the high agreement on the general principles does translate into high agreement on individual cases.

As you can imagine, we found the BKH article quite difficult to understand, given its false assumptions about our claims. For example, it has an entire section showing disagreements in cases of deception in exchanges. Whether somebody is deceived in an exchange obviously is a function of one’s expectations about the terms of the exchange, and those expectations could be highly culturally dependent or, even within a culture, highly dependent on context. The case we used in the study was one of a store clerk shortchanging a customer. We used it precisely because it seemed to us that such shortchanging offered an example of a violation of a nearly universal expectation of this most common form of exchange, a purchase. People who agree to make an exchange typically believe that they have a shared understanding of the terms of the exchange. Thus, within any culture, there are likely to be shared expectations when an exchange takes place but, obviously, not always.

Even more puzzling is that BKH have a section showing disagreements about sexual offenses, as if this were in contradiction of our claim. Recall that the C&C disagreement study itself demonstrates that there is high disagreement with respect to many kinds of sexual offenses. It is hard to know why BKH would want to lecture us on disagreement regarding sexual offenses when our study may be the best available empirical evidence in support of that disagreement.

---

33 Robinson and Kurzban, 91 Minn L Rev at 1867–80 (cited in note 1) (describing two studies that predict and demonstrate agreement among subjects who ascribe blameworthiness to specific scenarios).
35 Id at 1573–77 (noting that law concerning rape “has been a site of intense legal and political conflict for over thirty years”). BKH also discuss disagreement over the criminality of the sexual offenses of prostitution, bestiality, and unwanted sex. Id at 1555–56.
36 Robinson and Kurzban, 91 Minn L Rev at 1885 table 6, 1886 table 7 (cited in note 1).
Generally, BKH have our claim reversed. The point of the C&C agreement study is not to show that people agree on everything, or even a lot, but rather to expose as false the common wisdom of the day that, at least in terms of individual cases, people agree on nothing. It is difficult for us to see the value in BKH presenting an exaggerated view of our claims, then criticizing them for being exaggerated.

The important point here is that, simply because one can find or create disagreement by introducing facts on matters in dispute, it does not take away from the fact that there does exist a core of high agreement. And, as discussed further in Part II below, that high agreement has implications for the malleability of people's intuitions of justice.

B. Regarding (3)

We see the same pattern of misunderstanding when BKH represent us as claiming that there exists high agreement as to what is criminal. A primary theme of the C&C disagreement study is to show just the opposite—it shows not just disagreements about relative blameworthiness, but also about whether the conduct should even be criminal. In two of the scenarios, more than 20 percent of the subjects assign no liability. In another, more than 40 percent find no liability. (This is so even though all three scenarios are criminal under current law.)

Indeed, even in our C&C agreement study, we specifically demonstrate that people can agree on the relative blameworthiness of different offense cases yet disagree as to where on the continuum of blameworthiness the line should be drawn marking off the minimum point for criminal liability and punishment. We designed one of our scenarios as an intentionally borderline case (taking food away from an “all you can eat” buffet in violation of the buffet rules). It was no surprise that people disagreed about whether the case should trigger criminal liability. The important point to us was to show that, despite this, there was still near consensus on where the case ranked on the continuum of relative blameworthiness. The larger point here is that it is judgments about relative blameworthiness on which people can have high agreement, not necessarily judgments about exactly

---

37 Id at 1831–32.
38 See id at 1885 table 6 (showing that 23.4 percent of participants would assign no liability for prostitution, 21.9 percent of participants would assign no liability for marijuana use, and 42.2 percent of participants would assign no liability for bestiality).
39 Id at 1876, 1900-01.
40 Robinson and Kurzban, 91 Minn L Rev at 1869, 1900 (cited in note 1).
where the line is to be drawn for minimum blameworthiness for criminal liability.

BKH similarly misunderstand our view as claiming that people agree not only on all aspects of what is criminal but also on what is justified. 41 That is, they go beyond what criminal law theorists call “prohibitory norms,” or notions about what should be prohibited as wrongful, to include “justificatory norms,” or notions about what admittedly wrongful conduct might be tolerated under special justifying circumstances. (Indeed, BKH give no indication that they are aware of this distinction or that they have crossed from one to the other.) They begin their article with a hypothetical about a person deciding whether to steal a bus ticket in order to make an important appearance at a wedding. They point out that Americans might think it wrong to take the ticket in this situation but that Indians might not, suggesting that we would claim that everyone would agree about this case.

Contrary to the way in which they present it, the hypothetical is not a simple case about whether people disagree on whether theft is wrongful. Presumably the person in the hypothetical believes that stealing is wrong. The issue presented is a different one: whether the conduct (theft of the ticket), which all agree is wrongful, may nonetheless be justified in this instance because of a special competing interest of sufficient importance (the need to get to the wedding) that might justify the otherwise wrongful conduct. The case is not a test of whether people think theft is wrongful, but rather a test of the comparative value of the competing justification interests.

If one compares interests on which people agree (or that are so disparate that people’s minor disagreements are irrelevant), then one will get agreement. On the other hand, if one compares interests on which people disagree—like the value of a father’s ashes or the importance of getting to a wedding—then people will disagree on the comparison. In this instance, the value one places on the importance of one’s wedding responsibilities is culturally dependent. BKH suggest that Indians give it greater value than Americans, and we do not know enough about it to disagree.

Consider, however, what else the hypothetical might illustrate. We suspect that there is high agreement on the general principle that special justifying circumstances can outweigh the prohibition of conduct that is itself wrongful. And, as we have shown in the C&C agreement study, that high agreement on a general principle can translate into high

41 See note 25.
42 Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1531, 1533–34 (cited in note 11) (suggesting that American participants tended to focus on “individualized justice and personal property,” whereas Indian participants focused on “social and relational responsibilities of friends”).
agreement in individual cases, if those cases involve only matters on which there is agreement. 43 Introducing into a case a factor on which there is disagreement obviously reduces the previous agreement.

BKH might ask themselves why it is that both Americans and Indians, and every other group on the planet that recognizes property rights, agree that taking another person’s property of value is wrongful and thus would require some justification. No doubt there is considerable cultural diversity about what counts as “another’s property,” as well as considerable diversity on the value that different cultures place on different kinds of property.4 But how is it that diverse cultures all seem to agree with the basic rule that taking another person’s property of value without consent is wrongdoing? Indeed, there is probably further agreement: the greater the value, the greater the wrongdoing. Do the “Punishment Realists” believe that they can find societies, or could create societies, in which people would believe that there was nothing wrong with taking another’s property without consent, or that, if people see such taking as wrongful, they could be made to believe that the greater the value of the property taken, the lesser the wrongdoing? More on this below.

III. POSSIBLE POINTS OF DISAGREEMENT WITH BKH

As one might imagine, we have found it difficult to engage with much of the BKH article, based as it is on a representation of our views that we often do not recognize. The experience has left us with little confidence that we understand BKH’s position, and thus we are hesitant to make claims about their views. On the points below, we suspect there may be real disagreement, but we think it prudent simply to identify possible points of disagreement and leave it to BKH to confirm or deny that these are their views.

A. There Is No Core of Agreement

It might be that BKH’s view is that there are no aspects of any offenses on which there is high agreement regarding an offender’s relative blameworthiness. This, of course, was the common wisdom prior to the C&C study.46 It is hard to see how this could continue to be a

43 Robinson and Kurzban, 91 Minn L Rev at 1880–81 (cited in note 1).
44 See, for example, Keith Aoki, Considering Multiple and Overlapping Sovereignties: Liberalism, Libertarianism, National Sovereignty, “Global” Intellectual Property, and the Internet, 5 Ind J Global Legal Stud 443, 462–63 (1998) (discussing the way that countries’ differing views about property and ownership can lead to significant disputes in the intellectual property context).
45 Robinson and Darley, Justice, Liability and Blame at 84–94 (cited in note 32).
46 Robinson and Kurzban, 91 Minn L Rev at 1847 (cited in note 1) (noting that many writers “have argued that people simply disagree in their notions of justice” in a way that prevents
plausible claim, yet BKH seem to act as if this were the case by continuing to deny that there is a difference between core and noncore cases."

B. There Is a Core of Agreement but It Is Meaningless

It may be more likely that the BKH position is slightly different. They may concede that we have shown a core of high agreement but may believe that this is in some sense a false appearance of agreement, or at least an appearance of agreement that has no real significance. If that is their view, then presumably BKH are claiming that the kinds of cases in our agreement study, which we identify as part of the "core" of high agreement, are no different in terms of their potential for agreement or disagreement from the kinds of noncore cases in our disagreement study. That is, there is nothing meaningful about the apparent core of agreement that we show, they may be arguing, because people's agreement or disagreement about relative blameworthiness is something that can be created and dissolved and recreated and redissolved at will. No set of issues relating to blameworthiness is any different from any other set of issues in this respect; all are subject to the same potential for agreement or disagreement. The results of our C&C agreement study (and presumably our disagreement study as well), they may argue, show just one of an infinite variety of patterns of agreement and disagreement that a researcher or reformer could create or dissolve at will by manipulating the facts or by giving reasons.

We think that BKH are simply wrong on the empirics. We offer them this challenge: Using the noncore cases in the C&C disagreement study, create the same level of agreement across demographics that we did using core cases in the C&C agreement study. If core and noncore cases have no meaningful difference, as BKH seem to believe, then they should be able to create the same level of high agreement using the noncore cases that we claim is unique to the core cases. We

agreement on blameworthiness). See also, for example, Michael Tonry, Obsolescence and Immanence in Penal Theory and Policy, 105 Colum L Rev 1233, 1263 (2005) ("How do we know how much censure, or 'deserved punishment,' a particular wrongdoer absolutely deserves? ... [A]s countless sentencing exercises have shown, people's intuitions about individual cases vary widely.").

