**Vanderbilt Law Review**

Volume 60 | Issue 6 Article 1

11-2007

# The Origins of Shared Intuitions of Justice

Owen D. Jones

Paul H. Robinson

Robert Kurzban

VANDERBILT LAW REVIEW

VOLUME 60 NOVEMBER 2007 NUMBER 6

# The Origins of Shared Intuitions of Justice

*Paul H. Robinson,* Robert Kurzban,** and Owen D. Jones****



* Colin S. Diver Professor of Law, University of Pennsylvania. The authors are indebted to Michael Orchowski, Alex Shaw, Peter DeScioli, Lindsay Suttenberg, Martha K. Presley, Shang Cao, and Susan Eisenberg for invaluable research assistance.

** Assistant Professor of Psychology, University of Pennsylvania.

*** Professor of Law & Professor of Biological Sciences, Vanderbilt University; Co-Director, Network on Decision Making, MacArthur Foundation Law and Neuroscience Project.

## I. THE PUZZLE: SHARED INTUITIONS OF JUSTICE

The role of justice in assigning criminal liability and punishment has been a matter of long-standing debate. A standard argument against a desert distributive principle—that is, against distributing punishment according to an offender's blameworthiness—

has been that such a concept of "desert" is simply too vague and the subject of too much disagreement to operationalize.[1] There may be some truth to these complaints when applied to a philosophical notion of desert, which has been the traditional basis of the desert school.[2]

But more recently, a utilitarian-based theory of desert has urged reliance on an empirical notion of desert, drawn from the community's shared intuitions of justice rather than from the reasoned concepts of moral philosophy.[3] The same objections about vagueness and lack of agreement are lodged against this empirical notion of desert. People's intuitions of justice, it is claimed, are simply too vague to be relied on and, in any case, there is too much disagreement about what constitutes deserved punishment to construct a workable criminal justice system. It is the common wisdom that little agreement exists among people's intuitions of justice:

> [E]ven assuming retribution in distribution is appropriate, there is a classic epistemological problem. How do we know how much censure, or "deserved punishment," a particular wrongdoer absolutely deserves? God may know, but as countless sentencing exercises have shown, peoples' intuitions about individual cases vary widely.[4]

> There is . . . reason to doubt that anything like a consensus exists on the seriousness of criminal conduct. While there may be some agreement on relative levels of harm, there appears to be great variation in perceptions of the absolute magnitude of harm represented by various criminal acts, and in either the relative or absolute level of culpability represented by various criminal actors.[5]

---

1. Paul H. Robinson, *Competing Conceptions of Modern Desert: Vengeful, Deontological, and Empirical* 67 CAMBRIDGE L.J. (forthcoming 2008) (on file with authors).

2. *Id.*

3. *Id.*; Paul H. Robinson & John M. Darley, *The Utility of Desert*, 91 NW. U. L. REV. 453 (1997).

4. Michael Tonry, *Obsolescence and Immanence in Penal Theory and Policy*, 105 COLUM. L. REV. 1233, 1263 (2005).

5. John Monahan, *The Case for Prediction in the Modified Desert Model of Criminal Sentencing*, 5 INT'L J.L. & PSYCHIATRY 103, 105 (1982). Other authors have made similar arguments:

> [Desert theorists argue that] we can work out one single, linear ordering of crimes, from least to most "serious." Yet that scarcely seems a credible assumption. Try, for instance, to rank the following crimes in order of their "seriousness": attempted residential burglary, trading stock on inside information, negligent vehicular homicide, bribing a mine-safety inspector, possessing an ounce of cocaine, and burning a cross on the lawn of black newcomers to a previously all-white neighborhood. To view this motley assortment along a single dimension of "seriousness" would seem no less difficult than to perceive the inner logic behind the apocryphal Chinese encyclopedist of Jorge Luis Borges's imagination.

David Dolinko, *Three Mistakes of Retributivism*, 39 UCLA L. REV. 1623, 1638-39 (1992). Similarly, Ernest van den Haag argues:

> [Desert theorists moreover appear] to believe that the comparative seriousness of crimes can be determined in all cases. Not so. Comparative seriousness can be determined only for some crimes, and it does not fully determine the comparative

In this respect, the common wisdom is simply wrong. Part of the confusion here arises from the failure to distinguish two distinct judgments: setting the endpoint of the punishment continuum and, once that endpoint has been set, ordinally ranking all cases along that continuum. The first of these judgments is not really a distributive judgment, at least in the sense that it is not part of the adjudication process. Rather, every society must decide what punishment it will allow for its most egregious case, be it the death penalty or life imprisonment or fifteen years. Once that endpoint is set, the challenge for the adjudication system is to determine whom to punish and how much punishment to impose. The process of distributing punishment requires only an ordinal ranking of offenders according to their blameworthiness. The result is a specific amount of punishment, but that amount of punishment is not selected because there is some intrinsic connection between that offense and that amount of punishment. Rather, it is selected because, given the large number of cases with distinguishable blameworthiness and the limited range of punishment, that specific amount of punishment is required to set the offender in his appropriate ordinal rank according to his relative blameworthiness. If the punishment continuum endpoint is changed, the appropriate punishment for each offender changes accordingly.[6]

A wide variety of empirical studies indicate that people broadly share intuitions that serious wrongdoing should be punished and also share intuitions about the relative blameworthiness of different transgressions.[7] In some studies, subjects were asked to put offenses or offense scenarios into one of a set of predetermined categories; in another kind of study, subjects were asked to rank-order offenses or offense scenarios; in a third kind of study, subjects were asked to assign numerical values to each of a number of offenses or offense scenarios.[8]

---

> punishment deserved. If rape is a crime and murder is a crime, rape-murder must be more serious than either. Does rape-murder deserve the sum of the punishments meted out for rape and for murder? More? Less? Even when crimes are nearly homogeneous, assigning seriousness is arbitrary: Is rape more serious than assault with a deadly weapon? Is burglary more serious than fraud when fraud does more harm? What about mishandling toxic waste? Ordinal determinations of seriousness become altogether arbitrary when the seriousness of heterogeneous crimes must be compared.

Ernest van den Haag, *Punishment: Desert and Crime Control*, 85 MICH. L. REV. 1250, 1254 (1987).

6. Robinson, *supra* note 1 (manuscript at 7-8).

7. For a general discussion of these matters, see Paul H. Robinson & Robert Kurzban, *Concordance & Conflict in Intuitions of Justice*, 91 MINN. L. REV. 1829, 1854-55 (2007).

8. *Id.* at 1837-40.

The results in all of these studies are consistent. First, subjects displayed a good deal of nuance in the judgments they made.[9] Small changes in facts produced large and predictable changes in punishment. Second, subjects indicated an interest in imposing punishment for serious wrongdoing.[10] Finally, subjects indicated considerable agreement about the relative amount of punishment that different offenders deserved.[11] Indeed, the ordinal ranking of cases generally is consistent across demographic categories, including different cultural groups, as evidenced in cross-cultural studies that replicated domestic studies.[12] Typical of the conclusions in these studies, and according to Graeme Newman,"it is apparent that there was considerable agreement as to the amount of punishment appropriate to each act,"[13] and looking at relative rankings indicates "general agreement in ranks across all countries."[14]

The striking extent of the agreement on intuitions of justice is illustrated in a recent study that asked subjects to rank-order twenty-four crime scenario descriptions according to the amount of punishment deserved.[15] Most researchers would consider this a quite demanding task, perhaps asking for more concentration and effort than most subjects are willing or able to provide. Yet the study's authors found that their subjects had little difficulty performing the task. Moreover, the task was also quite complex, requiring subjects to compare the deserved punishment for each scenario to the deserved punishment for each of the other twenty-three scenarios. Nonetheless, subjects displayed an astounding level of agreement in the ordinal ranking of the scenarios.

A statistical measure of concordance is found in Kendall's coefficient of concordance ("Kendall's W"), in which 1.0 indicates perfect agreement and 0.0 indicates no agreement. In the study just

---

9. *Id.* at 1846.

10. *Id.* at 1842-53.

11. *Id.* at 1854-65.

12. *Id.*

13. GRAEME NEWMAN, COMPARATIVE DEVIANCE: PERCEPTION AND LAW IN SIX CULTURES 140 (1976).

14. *Id.* at 141-43 (providing a table listing data by country); *see also id.* at 135-48 (discussing variability deriving from differences in views regarding how particular acts should be controlled or punished). People from different cultures might share the intuition that an act is wrong, and even a view on each act's relative seriousness, yet still differ in how punishment should be imposed, whether by the state, family, or some other source. This discussion highlights the importance of assessing intuitions regarding seriousness as distinct from preferred punishments meted out by the state. While the former might be correlated strongly with the latter in some contexts, it will be less so in others.

15. Robinson & Kurzban, *supra* note 7, at 1867-74.

described, the Kendall's W was 0.95 (with $p < .001$).[16] This is a striking level of agreement. One might expect a similarly high Kendall's W if subjects were asked to judge the relative brightness of different groupings of spots, for example.[17] In the context of more subjective or complex comparisons, such as asking travel magazine readers to rank the attractiveness of eight different destinations, a Kendall's W of 0.52 is typical.[18] When asking economists to rank the top twenty economics journals according to quality, one gets a Kendall's W of 0.095.[19]

How can it be that when asking people to perform a comparative task as complex and subjective as assessing the relative blameworthiness of twenty-four different offenders, their level of concurrence matches that for much more simple and objective comparative tasks? That is the puzzle that this Article explores.

One possible explanation is the effect of evolutionary processes, which can generate cognitive and behavioral predispositions that help individuals solve commonly encountered problems in ancestral environments. Part II explains how these processes might underlie the phenomenon of shared intuitions of justice. Part III examines evidence from animal behavior studies, brain science, and child development studies that are consistent with the evolutionary explanation. Part IV considers an alternative explanation: that shared intuitions of justice might arise through general social learning because particular intuitions of justice provide efficient norms for group functioning. However, the social learning explanation has a variety of difficulties. To note just one, consider the wide diversity in the life experiences of individuals, and in the social context and structure of groups. Given this, one would expect a social learning process to produce far more than the observed variations in intuitions of justice. On present

---

16. *Id.*

17. *See* Charles M.M. de Weert & Noud A.W.H. van Kruysbergen, *Assimilation: Central and Peripheral Effects*, 26 PERCEPTION 1217, 1219-24 (1997) (obtaining a Kendall's W of 0.95 when subjects compared brightness).

18. Baruch Fischhoff et al., *Travel Risks in a Time of Terror: Judgments and Choices*, 24 RISK ANALYSIS 1301 (2004).

19. The analysis of complex comparisons of journal rankings among economists reveals a relatively low Kendall's W:

> These results unveil significant diversity in the journal quality perceptions among groups of economists despite the fact that our sample focused on [American Economic Association] members. To test the robustness of this claim, using Kendall's W we examined the correlation in journal quality perceptions between any two randomly selected economists in our sample. We found Kendall's W for the top ten journals in our rankings to be 0.396, which demonstrates a relatively low level of agreement among economists. Once we extended this exercise to the top 20 journals in our rankings, Kendall's W dropped to only 0.095.

Kostas Axarloglou & Vasilis Theoharakis, *Diversity in Economics: An Analysis of Journal Quality Perceptions*, 1 J. EUR. ECON. ASS'N 1402, 1421 (2003).

evidence, we conclude, the evolutionary explanation of shared intuitions of justice seems more plausible than a universal social learning explanation.

## II. AN EVOLUTIONARY EXPLANATION

Part I paints a striking picture of human intuitions of justice: across demographics, even across cultures, humans share nuanced intuitions (1) about what constitutes serious wrongdoing, (2) that serious wrongdoing should be punished, and (3) about the relative blameworthiness of offenders. How can matters of such nuance and complexity show this degree of agreement despite dramatic differences in individuals' social and cultural context and upbringing? How can it be that persons of dramatically different religious, economic, and educational backgrounds and of different races, ages, and genders share the same intuitions of justice? We suggest that one explanation for this homogeneity of human intuitions of justice derives from that which all humans share by virtue of being human: their unique evolutionary history and resulting human nature.[20]

In this Part, we first briefly offer a primer for the role of genes in the development of an individual's traits; we then describe how evolutionary processes, most notably natural selection, operate to preserve both anatomical and behavioral traits that helped to solve challenges regularly encountered by ancestors over evolutionary time. We then explain how the process of solving those challenges can lead to a variety of common psychological preferences in humans, including some general moral preferences. And we subsequently discuss how, more specifically, those evolutionary processes can result in shared intuitions of justice in the context of punishments for wrongdoing.

