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.[1]

In *The Origins of Shared Intuitions of Justice*[2] we built off of prior work[3] 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[4]). Their critique includes a vehement disagreement with our preferred hypothesis,[5] 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).

2 Paul H. Robinson, Robert Kurzban, and Owen D. Jones, *The Origins of Shared Intuitions of Justice*, 60 Vand L Rev 1633 (2007).

3 See generally, for example, Owen D. Jones and Timothy Goldsmith, *Law and Behavioral Biology*, 105 Colum L Rev 405 (2005); Owen D. Jones, *Time-Shifted Rationality and the Law of Law's Leverage: Behavioral Economics Meets Behavioral Biology*, 95 Nw U L Rev 1141 (2001); Robert Kurzban and Mark R. Leary, *Evolutionary Origins of Stigmatization: The Functions of Social Exclusion*, 127 Psych Bull 187 (2001). See also, for example, Robert Kurzban and Steven Neuberg, *Managing Ingroup and Outgroup Relationships*, in David M. Buss, ed, *Handbook of Evolutionary Psychology* 653, 659–61 (John Wiley 2005). This builds off of work by many schoolars. See, for example, sources compiled on the website of the Society for Evolutionary Analysis in Law (SEAL), online at http://www.sealsite.org (visited Aug 30, 2010).

4 See generally Paul H. Robinson, Owen D. Jones, and Robert Kurzban, *Realism, Punishment, and Reform*, 77 U Chi L Rev 1611 (2010).

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."[6] We do not do so. BKH repeatedly frame our work as denying plasticity in human cognitive and moral development.[7] We do not do so. BKH assert that our hypothesis denies roles to culture and social environment.[8] 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[9]) 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,[10] 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

---

lines, see generally Christopher Jaeger, *Defending a Social Learning Explanation: A Comment on the Origins of Shared Intuitions of Justice*, 62 Vand L Rev En Banc 25 (2009), online at http://law.vanderbilt.edu/publications/vanderbilt-law-review/online-companion/download.aspx?id=3993 (visited Aug 30, 2010) (arguing that "general social learning" is the most plausible explanation for our shared intuitions about justice).

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.

10 See Robert Kurzban, *Alas Poor Evolutionary Psychology: Unfairly Accused, Unjustly Condemned*, 2 Hum Nat Rev 99 (2002), online at http://human-nature.com/nibbs/02/apd.html (visited May 17, 2010); Robert Kurzban, *Grand Challenges of Evolutionary Psychology*, 1 Frontiers Evol Psych 1 (2010).

> 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.[11]

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.[12] Because many mechanisms' functions can only be executed by acquiring information from others,[13] theories about functional mechanisms *necessarily* entail a commitment to the belief that key parts of psychological development depend on social environments.[14] 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.[15] (We use "innate" too, but quite differently.[16]) 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 Braman, 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).

15 Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1538–39 (cited in note 5). See also H. Clark Barrett and Robert Kurzban, *Modularity in Cognition*, 113 Psych Rev 628, 637–38 (2006); Daniel Sperber, *Modularity and Relevance: How Can a Massively Modular Mind Be Flexible and Context-Sensitive?*, in Peter Carruthers, Stephen Laurence, and Stephen Stich, eds, *The Innate Mind: Structure and Content* 53, 57–59 (Oxford 2005).

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.[17] 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.[18] Yet researchers and theorists in biology, psychology, and neighboring fields long ago rejected this dichotomy as false, both conceptually and empirically.[19] 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 . . . ?"[20] 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[21] that morality is *both* "not determined" and "innate."[22] 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*.

---

17 See Douglas Futuyma, *Evolution* (Sinauer 2005); John Tooby, Leda Cosmides, and H. Clark Barrett, *The Second Law of Thermodynamics Is the First Law of Psychology: Evolutionary Developmental Psychology and the Theory of Tandem, Coordinated Inheritances: Comment on Lickliter and Honeycutt (2003)*, 129 Psych Bull 858, 863–64 (2003); Timothy Goldsmith and William Zimmerman, *Biology, Evolution, and Human Nature* (John Wiley 2001); John Tooby and Leda Cosmides, *The Psychological Foundations of Culture*, in Jerome H. Barkow, Leda Cosmides, and John Tooby, eds, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* 19, 83–84 (Oxford 1992). See also Jones and Goldsmith, 105 Colum L Rev at 428 (cited in note 3) ("Behavior flows from brains that encounter specific environmental stimuli and possess a neural architecture that is as importantly shaped by environments as it is by genes.").

18 See Steven Pinker, *The Blank Slate: The Modern Denial of Human Nature* (Viking 2002).

19 For an overview, see generally Ridley, *Nature via Nurture* (cited in note 9). See also Owen D. Jones, *Sex, Culture, and the Biology of Rape: Toward Explanation and Prevention*, 87 Cal L Rev 827, 874–77 (1999).

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

21 See, for example, Alan Slater and Gavin Bremner, 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.[23] Most researchers believe that evidence supports a functionally specialized process in the brain—a "Language Acquisition Device."[24] 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).[25]

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.[26] 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*."[27]

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."[28] Second, they

---

23 See id at 1642.

24 See, for example, Pinker, *Language Instinct* (cited in note 14).

25 See Noam Chomsky, *Aspects of the Theory of Syntax* (MIT 1965); Pinker, *Language Instinct* (cited in note 14).

26 See generally Ridley, *Nature via Nurture* (cited in note 9).

27 Braman, Kahan, and Hoffman, 77 U Chi L Rev at 1604 (cited in note 5) (emphasis altered).

28 Id at 1551 (emphasis added).

claim that "[naturalism] assumes a lack of diversity in the core of wrongdoing . . . ."[29]

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.

30 Robinson, Jones, and Kurzban, 77 U Chi L Rev at 1621–23 (cited in note 4).

31 See John Tooby and Leda Cosmides, *On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation*, 58 J Personality 17, 60–62 (1990).

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 ex ante 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).

39 See Chomsky, *Aspects of the Theory of Syntax* (cited in note 25); Tooby and Cosmides, *Psychological Foundations of Culture* 19 (cited in note 17).

is important, but we said so in the article that serves as a main focus for BKH's critique.[40]

V. 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.[41] We agree that insights about punishment should be reconciled across many relevant disciplines (including evolutionary biology).[42] We agree that "moral judgments depend on numerous cognitive and physiological mechanisms that are presumably the product of evolutionary pressures."[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 *more* consistent with the evidence and more scientifically sound. To clarify, however, we reassert our actual conclusion in *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.[44]

Future work may help to resolve the causal question.

---

40 See Robinson, Kurzban, and Jones, 60 Vand L Rev 1633 (cited in note 2).

41 See, for example, id.

42 See, for example, Robinson and Kurzban, 91 Minn L Rev 1829 (cited in note 1).

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).

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.