47 See Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1566 (cited in note 11).

48 BKH say, for example: "None of [the categories that constitute 'the core of wrongdoing'] is composed of acts free from dissensus, and the nature of the systematic dissensus that pervades each of these categories is at least as interesting and informative as any agreement that can be found." Id at 1568-69 (emphasis altered).

49 For BKH's view that no meaningful distinction exists between the core and the noncore cases, see note 26. BKH also argue that there is disagreement both among the core cases and among the noncore cases, so the distinction is not meaningful with regard to disagreement. Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1566 (cited in note 11).
do not believe it can be done. Contrary to BKH’s claim, core and noncore cases are importantly different.

C. Even If There Is a Core of Agreement, It Has No Effect on the Malleability of People’s Judgments about Justice

We think the core of high agreement has practical significance, because these intuitions will be harder to modify than judgments on other cases and issues. That the high agreement on the core cases can be reduced if one adds factors about which there is disagreement does not make people’s intuitions concerning the core cases any more malleable. The issues at the core remain as difficult to modify whether they appear with or without factors on which there is disagreement.

BKH seem to think that all judgments about justice are malleable. They seem to think that social reformers should see all judgments about justice as equally fair game for modification. Again, we believe that BKH are simply wrong on the empirics. As we note above, the existence of a high degree of agreement across demographics shows that the view is sufficiently deeply embedded as to be insulated from the powerful forces of social influence inherent in the wide variety of demographic factors at work in the world. If these core intuitions are immune from the influence of these forces, why would one expect that they nevertheless would be susceptible to easy modification by reformers?

---

50 Nor do we believe that BKH can construct a set of scenarios on which there is low disagreement as to their proper ranking, then modify the scenarios by adding information and thereby create high agreement. If BKH are correct that there is no such thing as a core of agreement, surrounded by factors and offenses of increasing disagreement out from that core, then BKH should be able to create, dissolve, and recreate agreement simply by their manipulation of the facts. We do not believe that it can be done. Once the high agreement of the core is destroyed by adding facts on which there is disagreement, there is no getting it back by adding more facts. The only way to get it back is to drop the disagreement-inspiring facts that were added. There is a core of agreement that is different from other blameworthiness issues and cases.

51 BKH say, for example: “Punishment Realism is based on the premise that while individuals do hold deep and abiding intuitions regarding wrongdoing and responses to it, these intuitions depend on social constructs that are demonstrably plastic.” Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1533 (cited in note 11). “The diversity of positions political communities have adopted on such issues—over place and over time—makes us conscious of the plasticity of social norms.” Id at 1535. “Where we see mutability in norms, and hence the inescapability of collective responsibility for their content, the naturalists apprehend their stability and warn of the futility and even perversity of using criminal law as an instrument of norm reform.” Id at 1536. BKH want to “learn how our moral intuitions are shaped and develop means of fostering conceptions of justice that are both satisfying to us and compatible with our collective welfare.” Id at 1532.

52 See note 12 and accompanying text.
The most obvious data point here is the shared intuition that serious wrongdoing should be punished. There is an active abolitionist movement, which we discuss and document in Implications. But given the strength and near unanimity of the intuition that serious wrongdoing should be punished, it strikes us as unrealistic that this movement will ever gain much support, let alone be implemented by any society. If, as BKH claim, this intuition can be modified, and if, as the abolitionist arguments make clear, there are good reasons to think that punishing wrongdoing might not be the best course for a society, why have no societies (of which we are aware) ever rejected having a punishment system? Given the diversity of societies throughout the world and across time BKH owe us at least an explanation of why this particular intuition of justice has apparently never been overridden.

Indeed, one can read BKH as essentially conceding our point that there are some intuitions of justice so deeply embedded as to make it unrealistic to think that they can be changed. BKH suggest that reformers need not worry about widely shared intuitions about justice because they do not "address any live policy debate." Presumably, they mean that the current debate about the abolition of punishment, for example, is a "dead" debate—which is simply another way of saying that it is unrealistic to expect that the abolitionists could ever prevail. But if BKH think that all intuitions of justice are malleable, without regard to the degree of agreement on an intuition, why should the debate not be a "live" one? How can BKH see the abolitionist debate as not "live" without conceding that there are important and predictable differences in malleability that attach to core intuitions?

In fact, BKH also are wrong in claiming that intuitions of justice that are hard to change are at the heart of "live" debates. As we discuss in Implications, the abolitionist debate has taken on a more modern form in the guise of "restorative justice." There can be little doubt that the founder of that movement, John Braithwaite, has as his principal motivation an antipunishment agenda. There are many different forms

---

53 As we say above, where to draw the line of demarcation between criminal and noncriminal conduct can be a matter of dispute. See text accompanying note 39. For those cases that are seen as serious wrongdoing, however, there is little dispute that punishment should be imposed. See Robinson and Darley, 81 S Cal L Rev at 9 (cited in note 3).
54 Robinson and Darley, 81 S Cal L Rev at 11–18 (cited in note 3).
56 See Robinson and Darley, 81 S Cal L Rev at 12–13 (cited in note 3) (explaining that restorative justice seeks to help the victims of a crime, the offenders, and the communities that were affected).
57 See John Braithwaite, A Future Where Punishment Is Marginalized: Realistic or Utopian?, 46 UCLA L Rev 1727, 1746 (1999) (characterizing restorative justice "as a competitor to
of restorative processes, some more inclined to defeat deserved punish-
ishment than others. The debates over which processes to prefer are
in large part debates over the extent to which just punishment can and
should be frustrated. Consistent with our prediction, restorative
processes are currently typically limited in their application to cases
involving juveniles or minor offenses, precisely because broader ap-
lication to more serious offenses would conflict with people’s shared
intuitions that serious wrongdoing should be punished. One of us has
argued in print that there could be great value in using restorative
processes in cases involving serious offenses and that reformers would
be better advised to drop their antipunishment agenda in order to
promote such broader use.58

To go beyond the broad intuition that serious wrongdoing should
be punished, consider an example of a specific criminal liability rule.
People overwhelmingly have a strong intuition that resulting harm
matters—for example, that murder should be punished more than
attempted murder and that manslaughter should be punished more
than reckless endangerment, even if whether a death results is a mat-
ter of bad luck. The strength of this widely held intuition has been
repeatedly documented across demographics and cultures.59

It is easy enough to see the rational argument against correlating
punishment with the resulting harm, as every criminal law professor
has used to regale his or her class. Why should an offender’s liability
and punishment vary because of a factor over which he has no con-
tral—the intended target happens to bend down to tie his shoe at the
moment the shot is fired; the pedestrian who would have been killed
by the reckless driver is running a bit late that day and gets to the
crosswalk just after the reckless driver has passed.

No doubt some professors are able to persuade some members of
their criminal law classes that rational analysis supports ignoring re-
sulting death. But there is a difference between persuading a student
of the irrationality of the rule and getting the students to feel that jus-
tice requires that resulting death be ignored. And even if one could
persuade a student on the justice point—and some minority of the
particularly rational students are indeed persuaded each year—it does
not follow that this victory of intellectualization over intuition could
be repeated for the general public, and certainly not at the rate that

59 See, for example, Robinson and Darley, Justice, Liability and Blame at 14-28, 181-89
(cited in note 32); Robinson, 5 J Contemp Legal Issues at 306-07 (cited in note 16) (finding that
97.3 percent of a study’s participants believed that an offender who murders his victim should
receive a harsher punishment than an offender who attempts to murder his victim).
would change a shared societal view. Anyone who thinks this kind of reform of a core intuition is possible has not spent enough time talking with ordinary people.

If BKH were to begin a broad-based campaign to change people's intuitions about the significance of resulting death, it is likely that the vast majority of people would not take them seriously. Remember, people tend to see such intuitions as analogous to observable facts: there is no need for discussion; the fact is clear and obvious. If it is nearly impossible to persuade more than a minority of one's highly rational criminal law students on the justice point, it seems hard to imagine what it would take to create a consensus, or even strong support, among a broad community. It should give BKH some pause that, no matter how irrational it may seem to give significance to resulting death, we are aware of no societies on earth in which people support a conception of blameworthiness that ignores it.

We think BKH are wrong not only to dismiss the difference between core and noncore issues as relevant to malleability, but also to ignore the difference between intuitional and reasoned judgments. (We suspect that all aspects of the core are intuitional judgments, but matters out from the core also may be intuitional in part.) As we discuss in Implications, intuitions about justice have quite different qualities from reasoned judgments about justice. The former are perceived as facts, held with great confidence, and give the holder little access to why she holds them. It would seem obvious that the distinction between intuitive and reasoned judgments about justice would be important to reformers. As to the latter, the reformer can change the judgment simply by presenting a better-reasoned argument. But changing the former—the intuitive judgment—requires something more. To start, the reformer must get the person to think that her strongly held intuition is worth reexamining. And, even if that hurdle is cleared, it does not follow that the intuition can be changed simply by presenting a reasoned argument. Lay people are not going to disavow the significance of resulting harm simply because a law professor can demonstrate its irrationality. When BKH ignore the difference between intuitions of justice, like those at the core of agreement, and reasoned judgments about justice, like those out from the core, they do social reformers a disservice. To be effective, a reform program aimed at changing intuitions would likely be quite different from one aimed at changing reasoned judgments.