### *A. The Expression of Genes*

It is relatively uncontroversial that evolutionary processes, genes, and environments all influence human predispositions and

20. In this, we extend work that addresses the intersections of law and evolutionary processes. *See, e.g.*, THE SENSE OF JUSTICE: BIOLOGICAL FOUNDATIONS OF LAW (Roger D. Masters & Margaret Gruter eds., 1992); Owen D. Jones & Timothy H. Goldsmith, *Law and Behavioral Biology*, 105 COLUM. L. REV. 405 (2005); Owen D. Jones, *Proprioception, Non-Law, and Biolegal History*, 53 FLA. L. REV. 831 (2001); Michael T. McGuire, *Moralistic Aggression, Processing Mechanisms, and the Brain: The Biological Foundations of the Sense of Justice*, *in* THE SENSE OF JUSTICE, *supra* at 31, 44 (arguing that there is likely "a species-characteristic biological basis for assessing events as just or unjust"); Society for Evolutionary Analysis in Law (SEAL), Owen D. Jones, Useful Sources: Biology, Evolution, and Law, http://www.sealsite.org (located under "Scholarly Resources" tab) (last visited Oct. 7, 2007).

resultant behaviors. But precisely how those influences interact, on which predispositions and behaviors, and with what consequences and implications, represent important questions under continuing investigation. Even what is known is often misunderstood. Consequently, we start with a brief primer on the general principles by which evolutionary processes, genes, and environments can affect behaviors, and how they cannot.[21]

### 1. Genetic Influences versus Genetic Determinism

Genetic explanations do not imply that a trait is "hard wired" or "inflexible." All human traits exist because of the interaction between an individual's set of genes—a subset of all possible human genes—and the unique combination of environmental conditions a person encounters while developing.[22] The effect of any given gene depends on many factors, such as the presence or absence of other genes, the body's internal environment (e.g., cells), the individual's external environment, and so forth.[23] Gene expression, therefore, is extremely complex and is always influenced by the (broadly construed) environment the gene encounters.

It is clear that changing genes or changing environments changes traits, whether physical or psychological. And there is consequently widespread consensus that environmental determinism and genetic determinism are both incorrect.[24] Human minds are neither "blank slate[s]" to be written on by experience,[25] nor completely the product of genetics. They are, rather, the products of complex interactions between genes and their developmental environments, all under the relentless influence of evolutionary

---

21. Key works in evolutionary biology generally include DOUGLAS J. FUTUYMA, EVOLUTIONARY BIOLOGY (3d ed. 1998) and MARK RIDLEY, EVOLUTION (3d ed. 2004). Useful works with behavioral emphases include JOHN ALCOCK, ANIMAL BEHAVIOR: AN EVOLUTIONARY APPROACH (8th ed. 2005); RICHARD DAWKINS, THE SELFISH GENE (30th anniversary ed. 2006); and TIMOTHY H. GOLDSMITH & WILLIAM F. ZIMMERMAN, BIOLOGY, EVOLUTION, AND HUMAN NATURE (2001). Primers written explicitly for legal thinkers also appear in Jones & Goldsmith, *supra* note 20, at 426-31 and Owen D. Jones, *Evolutionary Analysis in Law: An Introduction and Application to Child Abuse*, 75 N.C. L. REV. 1117, 1127-57 (1997).

22. *See, e.g.*, MARY JANE WEST-EBERHARD, DEVELOPMENTAL PLASTICITY AND EVOLUTION (2003).

23. *Id.*

24. JEFFREY L. ELMAN ET AL., RETHINKING INNATENESS: A CONNECTIONIST PERSPECTIVE ON DEVELOPMENT (1996); STEVEN PINKER, THE BLANK SLATE: THE MODERN DENIAL OF HUMAN NATURE 76, 113 (2002); MATT RIDLEY, NATURE VIA NURTURE: GENES, EXPERIENCE, AND WHAT MAKES US HUMAN (2003); Jones & Goldsmith, *supra* note 20, at 485-88.

25. This still-common view was given powerful voice by John Locke. *See, e.g.*, JOHN LOCKE, AN ESSAY CONCERNING HUMAN UNDERSTANDING 121 (Mortimer J. Adler ed., Encyclopaedia Britannica 1952) (1690).

pressures that have favored some patterns of interactions over others.[26]

### 2. The Genetic Code as a Recipe for Development

Although a "blueprint" provides a tempting metaphor for the genetic code, a better metaphor is a "recipe."[27] Without the presence of the appropriate environment (a cook and kitchen), genes (recipes) have no important causal consequences. In the realm of biological development, the dynamic interplay between genes and environment is exceedingly complex.[28] Biochemical interactions at the cellular level and the building of connections in the brain are vastly intricate and still little understood. However, this complexity does not imply that no prediction can be made about an individual's development, nor does it imply that development is so complex as to be essentially random. Rather, even psychological traits that are "learned," such as language, are brought about by a predictable interaction between an individual's genes and the environment in which the individual develops.[29]

The genes that humans share cause each person's structures to be built in a predictable and systematic way. Consider the human visual system. Every human visual system receives different stimuli because every baby is born into a unique perceptual world with different objects (e.g., people and animals). However, the mature visual system for all normally developed adults is essentially the same. When a trait such as this emerges across a wide range of environments, it is said to be "reliably developing." Human genes ensure that even these complex elements of the body are "reincarnated" each generation.[30] Because the developmental process for human psychological traits is complex, it is easier to see such trait "reincarnation" in physical traits. Consider *Gray's Anatomy*.[31] A book

---

26. PINKER, *supra* note 24.

27. *See* RICHARD DAWKINS, THE BLIND WATCHMAKER: WHY THE EVIDENCE OF EVOLUTION REVEALS A UNIVERSE WITHOUT A DESIGN 52 (W.W. Norton & Co. reissue ed. 1996) ("The genes, as we shall see, are more like a recipe than like a blueprint . . . .").

28. For an overview of recent research in the field of evolutionary development, see SEAN B. CARROLL, ENDLESS FORMS MOST BEAUTIFUL: THE NEW SCIENCE OF EVO DEVO AND THE MAKING OF THE ANIMAL KINGDOM (2005).

29. John Tooby & Leda Cosmides, *The Psychological Foundations of Culture*, *in* THE ADAPTED MIND: EVOLUTIONARY PSYCHOLOGY AND THE GENERATION OF CULTURE 19, 83-84 (Jerome H. Barkow et al. eds., 1992). For this reason, attempting to dichotomize important human traits as only culturally or genetically influenced is misguided.

30. John Tooby et al., *The Second Law of Thermodynamics is the First Law of Psychology*, 129 PSYCHOL. BULL. 858, 863-64 (2003).

31. HENRY GRAY, GRAY'S ANATOMY: THE ANATOMICAL BASIS OF MEDICINE AND SURGERY (Susan Standring ed., Elsevier Churchill Livingstone 39th ed. 2005) (1858).

about "the human body" is possible only because all humans develop in the same reliable, systematic fashion, despite the obvious complexities of that process. And, important to our purposes here, even psychological traits, such as the capacity to learn a language, follow a similar pattern.

Of course, development is not necessarily uniform across all individuals. In the context of language learning, the developing child has a special capacity to acquire the ability to understand and produce language that the child hears spoken.[32] But all normal humans have this special language-acquisition mechanism. Other species do not. A cat exposed to the same linguistic environment learns to understand few words, if any. We will argue below that, just as they do with language, people have a specific ability to acquire intuitions of justice. The intuitions that result from this acquisition mechanism are not "innate" in a naïve sense of the term, such as "present at birth." Instead, we argue that there exists a predisposition by which people acquire these intuitions over the course of development both as a child and as an adult.

Which language a child acquires depends on the linguistic environment, but the genes that allow language acquisition play a role as important as the presence of linguistic sounds.[33] All normally developing humans in Topeka acquire the syntax and vocabulary of English, while those in Tokyo learn Japanese. In this instance, the environmental variation leads to differences in the developed adult. In short, the effect of experience varies from one trait to another, with some traits being relatively constant across environments (vision) and others being different across environments (language).

The reason for the differential effect of environment can be traced back to the logic of evolved adaptations (which we describe in the next section). The process of evolution resulted in the selection of a developmental process, which results in the development of the visual system in such a way that it does not depend on the details of the specific stimuli that are in one's visual world because, by virtue of the universal principles of optics, the same solution to vision applies in every setting. This is not the case for language, however, as the most useful language to learn depends on the language others around you

32. *See* NOAM CHOMSKY, KNOWLEDGE OF LANGUAGE: ITS NATURE, ORIGIN, AND USE 3 (1986).

33. *See* Gary F. Marcus & Simon E. Fisher, *FOXP2 in Focus: What Can Genes Tell Us about Speech and Language?*, 7 TRENDS COGNITIVE SCI. 257 (2003).

are speaking. For this reason, evolutionary processes left specific elements of language acquisition to vary across groups.[34]

### *B. Evolution as a Process of Solving Survival Problems*

Most of the significant genes that distinguish us as "human" are ones that were selected over the course of the last 250,000 generations because their effects led to reproductive success.[35] More specifically, current human genes are in large measure the ones that caused individuals to survive and reproduce better, on average, than did alternative genes. Evolution by natural selection, therefore, is the origin of all complex, functional human traits, whether physical or psychological.[36] The success of one gene over another is generally the result of its contribution to solving problems that humans faced, such as finding nutritious food, choosing good mates, raising a family, avoiding predation, and so forth. Genes that contributed to success in confronting these problems were selected (meaning they tended to appear in increasingly large percentages in successive generations). Thus, the traits these genes produced can be understood as having functions, or means by which these problems were solved.

Traits are typically specific in function. That is, to guide the organism toward positive reproductive outcomes, specific traits are needed to solve different problems. This is evident in physical organ systems, for example, as the heart is good at pumping blood but bad at filtering it for toxins, while the reverse is true for the liver. And the human visual system is good at using light to learn about what is out in the world, but it is not well equipped to acquire language. Specificity of function is the hallmark of natural selection because

---

34. STEVEN PINKER, THE LANGUAGE INSTINCT: HOW THE MIND CREATES LANGUAGE 240-43 (1994).

35. This estimate is commonly used. *See, e.g.*, Colm O'hUigin et al., *The Implications of Intergenic Polymorphism for Major Histocompatibility Complex Evolution*, 156 GENETICS 867, 873 (2000). It corresponds to a period of roughly five million years, the estimated time since human and chimpanzee ancestors diverged. Feng-Chi Chen & Wen-Hsiung Li, *Genomic Divergences Between Humans and Other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees*, 68 AM. J. HUM. GENETICS 444, 444 (2001).

36. Obviously, there is more to this issue than can be covered here. Changes in gene frequencies in a population can occur as the result of multiple processes, including genetic drift and natural selection. *See, e.g.*, DAWKINS, *supra* note 21. However, natural selection is commonly considered the one process that can generate both increases in complexity and a close fit between the features of an organism and its environment. "Natural selection" refers to the result of three conditions: a) variation, b) heritability, and c) differential reproduction of individuals as a function of their heritable variations. The process was famously identified and described in CHARLES DARWIN, THE ORIGIN OF SPECIES BY MEANS OF NATURAL SELECTION (Penguin Classics ed. 1985) (1859).

mechanisms such as the heart and the visual system tend to work best when they help to solve a narrow task, in the same way that tools work best when they are designed for one particular job. The same holds for many psychological traits. In sum, much of human psychology consists of evolved, reliably developing traits that have functions associated with the adaptive problems faced by our ancestors. Modern human minds contain psychological adaptations that successfully solved our ancestors' adaptive problems.[37]

### *C. The Evolution of General Moral and Proto-Legal Sentiments*

No one yet knows the full suite of psychological adaptations that human minds share. And given the complexity of the interactions between genes and culture, it is often unclear what can be said with confidence about how the human brain works. Yet in order to bridge from the general notion of evolved psychological adaptations to our specific suggestion that shared intuitions of justice reflect evolved adaptations, we need to lay a brief foundation for the idea that preferences—and indeed loosely "moral" sentiments—can evolve. Fortunately, there is ample reason, at both theoretical and empirical levels, to believe that human minds share a wide variety of preferences that range from the hedonic to the moral.

Like most evolved human capacities, the psychological adaptations that steer humans through the complex social world have functions that led to reproductive success in the past. The function and concurrent reproductive advantages conferred by some human social adaptations are relatively obvious. Few could doubt that our species-typical preferences to sleep at night (on average) rather than during the day, to provision our own children over the children of other people, or to seek pleasure in sexual activity with others of our own rather than another species, reflect complex cognitive processes that in turn reflect evolutionary adaptations as well as the cultural overlays that may adjust the contexts and patterns in which we do so. The evolutionary logic is clear. And other species behave similarly, given the physical and behavioral niches that they have evolved to exploit.[38]

The conclusion from the foregoing is that complex functional human psychological and behavioral traits are the results of adaptation through natural selection. These include not only traits

37. John Tooby & Leda Cosmides, *The Past Explains The Present: Emotional Adaptations and the Structure of Ancestral Environments*, 11 ETHOLOGY & SOCIOBIOLOGY 375 (1990).

