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.

¹ Donald Braman, Dan M. Kahan, and David A. Hoffman, *Some Realism about Punishment Naturalism*, 77 U Chi L Rev 1531 (2010). See also Donald Braman, Dan M. Kahan, and David A. Hoffman, *A Core of Agreement*, 77 U Chi L Rev 1655 (2010).

² Paul H. Robinson, Owen D. Jones, and Robert Kurzban, *Realism, Punishment, and Reform*, 77 U Chi L Rev 1611 (2010). See also Owen D. Jones and Robert Kurzban, *Intuitions of Punishment*, 77 U Chi L Rev 1633 (2010).

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).

⁶ See Daniel Kahneman, Maps of Bounded Rationality: Psychology for Behavioral Economics, 93 Am Econ Rev 1449, 1451 (2003) (describing the intuitive process as "governed by habit, and . . . therefore difficult to control"); Steven A. Sloman, Two Systems of Reasoning, in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds, Heuristics and Biases: The Psychology of Intuitive Judgment 379, 391 (Cambridge 2002); Keith E. Stanovich and Richard F. West, Individual Differences in Reasoning: Implications for the Rationality Debate?, in Gilovich, Griffin,

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.⁸

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. 10 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

and Kahneman, eds, *Heuristics and Biases* 421, 436; Steven A. Sloman, *The Empirical Case for Two Systems of Reasoning*, 119 Psych Bull 3, 6 (1996).

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).

⁹ See, for example, Jonathan Haidt and Fredrik Bjorklund, *Social Intuitionists Answer Six Questions about Moral Psychology*, in Walter Sinnott-Armstrong, ed, 2 *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity* 181, 188–89 (MIT 2008). See also Fiery Cushman, Liane Young, and Marc Hauser, *The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm*, 17 Psych Sci 1082, 1087 (2006).

¹⁰ See Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 Psych Rev 814, 814 (2001).

into action when the subject was asked to justify the decision. 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. 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 (retributional considerations). In another pair of groups, the varied element matters if, for instance, the readers are sentencing for incapacitive reasons. 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. 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

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.").

See, for example, Kevin M. Carlsmith, John M. Darley, and Paul H. Robinson, Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment, 83 J Personality & Soc Psych 284, 295 (2002); John M. Darley, Kevin M. Carlsmith, and Paul H. Robinson, Incapacitation and Just Deserts as Motives for Punishment, 24 L & Hum Behav 659, 676 (2000).

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

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 retributional 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. Those who disagree are at best uninformed, more probably under some bad influence.

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

[P]articipants 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.

¹⁵ Dale T. Miller, *Disrespect and the Experience of Injustice*, 52 Ann Rev Psych 527, 535 (2001) (citations omitted).

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:

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).

¹⁹ See Joshua D. Greene, et al, *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*, 44 Neuron 389, 397 (2004); Joshua D. Greene, et al, *An fMRI Investigation of Emotional Engagement in Moral Judgment*, 293 Sci 2105, 2107 (2001).

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.²² 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.²³

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.²⁴ 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.

²⁴ See Greene, et al, 44 Neuron at 390 (cited in note 19).

²⁵ For a more precise description of the harmful actions to which the intuitive system is responding, see Joshua D. Greene, et al, *Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment*, 111 Cognition 364, 365 (2009).

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.²⁸ 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.²⁹

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.³⁰ 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.³² To some extent, the two groups'

²⁸ See notes 6–7 and accompanying text.

²⁹ A recent review summarizes the processes involved in overcoming stereotype intuitions. See Gordon B. Moskowitz, *On the Control over Stereotype Activation and Stereotype Inhibition*, 4 Soc & Personality Psych Compass 140, 141 (2010).

³⁰ Donald B. Kraybill, Why the Amish Forgive So Quickly, Christian Sci Monitor 9 (Oct 2, 2007).

³¹ Id.

³² For an account of these differences and a theory of how they come about, see generally John T. Jost, et al, *Political Conservatism as Motivated Social Cognition*, 129 Psych Bull 339

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. The word was inserted into written narratives describing actions that most readers in the word cue 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.³⁵

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, 36 which can favor

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

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).

See P.N. Johnson-Laird, *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness* 10 (Harvard 1983) (describing how humans "understand the world by constructing working models of it in their minds" and thus different people can have different "models" of the same reality). One of the more famous studies in psychology shows how, seeing a film of a penalty-ridden football game, partisans of each side perceived that the other side "started it." See Albert H. Hastorf and Hadley Cantril, *They Saw a Game: A Case Study*, 49 J Abnorm & Soc Psych 129, 133 (1954). Consider also 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, 841–42 (2009).

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.³⁷

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").

³⁸ Joseph Henrich, et al, Markets, Religion, Community Size, and the Evolution of Fairness and Punishment, 327 Sci 1480, 1481 (2010).

³⁹ See id at 1483–84.

⁴⁰ Id at 1480.

⁴¹ See id at 1483.

⁴² See Henrich, et al, 327 Sci at 1482 (cited in note 38).

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. 44

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).