---

60 See Robinson and Darley, 81 S Cal L Rev at 4–8 (cited in note 3).
D. There Is No Value in Having Criminal Liability and Punishment Rules Track the Intuitions of Justice of Those Governed by the Rules

BKH appear to reject the notion that there is a cost to criminal law's reliance upon rules that conflict with people's judgments of justice, our primary claim in Implications. (Note that our argument does not depend upon the existence of global agreement on any justice judgments. It simply urges that a criminal-code reformer adopt rules that will maximize the code's moral credibility within the community that it governs. That is, it provides a workable guide for law reformers even if their community has different notions of justice than those of other communities with other criminal codes.)

BKH certainly had good company in their view a few decades ago; modern crime-control doctrines have been quite indifferent about producing results that conflict with community views of justice. But we think scholars' views are changing in light of common sense, anecdotal evidence, and empirical studies. On purely anecdotal grounds, most people would probably concede that a criminal justice system perceived as unjust will have little or no normative influence and will instead prompt resistance and subversion rather than acquiescence and assistance. The Soviet criminal justice system, for example, lacked moral credibility, and probably few were surprised that it seemed unable to gain much deference in the absence of direct coercive force. This is simply common sense: why would people defer to a system as a moral authority when it has shown itself to be so regularly indifferent to injustice?

Empirical studies have more clearly shown the connection between a criminal justice system's moral credibility and people's willingness to assist and defer to it. Studies reported in Implications, as

---

61 We claim that the criminal law's intentional conflicts with community views can undermine the law's moral credibility, producing detrimental practical consequences, but BKH refer sarcastically to these consequences as "admittedly terrifying prospects." Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1594 (cited in note 11).

62 Robinson and Darley, 81 S Cal L Rev at 31 (cited in note 3) (arguing that the effectiveness of the criminal justice system depends, in part, on whether it is "perceived as 'doing justice'").

63 Id at 25–38.

64 Robinson, Goodwin, and Reisig, 85 NYU L Rev at *7–30 (cited in note 19) (demonstrating that punishments for habitual offenses, drug use, juveniles, mentally ill individuals, and strict liability crimes are much harsher than societal intuitions about just punishments).

65 See Dina Kaminskaia, Book Note, Final Judgment: My Life as a Soviet Defense Attorney, 96 Harv L Rev 1762, 1762 (1983) ("[T]he Soviet legal system achieves legitimacy not through the integrity of the judicial process, but through the underlying coercive force of the state.").

66 Robinson and Darley, 81 S Cal L Rev at 26 (cited in note 3).
well as additional studies since, have shown that by manipulating the former, one can produce a change in the latter. If BKH disagree with this view, what studies showing contrary results have they presented?

E. Our Reform Realism Means We Must Be Conservative Antireformers

BKH claim that our views make us conservative antireformers. We find this claim particularly bizarre. BKH have it backward. Our program is designed to give reformers tools for more effective reform. BKH's "Punishment Realism," in contrast, offers damaging advice that can hurt reform efforts and provides excuses for keeping the status quo in the face of glaringly unjust punishment rules that reformers have long wanted to change.

The analysis set out in Implications suggests two recommendations to social reformers. First, it may often be unwise to invest limited reform resources on trying to change intuitions of justice that will be difficult to change, at least given the resources and authority available to reformers. In other words, one should be smart and pick one's fights carefully.

Second, when developing a program to change people's intuitions of justice, it will often be a better investment to harness people's core intuitions of justice rather than fight them; such core intuitions are a power reformers can use for their own purposes. If one wants people to take domestic violence more seriously, emphasize the violence part and deemphasize claims that it is somehow exempt because it is domestic. If one wants people to see downloading music as more condemnable, build up the analogy to physical taking that we know to be part of the core of agreement.

It is not just these two recommendations that can help reformers, but a more sophisticated understanding of the complexities of making


68 BKH say:

We do feel deep concern ... over what we take to be the politically conservative resonances with which the Punishment Naturalist has been needlessly infused. It is, simply put, extremely difficult to take in the corpus of work that the Punishment Naturalists have amassed without sensing a deep commitment on their part to the status quo—to popular retributive sensibilities as they are (or are depicted with a high degree of uniformity to be), and to laws that conform (or are depicted as conforming) to them.


69 See Robinson and Darley, 81 S Cal L Rev at 57-59 (cited in note 3).

70 Id at 60-66 (noting the success of public education programs that analogize noncore conduct—such as drunk driving or smoking—to core conduct).
strategic choices, including an appreciation for the need of reformers to build and harness the moral credibility of the criminal law in order to use it to help them change norms, as we discuss in Implications.71 We illustrate with an example on which one of us has written: rape reform.72

Imagine a rape reformer who is dissatisfied with the way in which young men are routinely indifferent to obtaining a clear expression of consent before having intercourse with a date. In promoting a revised formulation of a rape offense, should the reformer urge strict liability as to lack of consent, or prefer a culpability requirement of at least negligence? The danger of a negligence standard is obvious: the inquiry into what is “reasonable” may incorporate by reference, and thereby perpetuate, the existing norm of indifference to consent to which the reformer objects. Strict liability seems the more attractive option, because it ensures that defendants who continue to be indifferent will be held liable, thereby encouraging a change in conduct.

With the insights offered in Implications, however, a reformer might come to a quite different conclusion. Reliance upon a strict liability standard increases the chances that some defendants will be seen as blameless, transforming the offender into a “victim.” Further, and more importantly, the criminal justice system, and in particular the new rape offense, risks being seen as unjust, imposing potentially serious punishment on the most and least blameworthy offenders alike. That can be seriously problematic for the reformer, as Implications points out, because the reformer, more than anyone else, needs the criminal law to speak with moral authority if it is going to be able to effectively change people’s norms.73 By undermining the law’s moral credibility, by having it rely upon a strict liability standard that potentially invites perceived injustices, the reformer risks undermining the very quality of criminal law that the reformer most needs. For reformers, criminal law’s greatest effect is not in punishing the particular offender at hand but rather in shaping the norms of the rest of the society.74 Changing people’s internalized norms means influencing the conduct of two people in an intimate situation, even if neither of them would dream of reporting their conduct to legal authorities. Reformers interested in changing conduct must have as their ultimate goal changing norms, not simply changing law.

71 See id at 51–52, 60–66.
73 See Robinson and Darley, 81 S Cal L Rev at 25 (cited in note 3).
74 See id at 28–29.
The larger point here is that reformers have an interest in generally building up the criminal law’s moral authority by adopting rules that will avoid perceptions of injustice and that will enhance the law’s reputation for doing justice, so that they then can use the law’s moral authority to help shift the community’s norms. They have an interest in criminal law earning “moral credibility chips” that can then be “spent” by leading a community to changed norms.79

BKH, in contrast, stand as the protectors of the status quo. As we have argued elsewhere, our program challenges the dominant theory of crime control in the United States for the past several decades, one based upon intentionally and regularly doing injustice in the name of general deterrence and incapacitation by its reliance upon doctrines like three strikes, high penalties for drug offenses, adult prosecution of juveniles, abolition or narrowing of the insanity defense, the felony murder rule, and the use of strict liability. We show that, in judgments of relative blameworthiness, these criminal law doctrines dramatically conflict with the community’s intuitions of justice, and we argue that even the good utilitarian ought to reject these crime-control doctrines because of the injustice they produce.77

BKH’s “Punishment Realism” offers a quite different reaction to these injustices. By discounting the significance of perceived injustice and by offering instead the false lure of widespread malleability about justice judgments, BKH protect the status quo of injustice. They suggest, and some scholars in fact have argued,76 that we ought not worry about these injustices. People’s blameworthiness rankings are all malleable. The lesson that BKH teach is that we can keep the injustices and simply change people’s views about what constitutes injustice. We should train people to think that justice is really whatever most effectively deters or whatever is necessary to incapacitate dangerous offenders, as if such changes were a realistic possibility. Distributing criminal liability so as to optimize deterrence or incapacitation might be a legitimate goal, but, disconnected as it is from moral blameworthiness,

75 See Robinson, Criminalization Tensions at *4 (cited in note 72) (suggesting that a “sophisticated criminal law” system will take “every opportunity to build its moral credibility” so that it will, when necessary, be able to use that credibility to shift societal norms).

76 See Robinson, Goodwin, and Reisig, 85 NYU L Rev at *7–30 (cited in note 19).

77 Id at *41–62 (noting that, although some may worry that forgoing these crime control doctrines could “increase avoidable crime,” imposing punishments that accord with the community’s intuitions might, “in the long run, . . . be the most effective means of fighting crime”).

78 See, for example, Alice Ristroph, The New Desert, in Paul H. Robinson, Stephen P. Garvey, and Kimberly Kessler Ferzan, eds, Criminal Law Conversations 45, 46 (2009) (“Desert requires external values to give it content. If those values change and produce revised sentencing policies—if we decide to emphasize incapacitation over rehabilitation, for example—the assessment of how much punishment is deserved is likely to change as well.”).
it is different from doing justice. More importantly, it is not reasona-
table to think that one could persuade a community that these goals,
rather than moral blameworthiness, are the criteria for doing justice.
Yet it is these kinds of injustice-assuring programs that BKH’s ap-
proach protects, by offering the false hope of malleable community
intuitions of justice.

CONCLUSION

Our program is to make reformers smart and effective. How
BKH have transformed this into a conservative antireformer program
is unclear. We think that reformers ought to invest their limited time
and energy with due regard for the comparative difficulties and poten-
tial effectiveness of alternative strategies. Wasteful or ineffective
reform programs are not to be preferred.