38. *See generally* GOLDSMITH & ZIMMERMAN, *supra* note 21; RIDLEY, *supra* note 21.

relevant to mating and parenting and kinship, but also traits relating to the particular challenges of group living, which include aggression, competition, cooperation, deception, and moral sentiments enabling the evaluation of good and bad behaviors.[39] Previous work has suggested that these preferences resulted in proto-legal systems of social primate interaction.[40]

Consider the sentiments leading to judgments and, later, laws surrounding incest. Because of the detrimental effects of inbreeding, evolution appears to have selected for genes that cause organisms to develop behavioral systems that lead them away from mating with close genetic relatives.[41] In humans, this manifests itself as disgust and, perhaps concomitantly, a shared sense that committing incest is wrong.

The intuition has been shown through experimentation not to derive from logical rationales. Experimental subjects insist on the wrongfulness of incest, even if they are unable to articulate principled reasons for why incestuous relationships are wrong.[42] Incest-avoiding mechanisms serve their evolutionary function without awareness of this function on the part of the individual organism.[43] Such an

---

39. *See generally* FRANS B.M. DE WAAL, GOOD NATURED: THE ORIGINS OF RIGHT AND WRONG IN HUMANS AND OTHER ANIMALS (1996); THE HANDBOOK OF EVOLUTIONARY PSYCHOLOGY (David M. Buss ed., 2005); INVESTIGATING THE BIOLOGICAL FOUNDATIONS OF HUMAN MORALITY (James P. Hurd ed., 1996); ROBERT WRIGHT, THE MORAL ANIMAL: THE NEW SCIENCE OF EVOLUTIONARY PSYCHOLOGY (1994); Dennis Krebs, *The Evolution of Morality*, *in* THE HANDBOOK OF EVOLUTIONARY PSYCHOLOGY, *supra*, at 747, 768 ("The mechanisms that give rise to morality are biological adaptations . . . .").

40. *See, e.g.*, OSTRACISM: A SOCIAL AND BIOLOGICAL PHENOMENON (Margaret Gruter & Roger D. Masters eds., 1986).

41. *See* 2 EDVARD WESTERMARCK, THE HISTORY OF HUMAN MARRIAGE 218-39 (5th ed. 1971) (1891) (arguing that natural selection eliminated injurious inbreeding tendencies); Irene Bevc & Irwin Silverman, *Early Proximity and Intimacy Between Siblings and Incestuous Behavior: A Test of the Westermarck Theory*, 14 ETHOLOGY & SOCIOBIOLOGY 171 (1993).

42. *See, e.g.*, Debra Lieberman et al., *Does Morality Have a Biological Basis?: An Empirical Test of the Factors Governing Moral Sentiments Relating to Incest*, 270 PROC.: BIOLOGICAL SCI. 819 (2003); *see also* Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOL. REV. 814 (2001); Jonathan Haidt & Matthew Hersh, *Sexual Morality: The Cultures and Emotions of Liberals and Conservatives*, 31 J. APPLIED SOC. PSYCHOL. 191 (2001). This insistence applies even to cases in which, for example, the consummated sex act is between adult siblings, fully consensual, one time, and incapable of leading to pregnancy. *Id.* at 197, 213-14.

43. Neither a conscious motive to avoid deleterious inbreeding nor an ability to detect the degree of genetic relatedness is necessary. Even the inclination in many animals to leave their natal group upon reaching sexual maturity can serve this function. *See, e.g.*, T.H. Clutton-Brock, *Female Transfer and Inbreeding Avoidance in Social Mammals*, 337 NATURE 70, 70 (1989). In humans, as in many species, for example, natural selection appears to have favored a predisposition against mating with those who were co-resident opposite-sex sihlings during childhood—which has historically been a reliable proxy for close genetic relatedness. Lieberman et al., *supra* note 42, at 825.

intuition is highly adaptive: people who had this intuition, in contrast to those not having it, enjoyed greater long-term reproductive success, leaving more offspring who likewise successfully reproduced.[44] This was accomplished by directing the individual away from inbreeding and its concurrent genetic costs. Incest provides a clear example of how natural selection can favor the evolution of intuitions about wrongdoing whose particular functions are relevant to reproductive success.

### *D. The Evolution of Intuitions of Justice*

Given the influence of genes, environments, and evolutionary processes on human problem-solving through evolved behavioral predispositions—including moral ones—it is a small but significant logical step to suggest that shared intuitions of justice, specific to punishments for transgressions of varying seriousness, also reflect our common evolutionary history. We argue below 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. Further, we argue there is reason to believe that evolution has in particular contributed to intuitions that physical harm, the taking of property, and cheating in exchanges are matters for particular attention and condemnation.[45]

#### 1. Human Sociality and the Predisposition to Acquire Shared Intuitions of Justice

The predisposition to acquire human intuitions of justice likely arose from the deeply and innately social nature of the human species. Humans have been described as "ultrasocial" by anthropologists because of the many ways in which humans interact with one another, particularly in the context of cooperation between individuals and

---

44. 2 WESTERMARCK, *supra* note 41, at 220, 236-39,

45. Recently, Marc Hauser suggested how evolution could have led to the human capacity for morality. Though similar in locating the origins of justice in the theory of evolution by natural selection, and in using language as an analogy to motivate the analysis, our treatment differs in that it focuses on the advantages conferred to individuals in having particular intuitions regarding punishment for wrongful acts. Hauser links the intuition that wrongdoing be punished to intuitions about fairness. MARC D. HAUSER, MORAL MINDS: HOW NATURE DESIGNED OUR UNIVERSAL SENSE OF RIGHT AND WRONG 108-09 (2006) ("We can safely assume that these intuitions evolved prior to or during our life as hunter-gatherers . . . . In such small-scale societies, fairness was most likely an effective proxy for judging punishable acts.").

within groups.[46] Group cooperation is relatively uncommon in the biological world—though there are exceptions, such as certain group-living insects—and captures the attention of biologists because evolution, broadly, tends to favor genes that lead to selfish behavior. Humans, however, behave altruistically in many contexts: family relations, friendships, and within-group cooperation. How does such extensive cooperation come to pass? The evolution of cooperation in humans and other species is the subject of a large theoretical and empirical literature in evolutionary biology,[47] and the general outline is relevant to our argument.

Both theoretical and empirical scholarship demonstrates that cooperation can evolve through several independent but overlapping processes.[48] The one most relevant for our immediate purpose concerns the mutually beneficial effects of reciprocity: if you share with me today in exchange for my sharing with you yesterday, we are both better off than if neither of us shares. In social animals, reciprocity can involve such things as alerting other group members when food has been discovered, sharing food over time, and supporting a comrade in action against others.

But underlying this rosy picture is a darker shadow. While it is evident that reciprocators can outperform loners, a cheater (or defector) could theoretically outperform both if he were able to take benefits regularly without repaying them. Consequently, an

---

46. Peter J. Richerson & Robert Boyd, *The Evolution of Human Ultrasociality*, *in* INDOCTRINABILITY, IDEOLOGY, AND WARFARE: EVOLUTIONARY PERSPECTIVES 71, 71-72 (Irenäus Eibl-Eibesfeldt & Frank Kemp Salter eds., 1998).

47. That literature suggests that several causal processes are simultaneously at work. One, known as mutualism, reflects the fact that two working together can sometimes gain more than the sum of their actions, if they had acted alone. A second, known as kin selection, reflects the extent to which cooperatively furthering the reproductive success of kin increases the prevalence of any beritable predisposition to do so, since many kin will also carry the genes that contribute to that predisposition. A tbird, known as reciprocal altruism, reflects the fact that heritable predispositions to cooperate over time can yield mutual benefits, even between unrelated cooperators. On the extent to which these processes may be supplemented by other processes, see Herbert Gintis, *Strong Reciprocity and Human Sociality*, 206 J. THEORETICAL BIOLOGY 169 (2000).

There is debate surrounding why humans evolved to be the extremely social animals that they are. For purposes of the current discussion, we remain agnostic on this debate and simply assume that humans are social creatures, have been for many generations, and have many traits that are adapted for social life. *See, e.g.*, Leda Cosmides, *The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason?*, 31 COG. 187, 187-97 (1989); Tooby & Cosmides, *supra* note 29, at 19, 20; Robert L. Trivers, *The Evolution of Reciprocal Altruism*, 46 Q. REV. BIOLOGY 35, 35-39 (1971). On the possible origins of social exchange in humans, and the inverse relationship between social hierarchy and social exchange, see CHRISTOPHER BOEHM, HIERARCHY IN THE FOREST: THE EVOLUTION OF EGALITARIAN BEHAVIOR 197, 197-99, 213 (1999).

48. *See generally* LEE ALAN DUGATKIN, COOPERATION AMONG ANIMALS: AN EVOLUTIONARY PERSPECTIVE (1997).

evolutionary arms race ensues in social animals between various predispositions toward cooperation and exploitation. In the end, the most successful cooperators are not those who always cooperate, but those who cooperate *selectively* with other cooperators, thus discriminating (passively or aggressively) against those who are not reliable partners in cooperative endeavors.[49] Put another way, effective cooperation requires rewarding good behavior and punishing (or at least failing to reward) bad behavior.

For reciprocal altruism to work, though, one must be able to evaluate the costs and benefits of various kinds of things (such as different objects, acts, and information) and evaluate others' valuation of those objects. One also must be able to discern unfairness, such as an inequitable distribution of resources or a violation of reciprocity. Furthermore, one must be able to recognize individuals and remember recent interactions with them, so as to discriminate in the future in favor of some and against others.[50] These are precisely the features that we tend to observe in the most highly social animals.[51]

Those features are also, we argue, precisely those that likely underlie the functional advantage of the specific learning capacity in humans for acquiring shared intuitions of justice. Humans have a universal and uniquely nuanced propensity for engaging in social exchange.[52] Indeed, gains from social exchange form the basis of the modern economy and infiltrate nearly every aspect of our lives,[53] both in formal markets and in personal relationships. The psychology that underpins exchange requires deep intuitions and complex computational capacities to operate.

In particular, one critical capacity for successful social exchange is the intuition that one should punish individuals who injure others or cheat in an exchange. If one is engaged in transactions with the same person over time, then allowing another to injure or to cheat with no punishment is an invitation to exploit without end. Therefore, to be most successful in social exchange, one must have the capacity not only to detect but also to punish such persons.

---

49. *See, e.g.*, ROBERT AXELROD, THE EVOLUTION OF COOPERATION (1984).

50. For a detailed list of these requirements, see Tooby & Cosmides, *supra* note 29, at 110.

51. *See generally* DE WAAL, *supra* note 39 (discussing reciprocal altruism and its basis in the remembrance of favors given and received).

52. Social exchange is the simultaneous or sequential delivery of benefits between two individuals for mutual profit.

53. *See generally* ROBERT WRIGHT, NONZERO: THE LOGIC OF HUMAN DESTINY (2000). Exchange allows non-zero-sum interactions, in which each participant is better off exchanging than alone. *Id.* at 26.

This implies that there might have been selection in humans for the cognitive mechanisms designed to detect inequities and, similarly, for the cognitive mechanisms that yield intuitions that motivate the punishment of people who violate the most ancient and fundamentally necessary principles of social exchange.[54] In other words, the evolutionary history of social exchange has likely led to a reliably developing psychological system that is able to compute when someone has injured or cheated and to motivate punishment.[55]

2. Intuitions of Justice as Generating Benefits for Individuals

The argument sketched so far may explain why humans have intuitions that lead them to believe that people who have harmed or cheated *them* ought to be punished. However, this does not explain why people have intuitions that punishment should be imposed on wrongdoers in transactions in which they have no part. This question—why people care about third-party transactions—opens the door to an important mystery. Most non-human organisms attend very little to transactions that occur between unrelated others (though there are some important exceptions[56]). Why do humans care so much about these interactions? Our analysis follows along these lines:

Once individuals possess intuitions about being wronged, detecting when others have been wronged is possible using the same set of intuitions. This does not, by itself, explain why individuals should care that someone else has been wronged; it merely explains the origin of the ability to detect wrongful acts toward others.

What may explain the interest in wrongdoing to others is that individuals need to make decisions about the people with whom they interact, engage in social exchange, form groups and coalitions, and, more generally, socialize. One observing a third party committing a wrong against another person has important information about the perpetrator of the wrong; this information may make the observer less likely to choose that person for social interactions.