By contrast, what is the positive contribution that BKH’s “Pun-
ishment Realism” provides to enhance the program of social reform?
Is the BKH contribution the insight that the world is full of disagree-
ments about the nature of justice? Who would dispute this? In fact,
there is no danger that anyone would think otherwise given that, not
long ago, the common wisdom was that there existed only disagree-
ments about judgments of justice. Or is the BKH contribution the
insight that people’s views about justice are commonly influenced by
socially dependent factors? That point too seems to be well known
and long understood. How could it be otherwise? Or is the BKH con-
tribution the insight that reformers should try to understand the so-
cially dependent factors? We suspect that anyone in the social reform
business figured this out long ago. On the other hand, there is no
harm in repeating it. The same cannot be said, however, about BKH’s
other advice.

Is it a positive contribution of the “Punishment Realists” to ad-
vise reformers that there is no difference between core and noncore
issues, so reformers ought give no attention to the existence of high
agreement on some issues when they design their reform programs? Is
it good advice that reformers should go ahead and invest in eliminating
people’s demand that serious wrongdoing be punished and invest in
convincing people that resulting death should be ignored? Encouraging

79 BKH say: “Realists just want to know what those extralegal influences are and how they
manifest themselves so that they can better predict legal outcomes and manipulate policy to
enhance whatever social welfare, fairness, or expressive concern they favor.” Bрамan, Kahan,

80 This seems to be the main point of Punishment Realism: that justice judgments com-
monly are influenced by social factors. “For the most part, these extralegal influences will move
legal actors to agree, but sometimes they will move them to disagree.” Id at 1566.
reformers to invest limited resources in programs likely to be ineffective does not seem like a positive contribution.

Is the positive contribution of “Punishment Realism” to advise reformers that a proposed law’s potential for injustice can be ignored, and that such injustice will have no effect on the long-term success of the reform program? Common sense, history, and now empirical studies show this also to be bad advice. Reformers act at their peril when they promote liability rules that will be perceived as unjust by those governed by them.

Ultimately, it is likely that the primary contribution of BKH’s “Punishment Realism” is to debunk a notion of “Punishment Naturalism” to which none of us subscribes. But now that that straw man is on the ground, one may wonder whether “Punishment Realism” has any continuing value.
Intuitions of Punishment

Owen D. Jones† & Robert Kurzban††

INTRODUCTION

Recent work reveals, contrary to widespread assumptions, remarkably high levels of agreement about how to rank order, by blameworthiness, three kinds of wrongs: (1) physical harms; (2) takings of property; and (3) deception in exchanges. We refer to these collectively as the “core” wrongs.

In The Origins of Shared Intuitions of Justice 1 we built off of prior work 2 to propose explanations for the high levels of agreement. We raised two possibilities: (1) such agreement traces to general (not particular) social learning mechanisms; and (2) it traces to effects of evolutionary processes on species-typical brains, which predispose humans to develop intuitions about core wrongs. We concluded that, although present evidence does not exclude the former explanation, the latter is more likely.

In their article published elsewhere in this issue, Professors Donald Braman, Dan Kahan, and David Hoffman (“BKH”) critique an assemblage of articles by ourselves and others (to much of which we respond separately). Their critique includes a vehement disagreement with our preferred hypothesis, 3 to which we respond here. We thank

† New York Alumni Chancellor’s Chair in Law and Professor of Biological Sciences, Vanderbilt University; Director, MacArthur Foundation Law and Neuroscience Project.
†† Associate Professor of Psychology, University of Pennsylvania.
1 Paul H. Robinson and Robert Kurzban, Concordance and Conflict in Intuitions of Justice, 91 Minn L Rev 1829 (2007).
5 See Donald Braman, Dan M. Kahan, and David A. Hoffman, Some Realism about Punishment Naturalism, 77 U Chi L Rev 1531, 1551–66 (2010). For an earlier critique along similar
BKH for initiating an important discussion. But we believe that they have misunderstood key aspects of our hypothesis, causing them to misjudge how well their proposed alternative—Punishment Realism—fares in comparison. Below, we clarify five items.

I. Genes, Plasticity, Innateness, and Culture

BKH ascribe to us several views that neither we nor any scientist we know actually holds. Some are inconsistent with fundamental findings of biology and psychology, on which we necessarily rely.

For example, BKH claim that we argue that people’s shared intuitions of justice (in the three specific contexts) are solely or predominantly the product of “genetically programmed moral instincts.” We do not do so. BKH repeatedly frame our work as denying plasticity in human cognitive and moral development. We do not do so. BKH assert that our hypothesis denies roles to culture and social environment. Quite the contrary.

We emphatically share BKH’s opposition to genetic determinism, their commitment to plasticity in human cognition, and a deep (in fact scientifically unavoidable) commitment to recognizing the crucial role that social environment plays in each individual’s development of intuitions of justice. These misunderstandings, common among critiques of evolutionary perspectives, derive from two core confusions: (1) mistaken interpretation of hypotheses about functional design; and (2) false dichotomies about biological and psychological processes.

A. Functional Design

BKH describe Punishment Realism as based on the premise that while individuals do hold deep and abiding intuitions regarding wrongdoing and responses to it, these


6 Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1536 (cited in note 5).

7 See, for example, id at 1532–33.

8 See id at 1539–40.

9 See Matt Ridley, Nature via Nurture: Genes, Experience, and What Makes Us Human (HarperCollins 2003); Robinson, Kurzban, and Jones, 60 Vand L Rev at 1640–42 (cited in note 2) (citing sources); id at 1677.

Intuitions of Punishment

1635

Intuitions depend on social constructs that are demonstrably plastic. Thus, while there are a number of important (perhaps even universal) features of human cognition that shape our understandings of wrongdoing, they are features that interact with, and enable the construction of, varied social norms rather than produce them in a determinate manner."

This passage misframes our views and the science underlying them.

Evolutionary explanations typically focus on function, a specification of the information-processing systems of the human mind—that is, "cognitive mechanisms"—and what natural selection has designed these mechanisms to do. Because many mechanisms’ functions can only be executed by acquiring information from others, theories about functional mechanisms necessarily entail a commitment to the belief that key parts of psychological development depend on social environments. Put simply, human psychology is strongly influenced by what others do and say because our ultrasocial species evolved this way. Our view is, therefore, like "Punishment Realism," fundamentally premised on the idea that intuitions of justice depend on learning from others and are not "determinate" or otherwise developed independent of social input.

To be clear: a claim that computational mechanisms have functions in no way entails that such mechanisms are inflexible, genetically determined, "fixed," or even "innate," as BKH use this term. (We use "innate" too, but quite differently.) The evolutionary view necessarily incorporates (because it is demonstrably true) the view that every aspect of every organism is the joint product of genes and

---

11 Branan, Kahan, and Hoffman, 77 U Chi L Rev at 1533 (cited in note 5).
12 For a classic description, see generally George C. Williams, Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought (Princeton 1966).
13 See text accompanying notes 23–25.
14 See, for example, Steven Pinker, The Language Instinct: The New Science of Language and Mind (Penguin 1994).
16 For example, it is clear from context that when BKH assert that we believe "intuitions about . . . crime and punishment . . . are innate," 77 U Chi L Rev at 1532–33 (cited in note 5), they use "innate" as a synonym for "genetically determined" or "fixed." For example, they claim that "moral judgments . . . are not innate insofar as they depend crucially on social meaning that varies across cultural groups." Id at 1532. See also id at 1546 n. 59. In contrast, our meaning of innate, consistent with the meaning of the term in biology and psychology, is elaborated below. See text accompanying notes 21–26.
environment. Environments include, it bears repeating, the social environment.

B. False Dichotomies

The second misconception reflects the fact that BKH are mired in the old dichotomy between “nature and nurture” (and its close cousins, “cultural versus biological” and “evolved versus learned”). That was a key axis of debate in social sciences through the early twentieth century. Yet researchers and theorists in biology, psychology, and neighboring fields long ago rejected this dichotomy as false, both conceptually and empirically. Consequently, BKH artificially polarize our respective views, and misframe important issues.

For example, BKH ask: “Is our morality by and large determinate and innate, the product of evolutionary forces acting over millions of years, or do we acquire it within our lifetimes . . . ?” Posing a choice between evolved or acquired morality highlights a fundamental disagreement with BKH. But that disagreement is not about whether morality is evolved or acquired. It is about whether framing the question that way makes sense. We believe, along with most scientists, that it does not.

The question, stated as an either—or proposition, necessarily entails that the answer cannot be both. We hold the majority view that morality is both “not determined” and “innate.” We do not mean “innate” in the superficial sense (as if innate means “genetically determined” or “present from birth”) but rather in the scientific meaning that there are specialized systems that give rise to it as we acquire it within our lifetimes.

---


20 Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1532 (cited in note 5).

21 See, for example, Alan Slater and Gavin Bremer, eds, An Introduction to Developmental Psychology 61 (Blackwell 2003); Arnold Sameroff, A Unified Theory of Development: A Dialectic Integration of Nature and Nurture, 81 Child Dev 6 (2010).

22 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1646 (cited in note 2).
To illustrate, consider how people learn language. Most researchers believe that evidence supports a functionally specialized process in the brain—a “Language Acquisition Device.” This is a set of computational mechanisms designed to take in information—generally speech from other people—and use this information to acquire the ability to understand and produce the language used by others in the social environment. How this information from the world is used to generate full-blown language is specific to the language system, even though the precise form of language varies. That is, learning language requires both local linguistic input and (innate) learning systems specialized for language (as opposed to more general processes).

Consequently, the idea that there is a system in the brain designed to acquire language does not entail that such a system will be “fixed,” “determinate,” or lead to perfect uniformity across individuals. Similarly, the idea that evolution has equipped the brain with specialized processes for acquiring the predispositions commonly referred to as “morality” does not entail that the social world is irrelevant. Quite the reverse. Just as it makes no sense to ask whether language is innate or acquired, it makes no sense to try to force on readers a choice between morality as innate or acquired.

The supposed distinction that BKH attempt to resuscitate between learned and innate has intuitive appeal. This might explain why it persists in some fields. But it was long ago abandoned in the core fields of biology and psychology. It has no place in contemporary discussions.

II. VARIATION

BKH claim that there is sufficient variety in views about justice as to falsify our claim to substantial agreement (about the three categories of crimes). Specifically, they refer to the “politically consequential fact [that] intuitions of justice are characterized by immense cultural heterogeneity.”

One cannot identify immensity simply by labeling it as such (a label we dispute). Regardless, BKH make two incorrect claims concerning variation. First, they ascribe to us the view that “evaluations of serious wrongfulness do not vary across social conditions.” Second, they

23 See id at 1642.
24 See, for example, Pinker, Language Instinct (cited in note 14).
26 See generally Ridley, Nature via Nurture (cited in note 9).
28 Id at 1551 (emphasis added).
claim that "[naturalism] assumes a lack of diversity in the core of wrongdoing . . .".  

We have clarified separately that the obvious diversity in intuitions of justice when looking across all criminal acts is irrelevant to our narrower claim that intuitions are generally shared with respect to an important subset of them. 30 But here our concern is the underlying conceptual one: what is the role of variation, where it exists, in arbitrating among candidate psychological theories for that subset?  

First, contrary to the BKH portrayal, variation is not, by itself, a problem for naturalist theories. Well-designed mechanisms—physiological and psychological—are expected to vary in systematic (and predictably patterned) ways as a function of particular features of the environment. That is simply the way evolution tailors different behavioral outcomes (for example, "look elsewhere") to different environmental circumstances (for example, "no food here"). It is a mistake to assume that biological processes necessarily result in an absence of variability and flexibility. 31  

Second, just as language can vary, despite evolved predispositions underlying its acquisition, intuitions of justice can vary, despite evolved predispositions underlying their acquisition. 32 Demonstrating some variation, alone, does not undermine the claim that there are morality-specific and morality-specialized computational mechanisms.  

Third, measuring diversity depends on how one counts. 33 For example, if one measured the sound used to designate "dog," one would find enormous cross-cultural variation. If one instead measures whether "dog" is separately lexicalized at all, distinct from other mammals, then one would see virtually no variation. The same is true for moral intuitions, 34 which develop reliably, as we have previously described. 35  

29 Id at 1592.  
32 This does not mean that our hypotheses regarding the causes of broadly shared intuitions of justice are unfalsifiable. No one has yet specified precisely how to measure variation in intuitions of justice; nonetheless, there are thresholds for lack of concordance that would be inconsistent with our hypotheses.  
33 For example, should the denominator be the number of social groups (which would count a small tribe and an entire country the same in the balance), or should it instead—seemingly much more usefully—be the total number of individuals across the planet?  
34 For example, the very notion of "wrongness" is quite universal. As has been shown, general social learning theories cannot explain this fact; there are no smaller, teachable "components" out of which wrongness can be constructed. See John Macnamara, Development of Moral Reasoning and the Foundations of Geometry, 21 J Theory Soc Behav 125, 143 (1991); George E. Moore, Principia Ethica 223 (Cambridge 1903).  
35 Robinson, Kurzban, and Jones, 60 Vand L Rev at 1664–75 (cited in note 2).
III. PUNISHMENT REALISM FAILS AS A SCIENTIFIC THEORY

The BKH endorsement of Punishment Realism is inconsistent with their professed commitment to empiricism. Because there are no observations that could be inconsistent with their theory, it is unscientific. BKH state that “Punishment Realism . . . holds that while people agree on many cases . . . they also frequently disagree about both whether an act is so wrong as to be criminal and, if it is, how serious the criminal offense is.”36 That is, Punishment Realism simply predicts some agreement and some disagreement—somewhere, sometimes—as a function of some (unspecified) things. No data could challenge one’s belief that this hypothesis is true.

Our theory, in contrast, makes testable predictions: (1) there will be high levels of agreement (in the three identified domains); and (2) these intuitions will be less malleable than average. If either of these is wrong, then we will be wrong.

IV. “GENERAL LEARNING MECHANISMS” DO NOT WORK

Punishment Realism invokes “social constructs,” “cultural priors,” and “cultural outlooks.”37 Socially, these terms are easily recognized. But scientifically, they are too underspecified to be useful in hypothesis formation. Even if one does try to make them do explanatory work, however, the notion that there are “generic cognitive mechanisms” for learning these various constructs is known to be wrong. General social learning theories have appeared with some regularity in psychology—from the behaviorists in the early twentieth century to connectionists in the latter half of it—and have always been found lacking.38 They just cannot work, as Chomsky has shown.39 In order for systems to change in useful ways (that is, learn), they must have exact theories about how to change in response to new information. Otherwise, “learning” would be random, and therefore useless.

We believe that the psychological mechanisms underlying these intuitions of justice are likely structured in ways that yield considerable homogeneity with respect to certain subsets of harms, specifically physical harms, thefts, and violations of social contracts. It is crucial to note that, even here, we not only believe that learning and development

36 Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1578 (cited in note 5).
37 See id at 1533, 1598, 1599.
38 See Tooby and Cosmides, Psychological Foundations of Culture at 100–08 (cited in note 17).
is important, but we said so in the article that serves as a main focus for BKH’s critique.\textsuperscript{40}

V. \textbf{WHERE THIS LEAVES US}

BKH have offered a frequently insightful, though we think often incorrect, discussion of how and why theories about intuitions of justice can matter. Our disagreements should not, however, obscure the many matters on which we and BKH agree. For example, we agree that reality matters. (That is, we are all empiricists.) We agree that cross-cultural studies can aid deeper understandings of punishment.\textsuperscript{41} We agree that insights about punishment should be reconciled across many relevant disciplines (including evolutionary biology).\textsuperscript{42} We agree that “moral judgments depend on numerous cognitive and physiological mechanisms that are presumably the product of evolutionary pressures.”\textsuperscript{43} We agree that parsimony is generally a virtue when considering alternative hypotheses. And we agree that sharp distinctions must be drawn between explanations and justifications.

Among the points on which we disagree is the best causal explanation for the shared intuitions of justice that exist in three distinct criminal arenas. BKH’s explanation emphasizes culture, excluding a meaningful role for evolutionary processes that underlie the modern mind. Our explanation emphasizes culture and also includes a meaningful role for evolutionary processes. We think our view is \textit{more} consistent with the evidence and more scientifically sound. To clarify, however, we reassert our actual conclusion in \textit{Origins}:

On present evidence, we believe that the explanation for the “puzzle” of the existence of shared intuitions of justice is more likely a specific evolved human mechanism for acquiring these core intuitions than general social learning derived from some set of conditions and life experiences universal to all humans and all human groups. The latter cannot be ruled out on present evidence, but it seems implausible, while the former is consistent with all available data.\textsuperscript{44}

Future work may help to resolve the causal question.

\textsuperscript{40} See Robinson, Kurzban, and Jones, 60 Vand L Rev 1633 (cited in note 2).
\textsuperscript{41} See, for example, id.
\textsuperscript{42} See, for example, Robinson and Kurzban, 91 Minn L Rev 1829 (cited in note 1).
\textsuperscript{43} Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1532 (cited in note 5). See Robinson, Kurzban, and Jones, 60 Vand L Rev at 1646–49 (cited in note 2).
\textsuperscript{44} Robinson, Kurzban, and Jones, 60 Vand L Rev at 1687 (cited in note 2).
CONCLUSION

Professors Braman, Kahan, and Hoffman offer a thoughtful critique of our evolutionary hypotheses that seeks to explain puzzlingly consistent intuitions of justice about three categories of core wrongs. The value of their critique is limited by the extent to which they have misunderstood key components of those biological and psychological perspectives. Nonetheless, because Professors Braman, Kahan, and Hoffman are not alone in misperceiving the bases of these perspectives, we are grateful for the opportunity they have provided to clarify—and to engage in further discussions about—where, how, and why people's intuitions of justice so powerfully converge.
Realism on Change in Moral Intuitions

John Darley†

There are several issues in the exchanges in Some Realism about Punishment Naturalism and Realism, Punishment, and Reform. The primary issue is the degree to which individuals' moral intuitions differ regarding what counts as a "crime," the moral magnitudes of different crimes, and what type and duration of punishment a given crime deserves. A closely linked issue is the degree of fixedness versus malleability in a person's judgments on these matters, and what processes produce whatever malleability exists. Caught up in these issues is an evolutionary psychological stance that is at least initially interpreted as suggesting universally shared immutable intuitions but seems to be converging on agreement, stated by Professors Donald Braman, Dan Kahan, and David Hoffman ("BKH") as follows: "[C]ognition is, to be sure, shaped by a host of demonstrable and perhaps nearly universal cognitive biases and heuristics, many or all of which are the product of evolutionary pressures or accidents. [We] view[] these innate cognitive traits as interacting with and generating a variety of social meanings that ultimately determine our understanding of and reaction to wrongdoing."³

The statement that I want to contest is the following:

We do feel deep concern, however, over what we take to be the politically conservative resonances with which the Punishment Naturalist has been needlessly infused. It is, simply put, extremely difficult to take in the corpus of work that the Punishment Naturalists have amassed without sensing a deep commitment on their part to the status quo—to popular retributive sensibilities as they are (or are depicted with a high degree of uniformity to be), and to laws that conform (or are depicted as conforming) to them.⁴

† Warren Professor of Psychology, Princeton University.
³ Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1567 (cited in note 1).
⁴ Id at 1602 (emphasis omitted).
Paul Robinson and I have suggested that people in a culture have well-developed intuitions about what constitutes a morally wrong action that requires punishment. Criminal codes that are broadly in agreement with those shared intuitions are seen as enacting justice and gain credibility as guides to moral behavior that citizens will be motivated to follow. Those seeking to change portions of the legal codes—often a morally appropriate enterprise—should seek to persuade citizens of the moral superiority of the changes proposed, rather than simply engage in elite efforts to rewrite the legal codes. The latter move risks delegitimizing legal codes if citizens perceive the novel codes as consistently and seriously at odds with their moral intuitions. Given the particular psychological character of the often-intuitive judgments that citizens form about wrongs, changing them is difficult, but possible in several ways.