Another puzzle exists: why do people have the intuition that wrongs to others should be punished? The evolutionary advantage of this intuition is debatable,[57] but we advance some ideas that help

54. For recent discussion, see Herbert Gintis, *Strong Reciprocity and Human Sociality*, 206 J. THEORETICAL BIOLOGY 169 (2000).

55. For suggestive nonhuman evidence, see *infra* notes 76-82 and accompanying text.

56. *See infra* notes 87-91 and accompanying text.

57. For early thinking on this subject, see Trivers, *supra* note 47, *passim*, which discusses how selection can operate against those who fail to reciprocate; see also MATT RIDLEY, THE ORIGINS OF VIRTUE: HUMAN INSTINCTS AND THE EVOLUTION OF COOPERATION (1996), which

explain the preference. The logic begins with the uncontroversial premise that humans evolved in relatively small groups consisting of people with whom they had a large number of transactions over a sustained period.[58] Consider two kinds of groups: first, one in which individuals whose acts of unprovoked violence, theft, or cheating in exchanges is left unpunished and, second, one in which these acts are punished. In the second group, the individuals in the group benefit because they are less likely to be subject to violence, theft, or cheating. Thus, each person, to gain that benefit, has an incentive to inflict punishment on those who commit such offenses. Of course, punishment is potentially costly, so the long-term benefits of punishing must outweigh the long-term costs if it is to be selected.[59] And very recent experimental evidence in humans suggests that under such circumstances, given a choice, even prior transgressors choose to be in a group that punishes over one that is punishment-free.[60]

However, personally punishing wrongdoers is not the only way to increase the chances of being in a group in which undesirable acts are punished. One need not inflict punishment oneself. One can decrease the cost of inflicting punishment on wrongdoers by supporting or defending those who inflict the punishment. A third-party punisher has a smaller risk of retaliation if group members endorse the punishment meted out by the third party. Accordingly, an individual benefits by having intuitions that support the punishment—by some individual—of those who transgress.[61] In the

---

traces, among other things, the extent to which moralistic aggression can police fairness in social exchanges, and thereby increase the prevalence of reciprocity by increasing the cost of cheating.

58. *See generally* BOEHM, *supra* note 47, *passim*.

59. *See* Richerson & Boyd, *supra* note 46, *passim*.

60. Özgür Gürerk, Bernd Irlenbusch & Bettina Rockenbach, *The Competitive Advantage of Sanctioning Institutions*, 312 SCIENCE 108, 110 (2006).

61. Alexander made similar arguments regarding reputation and the effects on the group of having cooperative individuals. RICHARD D. ALEXANDER, THE BIOLOGY OF MORAL SYSTEMS 94, 153 (1987). Trivers made a related argument. Robert Trivers, SOCIAL EVOLUTION 388 (1985) ("[A] sense of fairness has evolved in human beings as the standard against which to measure the behavior of other people, so as to guard against cheating in reciprocal relationships."); *see also* Krebs, *supra* note 39, at 766-67 (noting that the high individual cost of punishment has prompted members of society to create institutions to catch and punish cheaters); Dennis L. Krebs, *The Evolution of Moral Behaviors*, *in* HANDBOOK OF EVOLUTIONARY PSYCHOLOGY: IDEAS, ISSUES, AND APPLICATIONS 337, 342 (Charles Crawford & Dennis L. Krebs eds., 1998) (arguing that deference gives rise to moral behavior that involves adherence to and respect for the rules and laws of authority); Robert Kurzban & Steven Neuberg, *Managing Ingroup and Outgroup Relationships*, *in* THE HANDBOOK OF EVOLUTIONARY PSYCHOLOGY, *supra* note 39, at 659 (noting that punishment in social exchange is sensible if it prevents future cheating).

We have considered the benefits flowing to individuals, as is standard in biology. Some, proposing "multi-level selection," argue that considering benefits to the group yields an equal or

small groups in which humans have evolved, the marginal benefit of having each individual support punishment for wrongdoings might have reduced the number of transgressions in the group and, thus, protected an individual's health, property, and ability to make contracts.[62]

If it is true that (1) individuals prefer that wrongdoers be punished and (2) individuals may not want to bear the costs of inflicting punishment, it is not surprising that means have been developed to satisfy these two preferences, though the means may differ across the centuries and cultures. For example, in some cultures,[63] one person has special responsibilities and duties in the context of such punishment. In the West, punishment is administered by a system of criminal justice, which includes legislatures, police, judges, and so on.

In short, shared intuitions of justice contribute to the ability of an individual or group to punish, which in turn provides an evolutionary advantage to all. In other words, there is an evolutionary advantage to understanding victimhood and to the concurrent

---

superior analysis. *See, e.g.*, ELLIOT SOBER & DAVID SLOAN WILSON, UNTO OTHERS: THE EVOLUTION AND PSYCHOLOGY OF UNSELFISH BEHAVIOR 331 (1998). In that view, groups that effectively cooperated because of widely shared and specific intuitions of justice that benefited the group could have outperformed groups without those intuitions, leading over time to more groups composed of more people who share those intuitions. Game theoretic analysis suggests that group-benefiting norms, and punishment through inflicting costs on those who do not follow such norms, can be a powerful evolutionary force in the adoption, spread, and maintenance of these norms. *See, e.g.*, Herbert Gintis, Samuel Bowles, Robert Boyd & Ernst Fehr, *Explaining Altruistic Behavior in Humans*, 24 EVOLUTION & HUM. BEHAV. 153, 154 (2003); *see also* Peter J. Richerson & Robert Boyd, *The Evolution Of Subjective Commitment to Groups: A Tribal Instincts Hypothesis*, *in* EVOLUTION AND THE CAPACITY FOR COMMITMENT 186-220 (Randolph M. Nesse ed., 2001).

62. This last element, contracts, is worth special consideration. Because of humans' abilities to represent abstract costs and benefits, the number of social exchanges that are possible is much, much larger than in other organisms, who can exchange only narrowly delimited commodities. If an organism cannot translate how beneficial or costly an object or act is compared to another object or act, she cannot make good decisions about what counts as a good exchange. As the range of possible exchanges increases, the advantage of the ability to enforce contracts increases. *See* Leda Cosmides & John Tooby, *Cognitive Adaptations for Social Exchange*, *in* THE ADAPTED MIND: EVOLUTIONARY PSYCHOLOGY AND THE GENERATION OF CULTURE 163, 177 (Jerome H. Barkow et al. eds., 1992) (discussing the computational requirements of social exchange).

63. The punishment of wrongdoers can be implemented in multiple ways. One way is for the individual who has been harmed to punish. Another way is for individuals not directly involved to inflict punishment. Among the Huron, for example, the obligation to slay a murderer reportedly fell upon the kin of the murderer. BRUCE G. TRIGGER, THE CHILDREN OF AATAENTSIC: A HISTORY OF THE HURON PEOPLE TO 1660, 59-60 (1976). Moreover, another way is for institutions to inflict punishment on those who have committed wrongs. This comes the closest to the way in which modern states impose punishment: through an official criminal justice system tbat formally represents all citizens of the state.

condoning of the punishment of one who has engaged in wrongful conduct.[64] This is important because the imposition of punishment—the imposition of costs against the wrongdoer—is itself something that normally would be seen as a wrongful act. To produce conditions that avoid a never-ending spiral of wrongful acts, intuitions of justice must specify that wrongful acts can be punished without eliciting the further intuition that the punishers should themselves be punished.

To distinguish deserved punishment, which is permitted, from undeserved punishment, which is not, the group obviously must share some sense of what constitutes wrongful conduct and the amount of punishment appropriate in any one case relative to other cases.[65] If everyone in a group agrees on this—what constitutes a wrongful act and its relative wrongfulness as compared to other acts, hence the relative amount of punishment to be permitted—stability and cooperation, and the consequent benefits to members of the group, can be achieved.[66]

Contributing to the institutional arrangement for punishment may not be the only reason that intuitions of justice evolved. Because social exchange is mutually beneficial, participation in exchange is desirable. When there are many social interactants in a group, there is competition to be selected as a partner in exchange. We should therefore expect that natural selection would favor predispositions that signal that one is a good candidate for participation in such exchanges. If one shares others' moral intuitions—that cheating in an exchange is condemnable, for example—then one can expect selection to favor those who signal having such an intuition.[67] Dennis Krebs and Kathy Denton recently made this argument: "[p]ersuading others that you are a fair, honest, generous, responsible and moral person who

64. *See* JAMES Q. WILSON, THE MORAL SENSE 40 (1997) (quoting John Stuart Mill's belief that our sense of justice involves a "desire to punish wrongdoers, even when we are not the victims").

65. The problem alluded to above, *supra* note 62, is essentially identical for punishment because the means of punishment is often not in the same coin as the offense. If intuitions of justice imply that harm should be inflicted on a perpetrator, then computing what constitutes a harm of appropriate magnitude is crucial. This is a richly debated topic in many modern areas of law (e.g., drug offenses). However, we believe that there are strong intuitions about the relative severity of offenses and suggest that the complexity of the psychological systems necessary to make such comparisons should not be underestimated. The cycle of harm, retaliation, and retaliation for retaliation can escalate if the original punishment is judged too great, given the magnitude of the offense.

66. *See* Robert Boyd & Peter J. Richerson, *Solving the Puzzle of Human Cooperation*, *in* EVOLUTION AND CULTURE 105, 114-19 (Stephen C. Levinson & Pierre Jaisson eds., 2006).

67. *See generally* ROBERT H. FRANK, PASSIONS WITHIN REASON: THE STRATEGIC ROLE OF THE EMOTIONS (1988).

will make an attractive exchange partner may induce them to bestow benefits on you."[68]

The emergence of intuitions of justice has a self-sustaining character. Once the intuitions exist in a group, actions that violate others' intuitions invite censure and punishment. Once the intuitions of justice exist, it is disadvantageous to reject publicly the principles of that system or to behave in ways that conflict with others' intuitions.[69] Thus, persons who do not reliably behave consistently with those intuitions or do not signal their agreement in such shared intuitions will be disadvantaged as against those who do.

To illustrate that "core" intuitions of justice should reliably develop in all humans, consider the fate of a mutation that causes individuals not to develop such intuitions. These individuals would be punished by those they injured, stole from, or cheated, and they would not be desired as social exchange partners or as members of groups. For a social creature, such a fate would have been a reproductive disaster.[70] The individual costs associated with having diverging intuitions from those shared by others are a powerful force for homogenizing the intuitions of justice.[71] Thus, once intuitions of justice are in place, there would be strong pressure favoring any variant in humans that acquires those intuitions, thus stabilizing them and making them essentially universal in the species.[72]

One might wonder why some details of human intuitions of justice are widely shared, while others are not.[73] Intuitions of justice that were advantageous no matter what the context should be

---

68. Dennis L. Krebs & Kathy Denton, *Toward a More Pragmatic Approach to Morality: A Critical Evaluation of Kohlberg's Model*, 112 PSYCHOL. REV. 629, 642 (2005).

69. Indeed, formal game theoretical analyses have shown that punishment can lead individuals to behave in accordance with an arbitrarily wide array of norms. Robert Boyd & Peter J. Richerson, *Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups*, 13 ETHOLOGY & SOCIOBIOLOGY 171 (1992).

70. Robert Kurzban & Mark R. Leary, *Evolutionary Origins of Stigmatization: The Functions of Social Exclusion*, 127 PSYCHOL. BULL. 187, 192 (2001); *see also* Roy F. Baumeister & Mark R. Leary, *The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation*, 117 PSYCHOL. BULL. 497, 521 (1995).

71. *See, e.g.*, Joe Henrich & Robert Boyd, *The Evolution of Conformist Transmission and the Emergence of Between-Group Differences*, 19 EVOLUTION & HUM. BEHAV. 215, 236 (1998).

72. Our argument does not turn on universality without exception. Most features of most species appear in ranges reflecting variation in the genes and environments that give rise to unique organisms. The breadth and distributions of features within those ranges is sensitive to (among other things) the strength of selection pressures. Our argument is that there is good reason to hypothesize that selection pressures in this context would yield very similar intuitions of justice that are widespread across the human species.

73. See Robinson & Kurzban, *supra* note 7, at 1892-93, for a comparison of Parts I-III with Part IV, noting that some intuitions of justice seem to be universal, while others outside the "core" wrongdoings spark much disagreement.

relatively consistent across individuals and should develop reliably regardless of one's local environment.[74] These "core" intuitions probably include the notions that unjustified physical aggression, the taking of another's property, and cheating in exchanges are all wrong and should be punished.[75] In contrast, it is unlikely that conduct that less directly prevented beneficial social interaction—or that was more dependent on social, economic, or other contexts—will be included in the "core" intuitions that reliably develop today in all environments. This is reflected in the cross-cultural, and even within-group, divergence on such issues as drug use or prostitution, in which the core intuitions regarding harm, theft, and deception are not violated.