I. PUNISHMENT JUDGMENTS ARE MADE BY DUAL PROCESSES

Researchers now recognize that people make decisions via many different processes. It is increasingly common to array these processes along a continuum that ranges from intuitive to reasoned processes. In this way, researchers distinguish between two broadly different ways that people come to decisions and judgments: one involves heuristic, intuitive processes, and the other involves reasoning processes.

A. Intuitions

What are the characteristics of intuitive punishment judgments? First, descriptions of crimes automatically and non-optionally trigger intuitive processes into action. Once triggered, they progress rapidly to their conclusions. Intuitive processes can proceed in parallel with other mental processes. This means that they can run "in the background" when a person's consciousness is directed elsewhere. They are implicit; that is, they are not available for introspective analysis and are frequently emotionally loaded. They are often the product of

---

5 See Paul H. Robinson and John M. Darley, The Utility of Desert, 91 Nw U L Rev 453, 456 (1997) (arguing that a legal code that "tracks the community's perceived principles of justice" leads to greater social compliance with the law).

what decision researchers call "heuristics."  

B. The Reasoning System

Decisions can also be the outputs of the reasoning system. But the reasoning system produces those decisions and judgments using very different processes. "Reasoning" is what a person does when she considers alternatives, thinks carefully about possible options, and applies problem-solving procedures. This is done with conscious monitoring, and thus the steps in the process can be directed, checked, and controlled.  

C. Punishment Judgments Are Often Intuitive

Many researchers suggest that desires to punish that come rapidly to mind when one person harms another are products of intuitive rather than reasoned processes. Jonathan Haidt's well-known demonstrations of "moral dumbfounding" show that when various disgusting but not obviously harmful actions are described, listeners generally respond quickly with a near instantaneous flash of negative affect and a confident judgment that it is wrong to take those actions. But when pressed by the experimenter for the reasons why they feel that the act is wrong, they generally cite harms that the sort of action described generally could cause. The experimenter has anticipated those harms, however, and constructed the story to rule them out. So the experimenter points out that the story makes the subject's suggested harm impossible, and again asks why the action is wrong. Finally, dumbfounded, the subjects continue to maintain the wrongness of the actions while admitting that they cannot at the moment give reasons for that judgment. The affective intuitive response is driving the judgment of wrongness. Reasoning processes were not engaged before making the judgment, but did come


7 Kahneman, 93 Am Econ Rev at 1460 (cited in note 6) (explaining that people rely on heuristics in decisionmaking, "which reduce the complex task of assessing probabilities and predicting values to simpler judgmental operations"), quoting Amos Tversky and Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, 185 Sci 1124, 1124 (1974).

8 Kahneman, 93 Am Econ Rev at 1450 (cited in note 6).


into action when the subject was asked to justify the decision.\(^\text{11}\) This is characteristic: people often act on their intuitions, without subjecting those intuitions to reasoned scrutiny.

D. Punishment Intuitions Are Retributive

Studies suggest that the goal of intuitive punishment judgments is retribution.\(^\text{12}\) In these studies, one group reads a short description of a crime and assigns a prison sentence to that crime. Another group reads the same core story, but with one element varied. For one pair of groups, the element that is varied is important if the readers are punishing on the basis of just deserts (retributitional considerations). In another pair of groups, the varied element matters if, for instance, the readers are sentencing for incapacitive reasons.\(^\text{13}\) To summarize the results, variations in the scenarios' just deserts elements caused corresponding variations in the durations of sentences; incapacitation and other goal variations made little or no difference.\(^\text{14}\) The suggestion here is that, for people in our society, punishment is generally driven by retributive concerns.

E. The Phenomenology of the Retributive Impulse

Dale Miller, writing about justice, realized that the unifier of the various arenas in which people speak of justice is the common feeling of injustice that is generated when justice principles are flouted. His summary of this response captures the intuitive reaction that transgressions produce:

The arousal of moralistic anger is not confined to injustices perpetrated against one's self. Witnessing the harming of a third party can also arouse strong feelings of anger and injustice. Even so-called "victimless" crimes, such as prostitution or pornography, can arouse strong moralistic and punitive impulses. These "disinterested" feelings of injustice are not necessarily identical to those that arise in response to a direct offense against one's self, but they also depend greatly on the perception of disrespect. Individuals are committed to the "ought forces" of their moral community, as

\(^\text{11}\) See id at 817 ("The central claim of the social intuitionist model is that moral judgment is caused by quick moral intuitions and is followed (when needed) by slow, ex post facto moral reasoning.").


\(^\text{13}\) See Darley, Carlsmith, and Robinson, 24 L & Hum Behav at 662 (cited in note 12).

\(^\text{14}\) See id at 671.
Heider termed them, and people believe that these forces deserve respect from all members of the community. The violation of these forces represents an insult to the integrity of the community and provokes both moralistic anger and the urge to punish the offender in its members. Viewed from this perspective, disinterested justice reactions are not disinterested at all, because everyone has a stake in seeing that the rules and values of the authority structure under which they live are respected.15

The notion of the "ought force," an externally acting force that requires punishment of offenders, bears family resemblance to the concept of "natural law." The fact that retributinal decisions are arrived at intuitively, without any sense of conscious cognition having been employed, adds to the sense that the punishment intuition is an objective and external demand, rather than a potentially fallible judgment of the individual.

People become aware that others differ on the morality of, for instance, sodomy, first trimester abortion, and mercy killings. Nevertheless, that generally does not detract from their belief that their moral convictions are the correct ones. People are usually "naïve realists" about their moral beliefs.16 Those who disagree are at best uninformed, more probably under some bad influence.17

II. HOW PUNISHMENT JUDGMENTS CAN DIFFER WITHIN AND BETWEEN CULTURES

To summarize, rapid intuitions are automatically generated in response to instances of "criminal" actions and often drive a person's

16 The case for naïve realism in opinion disagreement has been most extensively made in Richard Nisbett and Lee Ross, Human Interference: Strategies and Shortcomings of Social Judgment 196-227 (Prentice Hall 1980). Emily Pronin has demonstrated that the years psychologists have spent teaching that decisions are often made by heuristics and therefore possibly biased and wrong have had an influence—each person believes that others are swayed by biases, while he himself is much less so. Emily Pronin, How We See Ourselves and How We See Others, 320 Sci 1177, 1178 (2008).
17 BKH make this point better than I do:

[Participants may recognize that the moral hierarchies of others vary, but they are unlikely to prize other people's mores and commitments more highly than their own; at best they may view other value structures as strange or foreign, at worst as false and debased. And while those involved in such moral disputes may understand that their preferred outcomes derive from their values, they will often have trouble articulating the source of their values. Their values will seem, at least to them, to be natural.

Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1567 (cited in note 1). I therefore converge with BKH on the recognition that actors are phenomenological naturalists about many of their moral intuitions.
judgment that punishment is required. Were we to stop there, it would suggest that judgments about wrongs are likely to be shared at least within a society and that they are retributive in character. One understands why BKH comment on “a deep commitment . . . to the status quo.” But they misplace who has the commitment. We suggest that it is the commitment of the citizens to the general moral correctness of the system linking moral wrongdoing to criminal punishment, not our own.

Happily, further evidence suggests that the story sketched above is not the entire story. We now turn to how punishment judgments can vary, within the same individual at different times, among different individuals in the same culture, and between different cultures. These processes create possibilities of differing case judgments among citizens.

A. The Reasoning System Overrides

First, although these immediate, affect-laden intuitions often guide judgments, they do not always do so. The reasoning system can overrule the guidance of the intuitive system. A set of studies by Joshua Greene and his colleagues demonstrates this principle. Briefly, subjects undergoing fMRI brain imaging make a series of judgments about whether certain activities are morally acceptable. Shooting a misbehaving young child, for instance, is rapidly judged unacceptable. In one study, the researchers found that a set of problems titled “personal moral dilemmas” activated brain regions that previous research had associated with both emotion and social cognition activities.

These personal violation cases generally drew quick decisions: if the action in question was judged wrong, fMRI patterns showed heightened brain activity in the emotion and social cognition areas and a near-instant reaction of “don’t do it.”

Certain other decision cases brought very different reaction patterns. One such case was the “crying baby” case. In a group of people, a baby begins to cry. Is it acceptable to smother the baby? Other elements of the story horribly raised the stakes. The group was hiding from vicious soldiers, who would kill the entire group, including the baby, if the group were discovered. The baby’s cries would lead the

18 Id at 1602 (complaining that this outlook leads to needless paralysis among reformers).
20 See Greene, et al, 44 Neuron at 390 (cited in note 19) (indicating that the long reaction times for such dilemmas supported an inference that “cognitive control” was engaged).
21 See id at 393–94.
soldiers to the group. This case provoked the usual, rapid emotional responses, presumably produced intuitively by one’s repugnance at inflicting lethal harm to the baby. But all would be killed, including the baby, if action were not taken. This engaged slower reasoning processes in areas associated with higher-order reasoning and conflict management in decisionmaking.

Interestingly, different subjects reasoned to opposite conclusions—some that it still was impermissible to smother the baby, others that it was permissible, perhaps obligatory to do so. Choices between two bad courses of conduct, then, can produce important disagreements on punishment judgments.

According to this account, dual processes can contribute to moral judgments. One process, produced rapidly, takes place non-optionally. This is the intuitive system discussed above. The second set of processes involves abstract reasoning areas of the brain, and is not always triggered into action. Furthermore, when this reasoning system is activated, it sometimes overrides intuitions.

It is this possibility that creates one mechanism through which intuition-overriding punishment reactions can be produced in a person who originally followed intuitions. Other punishment decisions may also involve reasons overruling intuitions. Some restorative justice scholars argue that the concept of punishment is “barbaric,” and societies must give up that impulse to substitute restorative practices. Several scholars have implied that general deterrence is the only morally possible justification of punishment. They too reject retributive impositions.

B. Reasoned Decisions Become Intuitive

A second possibility exists here. Cognitive theorists suggest that intuitions can be changed. Intuitions, like perceptions, are the results of information processing, although the processing is rapid and not

22 See id at 390 (explaining how the hypothetical exposes tension between an emotional aversion to killing a baby and a rationalist desire to minimize the overall number of deaths).
23 See id at 393–94.
26 See, for example, Daniel W. Van Ness, New Wine and Old Wineskins: Four Challenges of Restorative Justice, 4 Crim L F 251, 257–60 (1993) (arguing that the modern criminal justice system is overly concerned with “public order” to the detriment of other communal values).
27 See, for example, Daniel M. Farrell, The Justification of General Deterrence, 94 Phil Rev 367, 368–69 (1985).
accessible to consciousness. So it is possible for an individual to change her information processing, and a related area of research has demonstrated that this can happen. Racial or gender stereotypes generally are produced intuitively in response to cues. But people committed to an egalitarian stance have managed to overcome the intuitive processes that conjure up stereotypes and instead activate egalitarian goals.

C. The Role of the Group

It is likely that a person's intuitive judgments, whether about punishment or other issues, are best sustained if they exist within a community that shares the relevant intuitions. It is a general observation of social science that people form communities of shared opinions, and that this may allow them to preserve their intuitions about certain issues, particularly political issues, because their intuitions are shared rather than challenged. Some groups may form for this sort of purpose—to support members retaining deviant intuitions. One recent tragic event hints at this. In Lancaster County, Pennsylvania in 2006, a neighbor entered a small Amish schoolhouse and killed five young girls before shooting himself. Members of the Amish community quickly took actions to signal forgiveness to the family of the killer. The author of an article about the incident has studied the Amish attitude toward forgiveness, and points out that it is ingrained. Their religious tradition "predisposes them to forgive even before an injustice occurs."

This quote contains an interesting suggestion: that the community has managed to so internalize the response of forgiveness that it has transformed it from a reasoned override of a retributive intuition to the intuitive response to at least some moral wrongs. It is likely that this kind of transformation occurs most easily within a community that rehearses its alternative account in its religious practices and daily belief enactments.

D. Persistent Differences within a Society

One of the most visible differences in punishment judgments exists between liberals and conservatives. To some extent, the two groups'
abilities to sustain their differing intuitions depend on their having circles of like-minded others with whom they validate their judgments.

Several other processes may be implicated as well. There are often strong emotional components to intuitive reactions to possible cases of wrongdoing, and they can cue and amplify moral reactions in the case readers. Reactions of disgust are known to have this propensity. Thalia Wheatly and Jonathan Haidt hypnotized subjects to react with disgust when they read a normally neutral word. The word was inserted into written narratives describing actions that most readers judged to be somewhat wrong. Those hypnotized to produce disgust elevated their ratings of wrongness for the stories in which a word cued disgust.

Conservatives may regard actions that disgust them as actions that are wrong and therefore appropriate candidates for punishment. Liberals, on the other hand, may react with disgust to events such as sodomy because of the social milieu in which the perceiver was socialized, but, reasoning that the actions in question are consensual and do not cause harm, judge the actions as allowable.

E. Mental Representations

There are other mental mechanisms that produce predictable disagreement about the relative morality of commonly contested cases. To have an intuitive reaction to an event requires first forming some mental representation of that event based on the information inputs, even though the rapidity of its formation obscures this fact. Differential identifications with the various actors in crime scenarios can lead to differential representations of the events, which can favor

(2003) (identifying several psychological characteristics, such as fear of loss and intolerance of ambiguity, that describe conservatives across all cultures).

33 Thalia Wheatley and Jonathan Haidt, Hypnotic Disgust Makes Moral Judgments More Severe, 16 Psych Sci 780, 780 (2005) ("Half of the groups were instructed to feel disgust when reading the word often; half were instructed to feel disgust when reading the word take.").

34 See id at 781.

35 In this light, consider the Supreme Court's decision striking down a Texas sodomy statute, with Justice Antonin Scalia in dissent. See Lawrence v Texas, 539 US 558, 599 (2003) (Scalia dissenting).

mitigating or aggravating perspectives on the moral meanings of the actions described.

This is often done casuistically. People are often motivated to form mental representations that lead to moral judgments favorable to those in the story with whom they identify. Psychologists have characterized "the intuitive prosecutor" as one who is looking for blaming interpretations."

F. Societal Effects on Moral Codes

Evolutionary and social learning perspectives converge on the suggestion that the society within which one is raised will have strong shaping effects on moral perspectives. Researchers have recently explored prosocial morality behavior across fifteen cultures. They find first that the amount of sharing of goods, and the degree of punishment of those who fail to share, co-varies in the fifteen cultures. Further, the degree of sharing also reflects certain cultural norms and the presence of certain cultural institutions.

Market exchanges, the researchers argue, build "trust, fairness, and cooperation. This lowers transaction costs, raises the frequency of successful transactions, and increases long-term rewards." Results showed that the more the members of the community obtained their daily caloric intake via market exchanges, the more sharing took place between individuals in the researchers' games. Further, cultures that had world religions in place that were likely to emphasize moralistic behavior also evidenced increased sharing behaviors. Importantly, moral norms vary as a function of a culture's moral-socialization practices and the underlying structure of the moral norms into which people are being socialized.

As this indicates, societies transmit moral rules to their members and thus the moral rule set can and does differ among cultures. A familiar example of this concerns whether moral rules against theft and murder are universal, or extend only to the ingroup such that people in other societies, clans, or tribes are "fair game." This indicates one quite central way in which the moral rules in force can be drastically

37 See, for example, Julie H. Goldberg, Jennifer S. Lerner, and Philip E. Tetlock, Rage and Reason: The Psychology of the Intuitive Prosecutor, 29 Eur J Soc Psych 781, 783 (1999) (claiming that "intuitive prosecutors" are created when observers learn that "justice was not served after an anger-eliciting event").


39 See id at 1483–84.

40 Id at 1480.

41 See id at 1483.

altered: denying that the other actors are within the community to which moral rules apply.

CONCLUSION

The folk theory that people hold about punishment is a naturalistic one, in the sense that people intuitively feel that the core prohibitions against physical harms, unauthorized takings, and deception in exchanges have a moral rightness that stems from forces that exist beyond the mere agreements of interacting individuals. Further, given that shared understandings are transmitted through the socialization practices of the culture, many, if not all, judgments are likely to be jointly held by the culture’s members. The sense of taken-for-granted certainty created by this uniformity will be strong, so broad and stable changes in moral rules will be difficult to produce.

Moral rules can be changed, for instance, by convincing people of the moral appropriateness of change, thus equipping them with reasons that will override their intuitions and might eventually convert their intuitive responses to accord with their reasoned decisions. In centuries past, prisoners were put to death, often in horrible ways, for offenses we would now regard as minor or no offense at all, such as heresy. We have moved far from those practices. We now find them intuitively repugnant. Further change is possible—but difficult.

43 David Garland offers a persuasive account of the transformations in our thinking that brought about these changes in punitive practices. See David Garland, Punishment and Modern Society: A Study in Social Theory 287 (Chicago 1990).

44 See Robinson and Darley, 91 Nw U L Rev at 471–88 (cited in note 5).
We are deeply gratified by this exchange with Professors John Darley, Paul Robinson, Owen Jones, and Robert Kurzban. We have benefited a great deal from their research, and this encounter only adds to our appreciation. Their work has always been exceptional in its devotion to empirical exploration and experimentation. We are grateful to them for taking the time to share their thoughts with us and with the readers of this journal. In responding, we are unsurprised to find that we are in agreement with quite a bit of what they have to say.

Indeed, there is very little that we can find in the nuanced and learned account that John Darley individually presents that is inconsistent with our conception of Punishment Realism. As we understand him, he also rejects most of what we found most objectionable in the accounts of Punishment Naturalism that we criticize. What one perceives to be right or wrong—and precisely how right or wrong one perceives it to be—will depend in large part on socialization, which can vary culturally. Through this socialization process, individuals develop very speedy moral evaluations that are consistent with norms in their culture. These rapid intuitions can, as he notes, be countered through conscious reflection and reasoning, but this type of critical reflection is difficult to prompt, and our intuitions can be quite difficult to revise. The cognitive mechanisms on which people draw to make moral assessments are highly uniform across individuals; but the content of those assessments varies across groups and within them over time as a result of local social influences. Darley’s is an account that we embrace as entirely consistent with Punishment Realism as we

---

† Associate Professor of Law, The George Washington University Law School.
†† Elizabeth K. Dollard Professor of Law, Yale Law School.
‡ Associate Professor of Law, Temple University Beasley School of Law.
1 John Darley, Realism on Change in Moral Intuitions, 77 U Chi L Rev 1643, 1652 (2010).
2 Id at 1643 (quoting our original article with approval on this point).
3 Id at 1644–45.
4 See, for example, id at 1652.
describe it—indeed, it is an admirably sophisticated and clear conception that improves our own understanding.