## III. EMPIRICAL SUPPORT FOR AN EVOLUTIONARY EXPLANATION

In Part II, we provided an evolutionary explanation for the little-acknowledged but stunningly consistent and widely shared intuitions of justice. On present evidence, one cannot know whether our specific evolutionary explanation is correct. Unlike bones, behavior does not fossilize, and one can only test the accuracy of evolutionary accounts of behavior by triangulating from available sources of information. This requires combining theories, offered in Part II, with observations and experiments, which we take up in this Part.

There is a good deal of empirical work that needs to be done, but in the meantime, we can look to three existing areas that shed some light on our hypothesis. First, we consider animal studies that suggest evolved, rudimentary notions of fairness and blaming. Second, we consider brain science studies that reveal identifiable physiological processes at work when humans make moral judgments. Finally, we examine the field of developmental psychology, which demonstrates a common path in the development of moral reasoning in the human child across demographics and cultures. We conclude that all three

74. *See* JEROME KAGAN, THE NATURE OF THE CHILD 123, 189 (1984) (discussing the existence of universal moral standards); *see also* Richard A. Shweder & Jonathan Haidt, *The Future of Moral Psychology: Truth, Intuition and the Pluralist Way*, 4 PSYCHOL. SCI. 360, 363 (1993) (discussing cognitive intuitionism's assumptions regarding the objectivity of moral qualities).

75. *See* Jones, *supra* note 20, at 847-67, 860 fig.9 (arguing that these three notions, alongside several others, have evolutionary roots). As a consequence, and to the extent that evolutionary processes influence the content of predispositions, emotions, intuitions of justice, and morality, the major features of legal systems will tend to reflect that influence. *See id.*; Jones & Goldsmith, *supra* note 20, at 466-475, 474 tbl.1. For recent elahoration, see Rohin Bradley Kar, *The Deep Structure of Law and Morality*, 84 TEX. L. REV. 877 (2006).

areas provide evidence generally consistent with our evolutionary hypothesis. In contrast, we know of no data in direct conflict with it.

### *A. Animal Studies Evidence*

If the evolutionary explanation for shared intuitions of justice is correct, one might expect to find in other animals—especially our close primate relatives—rudimentary forms of humans' core intuitions about what constitutes wrongdoing and how it should be punished. In fact, a number of socially cooperative species appear in some circumstances to exhibit characteristics of "punishing" aggressors and cheaters. And a number of researchers now suggest that such behaviors may reflect a rudimentary moral sense or intuition of justice.[76] Victims, the victims' relatives, and others sometimes avoid or act aggressively towards individuals that deviate from various group or dyadic norms and expectations.

These phenomena are not limited to primates, or even to mammals Here are a few of many examples. In one social species of wren, "helpers" assist by providing food when the young are being raised. Helpers experimentally removed from the group during that period are usually attacked and harassed upon their return, while helpers absent at other times of the year are never attacked.[77] Within the highly social naked mole rat communities, queens appear to focus attacks on lazy workers.[78] Wolves apparently refuse to play with those who violate the social rule against injurious play-fighting, and those wolves leave the groups and die at higher-than-average rates.[79]

Researchers studying animal punishment patterns have found that behavior akin to theft is one particular target for retribution.[80] For example, elephant seal pups caught trying to nurse from a female who is not their mother are not just shooed away. Often, they are bitten severely and sometimes killed.[81] Young male deer attempting to

---

76. For an alternative view, see Jeffrey R. Stevens & Marc D. Hauser, *Why Be Nice? Psychological Constraints on the Evolution of Cooperation*, 8 TRENDS COGNITIVE SCI. 60, 61 (2004), which distinguishes punishment from harassment.

77. Raoul A. Mulder & Naomi E. Langmore, *Dominant Males Punish Helpers for Temporary Defection in Superb Fairy-Wrens*, 45 ANIMAL BEHAV. 830, 831 (1993).

78. Hudson K. Reeve, *Queen Activation of Lazy Workers in Colonies of the Eusocial Naked Mole-Rat*, 358 NATURE 147, 147-48 (1992).

79. Marc Bekoff, *Wild Justice, Cooperation, and Fair Play*, *in* THE ORIGINS AND NATURE OF SOCIALITY 53, 62 (Robert W. Sussman & Audrey R. Chapman eds., 2004).

80. T.H. Clutton-Brock & G. A. Parker, *Punishment in Animal Societies*, 373 NATURE 209, 212 (1995).

81. *Id.* (citing Joanne Reiter, Nell Lee Stinson & Burney J. Le Boeuf, *Northern Elephant Seal Development: The Transition from Weaning to Nutritional Independence*, 3 BEHAV. ECOLOGY

sneak copulations with females being guarded by adult males are regularly attacked.[82]

With respect to an animal sense of fair treatment, a leading researcher has concluded that a wide variety of animals is capable of discerning when a situation is not equitable.[83] For example, in a recent and widely reported experiment with capuchin monkeys, different combinations of two adjacent monkeys regularly alternated returning granite tokens for cucumber slices. When the experimenter began to provide one monkey with a grape (a more highly valued food) in exchange for the same token that continued to yield mere cucumbers for the other monkey, that other monkey often manifested considerable distress. It sometimes jumped up and down, throwing the token or the cucumber at the researcher, refusing to eat the cucumber, and the like.[84] This led the authors of the study to conclude that capuchins are capable of comparing their own reward to the reward others receive, and accepting or rejecting rewards according to their relative, not absolute, value.[85] Similarly, chimpanzees reportedly will often refuse to participate in an exchange once another chimpanzee is receiving a more valued reward for the same amount of effort.[86] Thus, both capuchins and chimpanzees behave in ways suggesting that they can perceive unfairness and that it often agitates them.

Behavior suggesting an ability to perceive inequities appears to underlie a great deal of social behavior in primates, in whom transgressive acts are most systematically punished. By way of background, many primates regularly exhibit sophisticated cooperation, which ranges from simple reciprocal grooming and food-sharing to complex tool-using and coalitional behavior. For example, in olive baboons there is a strong correlation between A's prior support

---

& SOCIOBIOLOGY 337, 344 (1978); Joanne Reiter, Kathy J. Panken & Burney J. Le Boeuf, *Female Competition and Reproductive Success in Northern Elephant Seals*, 29 ANIMAL BEHAV. 670, 676 (1981)).

82. Clutton-Brock & Parker, *supra* note 80 (citing T.H. Clutton-Brock, S.D. Albon, R.M. Gibson & F.E. Guinness, *The Logical Stag: Adaptive Aspects of Fighting in Red Deer (Cervus elaphus L.)*, 27 ANIMAL BEHAV. 211, 212 (1979); T.H. Clutton-Brock, D. Green, M. Hiraiwa-Hasegawa & S.D. Albon, *Passing the Buck: Resource Defence, Lek Breeding and Mate Choice in Fallow Deer*, 23 BEHAV. ECOLOGY & SOCIOBIOOLOGY 281, 287-88 (1988)).

83. Sarah F. Brosnan, *Nonhuman Species' Reactions to Inequity and Their Implications for Fairness*, 19 SOC. JUST. RES. 153, 181 (2006).

84. Sarah F. Brosnan & Frans B.M. de Waal, *Monkeys Reject Unequal Pay*, 425 NATURE 297, 298 (2003); Sarah Brosnan & Frans de Waal, Reply, *Animal Behavior: Fair Refusal by Capuchin Monkeys*, 428 NATURE 140, 140 (2004) [hereinafter Brosnan & de Waal, Reply].

85. Bosnan & de Waal, Reply, *supra* note 84.

86. Brosnan, *supra* note 83, at 177; Sarah F. Brosnan, Hillary C. Schiff & Frans B.M. de Waal, *Tolerance for Inequity May Increase with Social Closeness in Chimpanzees*, 272 PROC.: BIOLOGICAL SCI. 253, 255 (2005).

in a conflict for a fellow baboon B, and A's successful recruitment of B to a new conflict in which A becomes involved.[87] In macaques, individuals have been observed to support unrelated others, who have previously supported them by intervening in conflicts on their behalf.[88] Vervet monkeys tend to groom preferentially individuals who have groomed them in the past.[89] When one chimpanzee grooms another, early in the day, the latter is more likely to share food with the former, later in the day.[90] And chimpanzees experimentally engaged in tasks requiring collaboration quickly determine which among different potential collaborators is most effective and recruit that individual preferentially for subsequent collaborative tasks.[91]

On the other hand, failure to cooperate regularly yields sharp consequences. Negative acts are the flip side of the cooperation coin. Authors of a leading paper on the subject conclude that the intensity of punishments often increases with the severity of harm caused by transgressors.[92] For example, rhesus macaques who discover food and are caught having failed to alert the group to its discovery often become targets of significant aggression.[93] In chimpanzee societies, those reluctant to share when they have food are more likely to encounter aggressive responses when they later approach those who have food.[94] Chimpanzees will attack former allies who failed to assist them in conflicts with third parties.[95] And many primates, including chimpanzees, tend to intervene most often against those who have

87. C. Packer, *Reciprocal Altruism in Papio Anubis*, 265 NATURE 441, 442 (1977).

88. Jessica C. Flack & Frans B.M. de Waal, *"Any Animal Whatever": Darwinian Building Blocks of Morality in Monkeys and Apes*, *in* EVOLUTIONARY ORIGINS OF MORALITY: CROSS-DISCIPLINARY PERSPECTIVES 1, 12 (Leonard D. Katz ed., 2000); Frans B.M. de Waal & Lesleigh M. Luttrell, *The Similarity Principle Underlying Social Bonding Among Female Rhesus Monkeys*, 46 FOLIA PRIMATOLOGICA 215 (1986).

89. *See* Robert M. Seyfarth & Dorothy L. Cheney, *Grooming, Alliances and Reciprocal Altruism in Vervet Monkeys*, 308 NATURE 541, 542 (1984).

90. Frans B.M. de Waal, *Food Sharing and Reciprocal Obligations Among Chimpanzees*, 18 J. HUM. EVOLUTION 433, 433 (1989).

91. Alicia P. Melis, Brian Hare & Michael Tomasello, *Chimpanzees Recruit the Best Collaborators*, 311 SCIENCE 1297, 1299 (2006).

92. Clutton-Brock & Parker, *supra* note 80, at 211.

93. Marc D. Hauser, *Costs of Deception: Cheaters Are Punished in Rhesus Monkeys (Macaca mulatta)*, 89 PROC. NAT'L ACAD. SCI. 12,137, 12,137 (1992); Marc D. Hauser & Peter Marler, *Food-Associated Calls in Rhesus Macaques (Macaca mulatta): II. Costs and Benefits of Call Production and Suppression*, 4 BEHAV. ECOLOGY 206, 206 (1993).

94. De Waal, *supra* note 90, at 456; *see also* DE WAAL, *supra* note 39, at 160 (concluding such behavior "suggests a sense of justice and fairness").

95. FRANS DE WAAL, CHIMPANZEE POLITICS: POWER AND SEX AMONG APES 207 (1998).

most often intervened against them.[96] These patterns have prompted leading researchers to refer to such negative reciprocities as a "revenge system."[97]

Indeed, among chimpanzees (who, along with bonobos, are the closest relatives of humans) retribution is common enough that researchers consider retaliation "an integral part of [a] system of reciprocity."[98] Because sharing and other cooperative behavior exists in a "multi-faceted matrix of relationships, social pressures, delayed rewards, and mutual obligations,"[99] the most successful individuals are those who can distinguish good colleagues from bad and deal with each accordingly. Frans de Waal, a prominent primatologist, consequently describes their community as "a 'market' of reward and punishment,"[100] with "balance sheets" on social interactions.[101] Indeed, reciprocity rules consisted of " 'one good turn deserves another' and 'an eye for an eye, a tooth for a tooth.' "[102] "Not only are beneficial actions rewarded," de Waal concludes, but "there seems to be a tendency to teach a lesson to those who act negatively."[103] Reward good deeds and punish the bad.

The data available, of which the foregoing is only a sample, do not prove definitively that some core aspects of human intuitions of justice are evolved adaptations.[104] Nor, even if they are adaptations, did they necessarily arise from the same roots as related intuitions in

---

96. Flack & de Waal, *supra* note 88, at 8; Joan B. Silk, *The Patterning of Intervention Among Male Bonnet Macaques: Reciprocity, Revenge, and Loyalty*, 33 CURRENT ANTHROPOLOGY 318 (1992).