We agree, too, with Paul Robinson, Owen Jones, and Robert Kurzban ("RJK") in their insistence that the empirical evidence reveals "not just disagreements about relative blameworthiness, but also about whether the conduct should even be criminal." And we are heartened that, in their separate article, Jones and Kurzban share our "opposition to genetic determinism, [our] commitment to plasticity in human cognition, and a deep (in fact scientifically unavoidable) commitment to recognizing the crucial role that social environment plays in each individual's development of intuitions of justice."5

But the core of our agreement with them has a clear and definite periphery. As RJK now clearly explain, they used a method carefully designed to exclude from measurement any "aspects" of "core" offenses on which there is demographic or cultural disagreement.6 Accordingly, we simply disagree with them when they assert that their work has important implications for criminal law reformers.

Here is what they say:

What is the "core"? [Braman, Kahan, and Hoffman ("BKH")]
suggest that its contours are quite vague and difficult to identify,
but what constitutes the "core" is not a matter of speculation or
tory, or even of interpretation. It is a matter of empirics. The
"core" is, by definition, that on which there is high agreement
across demographics, like that demonstrated in the C&C agreement
study.

What cases are included in the core? Those cases on which there
is high agreement across demographics.

What aspects of these offenses are included in the core? BKH
seem to assume that we claim that all aspects, all cases, involving
any of these offenses are part of the core, but this could hardly
be the situation. Our research used factors upon which we
judged there was high agreement. To the extent that one substi-
tutes a factor on which there is disagreement, obviously the level
of agreement on the relative seriousness of the case would have
to decline.

The point of C&C's Appendix B is to show the reader just how we were able to construct the twenty-four scenarios on which our subjects had such high agreement: by relying upon, and only upon, principles that we knew were deeply embedded intuitions of near unanimity. . . .

As you can imagine, we found the BKH article quite difficult to understand, given its false assumptions about our claims. For example, it has an entire section showing disagreements in cases of deception in exchanges. Whether somebody is deceived in an exchange obviously is a function of one’s expectations about the terms of the exchange, and those expectations could be highly culturally dependent or, even within a culture, highly dependent on context. The case we used in the study was one of a store clerk shortchanging a customer. We used it precisely because it seemed to us that such shortchanging offered an example of a violation of a nearly universal expectation of this most common form of exchange, a purchase.8

As we emphasized in our article—and as RJK now say is “obviously” correct—people of diverse identities (within and across societies) are intensely divided about whether certain conspicuous, recurring forms of behavior count as instances of the offenses that punish core criminal wrongdoing.9 In the United States, for example, there are intense cultural divisions on whether battered women who kill their husbands in their sleep, or “true men” who stand their ground and kill attackers when they could easily flee, are murderers,10 whether male college students (and others) who persist in engaging in intercourse with a woman who repeatedly and emphatically objects are rapists,11 and whether squatters have property rights12 or digital versions of songs can be shared among friends.13

8 Id at 1616–18.
9 Id at 1621. RJK pose various empirical “challenges” to us, see id at 1622–23, but those have no connection to the only empirical point we are making: there is significant cultural variation, and resulting political conflict, over what count as instances of offense types included in the RJK core. We stand by the evidence in our article on that. See Donald Braman, Dan M. Kahan, and David A. Hoffman, Some Realism about Punishment Naturalism, 77 U Chi L Rev 1531, 1566–92 (2010).
10 See Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1581–92 (cited in note 9).
11 Id at 1574–75.
RJK unconvincingly try to deflect this argument by suggesting that the work we rely on shows the influence of cultural variation only in “justificatory norms” as opposed to “prohibitory” ones. But as the work of Mark Alicke has shown, people tend to conform their perceptions of the various components of culpable behavior—such as volition, action, causation, and harm—to social norms extrinsic to those concepts. It follows that people with different norms, even when they agree about what conduct is morally blameworthy (or otherwise worthy of “prohibition”) generally, will systematically disagree about what counts as an instance of that conduct. Our work on cultural cognition seeks to identify the particular norms that make the most conspicuous contribution to this form of motivated perception and hence to the highly politicized disputes we see in law and society generally over who should be blamed for wrongdoing and when.

RJK’s “core” definitely measures something on which diverse people agree. But because their methods deliberately exclude from the specification of “core” offense types precisely those “aspects” of them that provoke cultural dispute about what counts as murder, rape, and fraud, the construct they measure cannot predict or explain who sees what as wrong (indeed, criminally wrong) and why in the real world.

For the same reason, what they are measuring when they find a “core” of agreement has no normative or prescriptive consequence. Whether the fact of “a high level of agreement” is treated as evidence of an act’s wrongfulness or simply recognized as a political constraint

---

14 Robinson, Jones, and Kurzban, 77 U Chi L Rev at 1620 (cited in note 5).
on the possibility of reform,\(^{17}\) the fact remains that the kind of “agreement” RJK measure lacks sufficient connection to live controversies to matter in either of these ways. The admonition that one should not undertake reform in any area where there is “consensus” (regardless of whether it involves murder, rape, torture, theft, fraud, or anything else) is simply beside the point, because there is not consensus on the sorts of issues that are at the practical core of efforts to evaluate and reform criminal law in American society.\(^{18}\)

We agree, in short, that RJK are talking about something other than what we and many other academic and political commentators are talking about. The whole point of our article was to make this unmistakably clear, lest anyone think that Punishment Naturalism supplies a reason either to doubt the reality of profound political conflict over the content of the criminal law in our society or to resist particular positions about how that conflict should be resolved. We are glad that RJK acknowledge this point.

Still, in response to their bafflement about why it would even seem necessary for us to make it, we note that the RJK response itself risks perpetuating the sort of overreading of their work that we warned against. To rebut the charge of conservatism, RJK insist that their “program is designed to give reformers tools for more effective reform.”\(^{19}\) These “tools” consist of pieces of advice such as “it may often be unwise to invest limited reform resources on trying to change intuitions of justice that will be difficult to change,”\(^{20}\) and “when developing a program to change people’s intuitions of justice, it will often be a better investment to harness people’s core intuitions of justice rather than fight them.”\(^{21}\) But to whom exactly are they addressing this prudential counsel? Presumably it cannot be anyone, for example, who is currently proposing reforms relating to “aspects”\(^{22}\) of murder, rape, and theft on which there is cultural dissensus, for RJK insist there is nothing in their research that speaks to such issues. Yet, in fact, they proceed to draw a “conclusion” from their work for those who want to reform rape law to combat the contested norm that “no means yes”: avoid a “strict liability” standard lest the conflict between law and “internalized norms” cause “defendants . . . [to] be seen as blameless” and vitiate the “moral credibility” of law generally.\(^{23}\) Can they really be surprised if readers

---

\(^{17}\) Robinson, Jones, and Kurzban, 77 U Chi L Rev at 1613 (cited in note 5).

\(^{18}\) See Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1534–36 (cited in note 9).

\(^{19}\) Robinson, Jones, and Kurzban, 77 U Chi L Rev at 1628 (cited in note 5).

\(^{20}\) Id.

\(^{21}\) Id.

\(^{22}\) Id at 1617.

\(^{23}\) Robinson, Jones, and Kurzban, 77 U Chi L Rev at 1629 (cited in note 5).
(and not just us, as we pointed out in our article\textsuperscript{24}) see their work as evincing resistance to reform on culturally disputed issues when they themselves read their work that way?

But we will not dwell on the possibilities for misunderstanding that persist. Instead, let us offer our own advice to would-be reformers of the criminal law.

First and foremost, contemporary debates in criminal law are characterized by dissensus over what deserves—and what should comprise—punishment. If you are involved in such a debate over torture, rape, self-defense, intellectual property, eminent domain, consumer fraud protection, or any other contentious legal issue, if you are an advocate for reform, or if you feel the law is unjust, you should be utterly undissuaded from attempting to reform the law by any notion that the content of the current law reflects a \textit{universal} or \textit{innate} intuition about justice.

Second, any attempt at legal reform is likely to be quite difficult and culturally fraught. The difficulty, however, has little to do with an innate "moral organ,"\textsuperscript{25} and everything to do with the cultural significance that those on both sides of the debate invest in the law. Recognition of the unavoidable connection between the law’s position in such conflicts and the status of contested visions of the good life should make you circumspect about the prospects of reform. It should also make you anxious when assessing the proper scope for norm shaping in a liberal society and intent on discovering means for avoiding cultural domination and accommodating difference.

Third, and finally, nothing about the innate structure of our minds will absolve you of the hard work of determining what should populate the categories of offenses that we condemn and punish, or assessing what the law will convey about the status of the communities to which it speaks. Intuition is often a poor guide for understanding the motivation and reasoning of those who oppose the social reordering you desire. Understanding and overcoming opposition in culturally contested battles over the law is profoundly difficult work, but it is also deeply important work.

\textsuperscript{24} See Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1602–04 (cited in note 9).

\textsuperscript{25} See id at 1539–40.