97. *See, e.g.*, Fillipo Aureli et al., *Kin-oriented Redirection Among Japanese Macaques: An Expression of a Revenge System?*, 44 ANIMAL BEHAV. 283, 289-90 (1992) ("Macaques might have an indirect revenge system in which kin relationships play a decisive role."); Frans B.M. de Waal & Lesleigh M. Luttrell, *Mechanisms of Social Reciprocity in Three Primate Species: Symmetrical Relationship Characteristics or Cognition*, 9 ETHOLOGY & SOCIOBIOLOGY 101, 114 (1998) ("Only this species exhibits what may be called a revenge system: chimpanzees tend to intervene against individuals who intervene against themselves.").

98. DE WAAL, *supra* note 39, at 157-58.

99. De Waal, *supra* note 90, at 452.

100. *Id.*

101. DE WAAL, *supra* note 39, at 157-59.

102. DE WAAL, *supra* note 95; *see also* Flack & de Waal, *supra* note 88, at 9 ("Monkeys and apes appear capable of holding received services in mind, selectively repaying those individuals who performed the favours. They seem to hold negative acts in mind as well, leading to retribution and revenge.").

103. DE WAAL, *supra* note 39, at 159; *see also* Flack & de Waal, *supra* note 88, at 9 (describing similar behavior among monkeys and apes).

104. Indeed, some dispute whether primates actually use negative sanctions to shape behavior of third parties or to punish deviation from social norms. Joan B. Silk, *The Evolution of Cooperation in Primate Groups*, *in* MORAL SENTIMENTS AND MATERIAL INTERESTS: THE FOUNDATIONS OF COOPERATION IN ECONOMIC LIFE 43, 60-64 (Herbert Gintis et al. eds., 2005).

other species. However, indifference to wrongdoing in highly cooperative species is exploited easily by cheaters and free-riders, and therefore potentially unstable.[105] Consequently, one would expect frequently to observe punishment (as well as retaliation, selective ostracism, and the like) of wrongdoing in ultra-social species such as ours. And the intersecting vectors of available animal evidence point toward that conclusion. Moreover, the close relationship of humans, chimpanzees, and other primates suggests that we ought not gratuitously, uneconomically, and unparsimoniously assume that there are different causes in humans for the same behaviors in other animals.[106] It is not surprising, then, that the leading animal researchers believe that evolution has supplied the "building blocks" of human morality[107] and that the question is not whether biology has influenced the development of human moral systems, but rather to what degree.[108] "Evolution," they believe, "has produced the requisites for morality: a tendency to develop social norms and enforce them."[109]

### *B. Brain Science Evidence*

Some people may balk at the notion that intuitions of justice could be the product of evolution, even indirectly, because views of justice are obviously matters of complex judgment making. The design of one's visual system might be the product of evolution, but how could evolution affect a person's judgment concerning what punishment is deserved?

Specialized brain structures and brain processes determine all decisionmaking, no matter how complex. And neuroscience has begun to provide details on what structures do what. As demonstrated below,

---

105. In game theoretic terms, a behavioral "strategy" is evolutionarily unstable if it can be outcompeted by an alternative strategy. *See generally* JOHN MAYNARD SMITH, EVOLUTION AND THE THEORY OF GAMES (1982).

106. DE WAAL, *supra* note 39, at 64.

107. Flack & de Waal, *supra* note 88, at 19-24; *see also* HAUSER, *supra* note 45 ("We can safely assume that these intuitions evolved prior to or during our life as hunter-gatherers . . . . In such small-scale societies, fairness was most likely an effective proxy for judging punishable acts."); DE WAAL, *supra* note 39, at 218 ("The fact that the human moral sense goes so far back in evolutionary history that other species show signs of it plants morality firmly near the center of our . . . nature.").

108. Flack & de Waal, *supra* note 88, at 3.

109. DE WAAL, *supra* note 39, at 39; *see also* RICHARD JOYCE, THE EVOLUTION OF MORALITY 140-42 (2006) (explaining the evolution of fairness); Sarah F. Brosnan, *Fairness in Monkeys*, *in* ENCYCLOPEDIA OF ANIMAL BEHAVIOR 288, 288-89 (Marc Bekoff ed., 2004); Brosnan, *supra* note 83, at 160-61; Brosnan & de Waal, Reply, *supra* note 84; Frans B.M. de Waal, *The Chimpanzee's Sense of Social Regularity and its Relation to the Human Sense of Justice*, 34 AM. BEHAV. SCIENTIST 335, 345-49 (1991).

we now know that specific brain regions perform certain kinds of reasoning tasks, and moral judgment, in particular, has been shown to involve specific brain regions that interact differently depending on the exact nature of the moral decision presented.

Contrary to prior assumptions,[110] we also know today that the brain is anatomically and functionally specialized. Different parts of the brain acting in varying combinations perform different information-processing tasks that in turn influence a person's behavior. Scientists have investigated the functions of different parts of the human brain by observing what happens when different brain locations are experimentally stimulated, different regions of the brain are removed or damaged, or a person within a brain scanning device is engaged in decisionmaking.

A wealth of experimental evidence demonstrates that artificial stimulation of different areas of the brain affects perception, cognition, and behavior differently. Using electrical stimulation in non-human primates, for example, researchers have manipulated the activity of small groups of neurons, ultimately establishing causal links between the activities of those cells and specific aspects of perception or cognition.[111] And with non-invasive stimulation studies on humans (such as transcranial magnetic stimulation), researchers have discovered predictable interference with different cognitive tasks when the activity of a specific portion of the brain is disrupted artificially.[112]

With respect to discoveries arising from cases of brain damage, consider the well-studied case of Phineas Gage, a reliable and personable railroad worker of normal social disposition, who had a metal tamping rod pass cleanly through his skull and brain after a spark ignited the dynamite he was tamping into its hole.[113] The rod removed Gage's medial prefrontal cortex and, although he appeared to retain all of his essential mental faculties, he became bizarrely anti-social.[114] Similarly, a variety of recent studies have identified ways in

110. *See generally* PINKER, *supra* note 24, *passim* (discussing such assumptions). For more on functional specialization, see STEVEN PINKER, HOW THE MIND WORKS (1999).

111. Marlene R. Cohen & William T. Newsome, *What Electrical Microstimulation has Revealed about the Neural Basis of Cognition*, 14 CURRENT OPINION NEUROBIOLOGY 169, 169-75 (2004) (providing overview of studies).

112. Alvaro Pascual-Leone et al., *Transcranial Magnetic Stimulation in Cognitive Neuroscience—Virtual Lesion, Chronometry, and Functional Connectivity*, 10 CURRENT OPINION NEUROBIOLOGY 232 (2000).

113. ANTONIO DAMASIO, DESCARTES' ERROR: EMOTION, REASON, AND THE HUMAN BRAIN 8-10 (2005) (describing Gage's injury and its effect on his social behavior).

114. *Id.*; Hanna Damasio, et al., *The Return of Phineas Gage: Clues about the Brain from the Skull of a Famous Patient*, 264 SCIENCE 1102 (1994).

which relatively localized and discrete brain damage and lesions in particular brain areas can result in severe impingement on the acquisition of social knowledge, on social behavior, and on moral reasoning.[115] For example, damage to the ventromedial prefrontal cortex, particularly of a young child, renders that person morally incompetent.[116] And a condition known as "acquired sociopathy" often results from damage to the orbitomedial or polar frontal cortex, the anterior temporal lobe, or the superomedial frontal lobe.[117]

Even more significant are the results of brain imaging studies. Functional magnetic resonance imaging ("fMRI"), which tracks both locations and amplitudes of activity in the brain,[118] has shown that specific parts of the brain—the orbital and medial sectors of the prefrontal cortex, as well as the superior temporal sulcus, among others[119]—are involved in moral reasoning and also show how the subareas work together in predictable ways during the process of making a moral decision.[120]

Not only does this work reveal the underappreciated extent to which processes of evolution and development have yielded highly specialized brain operations, evolutionary premises explicitly underlie the hypotheses of some of the most significant studies of moral decisionmaking. For example, Joshua Greene and colleagues started from the premise that elements of basic emotions (such as fear,

---

115. Joshua Greene, *Cognitive Neuroscience and the Structure of the Moral Mind*, *in* THE INNATE MIND: STRUCTURE AND CONTENTS 338, 339-41 (Peter Carruthers et al. eds., 2005).

116. DAMASIO, *supra* note 113, at 61 (describing defect in the ability to reason about choices and to make socially and ethically appropriate ones based on that reasoning).

117. Jorge Moll et al., *Morals and the Human Brain: A Working Model*, 14 NEUROREPORT 299, 300 (2003). Other implicated areas include "certain related subcortical neuclei, particularly the amygdala, ventromedial hypothalamus, dorsomedial thalamus, and head of the caudate nucleus or anterior limb of internal capsule." *Id.*

118. Useful introductions include: SCOTT A. HUETTEL ET AL., FUNCTIONAL MAGNETIC RESONANCE IMAGING (2004); MICHAEL I. POSNER & MARCUS E. RAICHLE, IMAGES OF MIND (1997).

119. *See, e.g.*, Jorge Moll et al., *The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions*, 22 J. NEUROSCIENCE 2730 (2002).

120. *See* Greene, *supra* note 115 (discussing "moral centers" of the brain and how they work together); Qian Luo et al., *The Neural Basis of Implicit Moral Attitude—An IAT Study Using Event-Related fMRI*, 30 NEUROIMAGE 1449, 1455-57 (2005) (providing useful bibliography); Moll et al., *supra* note 117. Other important works include Oliver R. Goodenough, *Mapping Cortical Areas Associated with Legal Reasoning and Moral Intuition*, 41 JURIMETRICS J. 429 (2001); Oliver R. Goodenough & Kristin Prehn, *A Neuroscientific Approach to Normative Judgment in Law and Science*, 359 PHIL. TRANSACTIONS: BIOLOGICAL SCI. 1709 (2004); Joshua Greene & Jonathan Haidt, *How (and Where) Does Moral Judgment Work?*, 6 TRENDS COGNITIVE SCI. 517 (2002); Joshua D. Greene et al., *An fMRI Investigation of Emotional Engagement in Moral Judgment*, 293 SCIENCE 2105 (2001); Joshua D. Greene et al., *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*, 44 NEURON 389 (2004); Hauke R. Heekeren et al., *An fMRI Study of Simple Ethical Decision-making*, 14 NEUROREPORT 1215 (2003).

jealousy, and anger) are evolved short cuts to behaviors that were adaptive for their bearers, on average, in deep ancestral environments. That is, compared to more cumbersome analysis and deliberation, emotions are generally "reliable, quick, and efficient" ways of solving commonly encountered circumstances.[121]

These researchers therefore predicted that the more closely a moral scenario aligns with regular features of ancestral environments, the more likely it is to invoke evolved emotional reactions, rather than dispassionately cognitive and deliberative ones. And the results were consistent with those predictions. Specifically, more physically personal dilemmas, such as whether to *push* one person into harm's way, thereby deflecting that harm from several people, yields greater activity in the brain areas associated with emotions, while a less personal but analogous dilemma, such as whether to *throw a switch* that would deflect an oncoming harm onto one person instead of several, yields greater activity in a different brain area, associated with less passionate analysis and calculation.[122]

Of course, such evidence does not suggest that there is a single, evolved, discretely boundaried "morality module" in the brain.[123] The situation is more complex. Some of the regions associated with moral decisionmaking are also associated with other brain activities. But what is clear is that an identifiable subset of the brain's regions, despite anatomical dissociability, are regularly and intimately networked in assessing moral dilemmas. And the interaction of these regions changes predictably as the nature of the moral dilemma changes.[124]

In another study, researchers found that people make faster decisions, and with less activity in the deliberating temporal poles,

---

121. Joshua Greene, *The Secret Joke of Kant's Soul*, *in* THE NEUROSCIENCE OF MORALITY: EMOTION, BRAIN DISORDERS, AND DEVELOPMENT (Walter Sinnott-Armstrong ed., forthcoming Jan. 2008) (manuscript at 33 *available at* http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-KantSoul.pdf.).

122. Specifically, the more personal dilemma generated comparatively greater activity in the posterior cingulate cortex, the medial prefrontal cortex, and the superior temporal sulcus. Greene, *supra* note 115, at 346.

123. *See, e.g.*, Jana Schaich Borg et al., *Consequenees, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation*, 18 J. COGNITIVE NEUROSCIENCE 803, 816 (2006) ("Our data highlight that morality is not represented in one place in the brain, but instead is mediated by multiple networks."); Greene, *supra* note 115, at 349 ("[I]t is clear from these studies that there is no 'moral center' in the brain, no 'morality module.' "); Greene & Haidt, *supra* note 120, at 522 ("What is becoming increasingly clear, however, is that there is no specifically moral part of the brain.").

124. *See* sources cited *supra* note 123.

when moral scenarios involve bodily harm than when they do not.[125] Another study showed significantly greater activity in the right amygdala, left medial orbitofrontal cortex, and medial frontal gyrus when subjects were shown interpersonal violence than when shown vandalism or violence to objects.[126] Compared to scenarios with unintentional harm, scenarios involving intentional harm yield greater activity in areas associated with emotion (such as the orbitofrontal cortex and the temporal pole) with correspondingly less activity in areas associated with deliberative cognition (such as the angular gyrus and superior frontal gyrus).[127] Moral scenarios in which action and inaction result in the same amount of harm yield greater activity in areas associated with deliberative cognition (such as the dorsolateral prefrontal cortex) and correspondingly less activity in areas associated with emotion (including the orbitofrontal cortex and temporal pole) than do analogous nonmoral scenarios. This suggests that "[i]ndividuals will utilize varying combinations of cognitive and emotive facilities to address moral challenges, but, overall, certain types of moral scenarios are likely to be processed in characteristic ways."[128] Further, removing emotive components of decisionmaking does not render people hyper-ethical, it instead renders them "unable to feel the rightness and wrongness of simple decisions and judgments."[129]

This area of research is relatively new. Nevertheless, the foregoing, in conjunction with other studies, has led a variety of researchers to conclude that our moral intuitions are built on evolved and widely shared building blocks.[130] Marc Hauser, for example, concludes that the evidence points to "all humans [being] endowed with a moral faculty."[131] Joshua Greene argues that the evidence suggests that "the form of human moral thought is importantly

125. Hauke R. Heekeren et al., *Influence of Bodily Harm on Neural Correlates of Semantic and Moral Decision-Making*, 24 NEUROIMAGE 887 (2005).

126. Borg et al., *supra* note 123, at 808-11; Luo et al., *supra* note 120, at 1454.

127. Borg et al., *supra* note 123, at 803.

128. *Id.* at 815-16.

129. Jonathan Haidt & Fredrik Bjorklund, *Social Intuitionists Answer Six Questions about Moral Psychology*, *in* 2 MORAL PSYCHOLOGY, THE COGNITIVE SCIENCE OF MORALITY: INTUITION AND DIVERSITY (Walter Sinnott-Armstrong ed., forthcoming Jan. 2008) (manuscript at 17).

130. *See, e.g.*, Greene, *supra* note 115, at 338-39 (discussing studies that help to reveal the foundations of morality). For an overview of recent work on the biology of punishment, see Ben Seymour, Tania Singer & Ray Dolan, *The Neurobiology of Punishment*, 8 NATURE REVIEWS NEUROSCIENCE 300 (2007).

131. Marc Hauser et al., *Reviving Rawls' Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions*, *in* 1 MORAL PSYCHOLOGY, THE EVOLUTION OF MORALITY: ADAPTATIONS AND INNATENESS (Walter Sinnott-Armstrong ed., forthcoming Jan. 2008) (manuscript at 1).

shaped by the innate structure of the human mind and that some basic, prosocial tendencies probably provide human morality with innate content."[132] And both argue that, while moral decisionmaking is networked, rather than unitary, it is networked in a way distinctly analogous to the brain's innate grammar (which is the basic grammar design one is born with and which underlies all human languages despite regional cultural variation).[133] That is to say, these authors and others conclude that the basic moral sentiments humans share are products of evolutionary processes.[134]

### *C. Child Development Evidence*

The child development literature provides much collateral support for the evolutionary theory offered in Part II. First, there is evidence that children's intuitions of justice progress through consistent stages of development across cultures. In the same way that baby teeth grow from gums and adult teeth replace baby teeth, intuitions about morality and justice seem to develop according to a relatively predictable sequence. Second, the cognitive skills that underpin these intuitions are precocious, with children able to make important conceptual distinctions relevant to morality and justice at young ages. Finally, the content of the intuitions of justice tracks the core intuitions discussed above, with injury, theft, and fairness being among the first principles of justice understood by young children.

#### 1. Predictable Stages of Development

There is a great deal of data about the development of moral reasoning in children. The classic—if somewhat dated—body of research in the area is Lawrence Kohlberg's work,[135] recently described as "the most widely researched description of moral development available."[136] Kohlberg's view is that people go through a

132. Greene, *supra* note 115, at 351.

133. *Id.* at 350-51; Hauser et al., *supra* note 131; *see* PINKER, *supra* note 34.

134. For example, Haidt and Bjorklund argue that "moral beliefs and motivations come from a small set of intuitions that evolution has prepared the human brain to develop . . . ." Haidt & Bjorklund, *supra* note 129 (manuscript at 2). Greene & Haidt, *supra* note 120, at 517, conclude that moral intuitions "are shaped by natural selection, as well as by cultural forces."

135. *See, e.g.*, 2 LAWRENCE KOHLBERG, ESSAYS ON MORAL DEVELOPMENT: THE PSYCHOLOGY OF MORAL DEVELOPMENT (1984); Lawrence Kohlberg, *From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away With It in the Study of Moral Development*, *in* COGNITIVE DEVELOPMENT AND EPISTEMOLOGY (Theodore Mischel ed., 1971); *see also* 1 ANNE COLBY & LAWRENCE KOHLBERG, THE MEASUREMENT OF MORAL JUDGMENT (1987); 2 ANNE COLBY & LAWRENCE KOHLBERG, THE MEASURE OF MORAL JUDGMENT (1987).

136. LARRY P. NUCCI, EDUCATION IN THE MORAL DOMAIN 81 (2001).

predictable sequence of phases of moral thinking.[137] As Jerome Kagan suggests, one would expect this if the capacity for learning moral reasoning were like the capacity for learning language.[138] That is, reliable, consistent, cross-situational development is a hallmark of the existence of a specific set of psychological mechanisms designed for a particular function by natural selection.[139]

For our purposes, it is not important whether Kohlberg's scheme is the best way of describing the various stages of

137. This view parallels Piaget's views on development. *See generally* 1 LAWRENCE KOHLBERG, ESSAYS ON MORAL DEVELOPMENT: THE PHILOSOPHY OF MORAL DEVELOPMENT (1981) (linking cognitive and moral development).

138. Jerome Kagan, *Introduction* to THE EMERGENCE OF MORALITY IN YOUNG CHILDREN ix, ix-x (Jerome Kagan & Sharon Lamb eds., 1987).

139. Kohlberg helieved that as children develop more sophisticated ways of thinking (for example, from the concrete to the abstract, see BÄRBEL INHELDER & JEAN PIAGET, THE GROWTH OF LOGICAL THINKING FROM CHILDHOOD TO ADOLESCENCE xxiii-xxiv (1958), which describes early work of Inhelder and Piaget studying the development of intelligence in humans), they also develop more sophisticated moral conceptualizations. Kohlberg's sequence is divided into three "levels" each with two "stages." His view might be summarized as:

| Level (age range) | Stage | Description of Moral Behavior |
|---|---|---|
| Pre-conventional (5-10) | Obedience/ Punishment | Conform to norms because of potential for punishment by authority figures. |
| | Individualism/ Exchange | Do what is "good" (i.c., feels good) for the self. This includes beneficial exchanges with others. |
| Conventional (10-14) | Good/Bad | Behave so as to elicit approval from others. |
| | Law and Order | Behave so as to discharge one's duty as part of the social order. |
| Post-Conventional (14-adult) | Social Contract | Acting in a way that benefits others because of rationally-based laws/norms in a society. Individual rights and utilitarianism. |
| | "Universal ethical principle orientation" | "Right is defined by tbe decision of conscience in accord witb self-chosen ethical principles appealing to logical comprehensiveness, universality, and consistency." |

(Quotations from KOHLBERG, *supra* note 137, at 16-19, 24-25, 409-12, are used for the stage name and description because it is difficult to gloss these ideas.) Kohlberg argued that people universally progressed from lower to higher stages of moral development. He did not argue that everyone reached the highest stage, "universal ethical principle orientation." The early stages, the "pre-conventional" stages, commonly from ages five through ten, are essentially about self-interest. The intermediate "conventional" levels, commonly ages ten through fourteen, in contrast, reflect an awareness of the benefits of having a positive reputation as a moral agent and of fulfilling one's duties in the context of social exchange. The highest level, the "post-conventional" stages, commonly from ages fourteen onward, includes a genuine interest in others' welfare, a respect for others' rights, and a recognition of universal moral principles. More recent approaches to Kohlberg's model maintain the idea that stages are reacbed in sequence, but instead of one stage "replacing" another, each stage is seen as supplementing the logic of previous stages. Dennis L. Krebs & Kathy Denton, *Toward a More Pragmatic Approach to Morality: A Critical Evaluation of Kohlberg's Model*, 112 PSYCHOL. REV. 629, 633 (2005). On this view, a person who has reached Stage 4 will still use Stage 3 reasoning in circumstances where doing so is useful or desirable. *Id.* at 645.

development. What is significant is the evidence, particularly from recent research (see below), of a predictable developmental path for all humans, however that path might be conceptualized and described. To the extent that children everywhere progress through similar stages of moral reasoning about justice at roughly the same ages, our evolutionary explanation for the origins of intuitions of justice receives collateral support.[140] As suggested by Kagan, "temporal concordance implies a biologically based preparedness to judge acts as right or wrong, where preparedness is used with the same sense intended by linguists who claim that two-year-old children are prepared to speak their language."[141]

Imagine the reverse case. 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 within cultures, such a general learning system would yield very different paths and timing in the acquisition of intuitions of justice for different individuals.[142]

### 2. Making Subtle and Complex Judgments at an Early Age

Early research in this area seems to have vastly underestimated the sophistication of children. More recent research demonstrates that children are able to make far more sophisticated judgments and much finer distinctions than those predicted by Kohlberg's proposed sequence. Although evidence now suggests that moral reasoning develops relatively early, it is likely that research still does not fully reveal the precociousness of moral reasoning.[143] John Darley and Thomas Shultz suggest in their broad review that "children are capable of making moral judgments at a much earlier

---

140. The first evolutionary perspective on Kohlberg's work of which we are aware appears in ALEXANDER, *supra* note 61, at 131-39.

141. Kagan, *supra* note 138, at x.

142. For a fuller discussion of general social learning as an alterative explanation, see *infra* Part IV.

143. *See* Lawrence Kohlberg, *A Current Statement on Some Theoretical Issues*, *in* LAWRENCE KOHLBERG: CONSENSUS AND CONTROVERSY 485, 491 (Sohan Modgil & Celia Modgil eds., 1986); *see also* John M. Darley & Thomas R. Shultz, *Moral Rules: Their Content and Acquisition*, 41 ANN. REV. PSYCHOL. 525, 537 (1990).

age than previously thought."[144] Summarizing recent literature, the authors conclude that "moral capacity is well developed although by no means completely developed in the third year of life."[145] The precocious abilities of young children support the evolutionary argument in Part II. To the extent that very young children have intuitions, acquire knowledge, and make conceptual distinctions, especially universally, the probability that each child acquires these by general learning processes decreases, and a more innate developmental sequence becomes more likely.[146]

*a. Distinguishing Moral, Conventional, and Prudential Rules*

In one early and crucial experiment, Judith Smetana tested very young children's beliefs about justice to determine if they make distinctions between violations of moral rules—acts that are wrong and deserve punishment (for example, one child hitting another)—and violations of conventional rules—acts that deviate from a convention (for example, failing to say grace before a snack).[147] Smetana used pictures indicating the acts to demonstrate violations. To elicit responses, a pictorial scale (different-size frowns) was used to gauge seriousness. Smetana also used a verbal assessment of how harshly the offender should be punished: not at all, a little, or a lot. One group of subjects consisted of children between two-and-a-half to roughly three-and-a-half years old. The other group consisted of children between roughly three-and-a-half and almost five years old. Moral offenses included both physical harm (hitting) and theft (taking someone else's apple). Both groups indicated that the moral transgressions were more serious and deserved more punishment than the conventional transgressions. The children indicated that the moral offenses, compared with the violations of convention, would be wrong even if "there [were] no rule about it."[148]

These results suggest that even very young children distinguish moral wrongs from violations of convention and apply a different logic to them. Moreover, children seem to consider moral

144. Darley & Shultz, *supra* note 143, at 552.

145. *Id.*

146. We take "innate" ideas to be those ideas that develop reliably in each member of the species given the broad range of plausible environments in which the organism develops. That is, an idea is innate if it does not rely on very general processes such as induction or social transmission for its acquisition.

147. *See* Judith G. Smetana, *Preschool Children's Conceptions of Moral and Social Rules*, 52 CHILD DEV. 1333, 1333-34 (1981).

148. *Id.* at 1334-35.

offenses to be serious: the majority of subjects in both groups rated the moral offenses as a four on the four-point scale. When tested two to three weeks later, subjects responded much as they had earlier, suggesting a consistency in judgments.[149]

The importance of these results should not be underestimated. The fact that such young children believe immoral acts are wrong, even in the absence of rules, indicates a precocious, universalist view of morality.[150] In sharp contrast, children believe that conventions apply to a particular group at a particular time and can be changed. According to Judith Dunn, the evidence suggests that "by three and four years old, children respond differently to different kinds of rule breaking or transgression" and that "[a]lthough the data do not show unambiguously a clear distinction between transgressions of social convention and of moral standards, they demonstrate that three-year-olds are sensitive to different kinds of cultural breach."[151] Other scholars take a stronger view:

> [C]hildren at early ages . . . judge moral issues to be obligatory, applicable across like situations, not contingent on specific social rules or authority-dictates, and not alterable on an arbitrary basis. They judge conventions as contingent on social organization—such as rules, authority, and existing arrangements.[152]

Somewhat older children similarly distinguish between moral rules and "prudential rules"—rules that are in place to protect people from harm (for example, a rule prohibiting climbing on the back of a sofa). Marie Tisak and Elliot Turiel gave children (average ages of the groups were roughly seven, nine, and eleven) stories about acts that violated either moral rules (regarding theft or hitting) or prudential rules (regarding running and falling).[153] Subjects reported that the violation of the moral rule was more wrong and that it would be more

---

149. Test-retest reliability was .66 for the fourteen subjects so tested, a reasonable number by traditional standards in experimental developmental psychology. *Id.* at 1334.

150. *See* Judith G. Smetana et al., *Preschool Children's Judgments about Hypothetical and Actual Transgressions*, 64 CHILD DEV. 202, 211 (1993) (reporting on a study that found preschool children differentiate between moral and conventional transgressions, finding moral transgressions more generally wrong regardless of the existence of rules).

151. Judith Dunn, *The Beginnings of Moral Understanding: Development in the Second Year*, *in* THE EMERGENCE OF MORALITY IN YOUNG CHILDREN, *supra* note 138, at 91, 93.

152. Elliot Turiel, Melanie Killen & Charles C. Helwig, *Morality: Its Structure, Functions, and Vagaries*, *in* THE EMERGENCE OF MORALITY IN YOUNG CHILDREN, *supra* note 138, at 155, 170; *see also* Jonathan Haidt et al., *Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?*, 65 J. PERSONALITY & SOC. PSYCHOL. 613, 621 (1993) (observing that children ten to twelve years old across the cultures investigated think that pushing another child off a swing should be punished and that this behavior would be wrong in other countries as well).

153. Marie S. Tisak & Elliot Turiel, *Children's Conceptions of Moral and Prudential Rules*, 55 CHILD DEV. 1030, 1031 (1984).

wrong to change moral rules.[154] In other words, it is more acceptable for the prudential rule to be modified. The comparison between moral rules and prudential rules indicates judgments that go beyond consideration of consequences of rule-violating actions. Again, the results of Tisak and Turiel suggest the development of nuanced views at early ages.

*b. Judging the Relative Seriousness of Wrongful Conduct*

Studies suggest that even young children intuitively appreciate the relative seriousness of different kinds of wrongful conduct and that the sophistication of the distinctions they make increases with age. One of the first principles of morality that children acquire is the idea that it is wrong to hurt others.[155] By age three, they understand this principle,[156] and by age four, they take into account the foreseeability of harm.[157] As Philip Zalazo and his colleagues put it: "substantive concepts of harm and welfare, however acquired, are guiding moral judgments by the fourth year of life."[158]

Findings also indicate that "children consider moral transgressions resulting in physical harm to be more wrong than moral transgressions resulting in property violations."[159] A study by Tisak and Turiel indicates that six-year-old children judge physical violence as more serious than theft.[160] This suggests that children have complex intuitions across different domains. Additional evidence comes from studies in which children are asked to give examples of moral transgressions. Children give physical acts of harm as the most common examples.[161] Acts of physical aggression are prototypical moral violations to children.[162]

---

154. *Id.* at 1028.

155. NUCCI, *supra* note 136, at 86; *see also* Kagan, *supra* note 138, at ix, xi.

156. Phillip David Zelazo et al., *Intention, Act, and Outcome in Behavioral Prediction and Moral Judgment*, 67 CHILD DEV. 2478, 2479 (1996).

157. Sharon A. Nelson-LeGall, *Motive-Outcome Matching and Outcome Foreseeability: Effects on Attribution of Intentionality and Moral Judgments*, 21 DEVELOPMENTAL PSYCHOL. 332, 336 (1985).

158. Zelazo et al., *supra* note 156, at 2488.

159. Marie S. Tisak et al., *Presehool Children's Social Interactions Involving Moral and Prudential Transgressions: An Observational Study*, 7 EARLY EDUC. & DEV. 137, 139 (1996).

160. Tisak & Turiel, *supra* note 153, at 1036.

161. Marie S. Tisak & Jeanne H. Block, *Preschool Children's Evolving Conceptions of Badness: A Longitudinal Study*, 1 EARLY EDUC. AND DEV. 300, 300 (1990).

162. *Id.* at 305.

Research on morality has shown that people cannot always articulate the basis for their moral judgments.[163] Even adults are, broadly, frequently unable to justify why they chose their actions[164] and it would be reasonable to expect children to have similar limitations. Nonetheless, research shows that by age seven, children can elaborate ideas about injury fairly well, and, when children make judgments about blame and punishment, their ideas include a calculation as to the *intention* of the person inflicting the injury.[165]

This conclusion—that the first moral concept to appear in children is that physical aggression is wrong—is relevant to our argument. It is likely more than coincidence that this is also the first step in our account of the evolutionary origins of intuitive justice.

In contrast, very young children might *not* have a firm grasp on notions of fairness in exchange or social contracts. Young children frequently announce that some outcome is "unfair." However, this often means that the outcome is unfavorable to the speaker, rather than violating a social contract or norm.[166] Nonetheless, William Damon has suggested that young children can show "consistent, patterned reasoning" about fair distribution of property, ownership, and the like.[167]

Scholars have argued that it is not until later in the development of a child, perhaps around nine years of age, that adult-like conceptions of fairness emerge.[168] Indeed, Larry Nucci asserts that "[t]he great accomplishment of early-childhood moral development is the construction of moral action tied to structures of 'just' reciprocity."[169] Tangentially relevant is research showing that young children (age five or less) allocate rewards equally, whereas older children tend to allocate rewards with a proportionality (equity)

---

163. Haidt, *The Emotional Dog and Its Rational Tail*, *supra* note 42.

164. *See* WILLIAM HIRNSTEIN, BRAIN FICTION: SELF-DECEPTION AND THE RIDDLE OF CONFABULATION (2005); Nisbett & Wilson, *supra* note 163 (concluding that subjects may not be able to evaluate the mental processes underlying a response and instead rely on prior perceptions of what would be an appropriate explanation).

165. *See* Dale T. Miller & C. Douglas McCann, *Children's Reactions to the Perpetrators and Victims of Injustices*, 50 CHILD DEV. 861, 866 (1979).

166. NUCCI, *supra* note 136, at 86.

167. William Damon, *Early Conceptions of Positive Justice as Related to the Development of Logical Operations*, 46 CHILD DEV. 301, 302 (1975).

168. Larry Nucci, *Because It Is the Right Thing to Do*, 45 HUMAN DEV. 125, 128 (2002). Nucci claims that at around age six, "children's moral judgments become regulated by conceptions of just reciprocity." *Id.* When fairness and reciprocity emerge can be debated, but it looks as though it is after intuitions regarding harm emerge.

169. NUCCI, *supra* note 136, at 87.

rule,[170] suggesting a developmental trend in issues associated with exchange (effort for reward).

The integration of multiple factors relevant to estimation of seriousness also begins early, perhaps around age seven. In a study performed by David Elkind and Ruth Dabek, children were divided into groups based on age (average ages of the groups were roughly five-and-a-half, seven-and-a-half, and nine).[171] The children listened to stories about different crimes. The offense described in each story varied in terms of whether it was intentional or not and whether the damage was to a person or to property. The children then assessed blame and were asked about punishment. On average, the children viewed damage to a person as more serious, and they judged intentional damage as more serious than non-intentional damage (though this varied by age). Children of all ages viewed intentional personal injury as more serious than unintentional property damage. Interestingly, the youngest group of children (average age roughly five-and-a-half) thought that unintentional personal injury was more serious than intentional property damage, suggesting the importance of personal injury.[172] This pattern was not found in the older children. These results are mirrored in later work that suggests that while very young children focus on either intention *or* harm, older children (age four and five) use both harm *and* intention when making decisions about punishment.[173] Such conclusions strongly imply that children have sophisticated views on desert and possess the ability to weigh multiple factors by age seven.

Charles Helwig and Urszula Jasiobedzka obtained similar results by having children aged six, eight, and ten evaluate laws that were either socially beneficial (such as traffic laws) or unjust (such as age discrimination).[174] They found that all children, regardless of age, considered "the perceived justice of the law, its social beneficial purpose, and its potential for infringement on individual freedoms and

---

170. *See* Jay Hook, *The Development of Equity and Logico-Mathematical Thinking*, 49 CHILD DEV. 1035, 1041 (1978); J.G. Hook & Thomas D. Cook, *Equity Theory and the Cognitive Ability of Children*, 86 PSYCHOL. BULL. 429, 441 (1979).

171. David Elkind & Ruth F. Dabek, *Personal Injury and Property Damage in the Moral Judgments of Children*, 48 CHILD DEV. 518, 519 (1977).

172. *Id.* at 521. There is evidence that even three-year-olds take intentions into account when making judgments about actors. *See* Sharon A. Nelson, *Factors Influencing Young Children's Use of Motives and Outcomes as Moral Criteria*, 51 CHILD DEV. 823, 828-29 (1980).

173. Zalazo et al., *supra* note 156, at 2478-92.

174. *See* Charles C. Helwig & Urszula Jasiobedzka, *The Relation Between Law and Morality: Children's Reasoning about Socially Beneficial and Unjust Laws*, 72 CHILD DEV. 1382, 1382 (2001).

rights."[175] Helwig and Jasiobedzka concluded that "children apply moral concepts of harm, rights, and justice to evaluate laws."[176] These findings imply that children develop textured views of legal issues by around age six.

*c. Judging Blameworthiness with Factors beyond Offense Seriousness*

Children's intuitions of justice are sophisticated enough to include more than just an assessment of the seriousness of the wrongful conduct. For example, they will take account of a person's culpable state of mind—whether the violation was intentional or accidental—as noted above. They also consider various exculpating or mitigating circumstances. For example, the punishment judged to be appropriate will change depending on the availability of information regarding mitigating circumstances (i.e., the offender had a head injury), an effect seen in experiments with children as young as seven years old.[177] Summarizing their findings, Darley and Shultz conclude:

> [R]esults indicated a fairly sophisticated use of a variety of the moral concepts by children from 5 years of age. Children revealed evidence of knowing that judgments of punishment presupposed judgments of moral responsibility and that moral responsibility judgments presuppose causal judgments. They also used information on intention and negligence to assign moral responsibility and information on restitution to assign punishment.[178]

More recently, research on children's comprehension of provoked harm versus unprovoked harm has revealed similar nuances. Studies by Judith Smetana, Nicole Campione-Barr, and Nicole Yell required children between the ages of six and nine to examine pictures that showed provoked and unprovoked transgressions involving either physical injury (hitting) or psychological harm (teasing).[179] The researchers concluded that "[c]hildren judged hypothetical moral transgressions to be more serious and more deserving of punishment, and they reasoned more

---

175. *Id.*

176. *Id.*

177. *See* John M. Darley et al., *Intentions and Their Contexts in the Moral Judgments of Children and Adults*, 49 CHILD DEV. 66, 66 (1978) (finding evidence in children as young as six that judgments regarding punishment vary depending on relevant circumstances, such as provocation); Adrian Furnham & Steven Jones, *Children's Views Regarding Possessions and Their Theft*, 16 J. MORAL EDUC. 18, 25-27 (1987); *see also* Cecilia Wainryb, *Understanding Differences in Moral Judgments: The Role of Informational Assumptions*, 62 CHILD DEV. 840, 847 (1991) (stating that new information can modify judgments of wrongness considerably among people aged eleven to twenty-one).

178. Darley & Shultz, *supra* note 143, at 535.

179. Judith G. Smetana et al., *Children's Moral and Affective Judgments Regarding Provocation and Retaliation*, 49 MERRILL-PALMER Q. 209, 216-17 (2003).

about concerns with others' welfare, for [unprovoked] than for provoked transgressions and when retaliation involved hitting rather than teasing."[180] The researchers also found that "all children judged that escalating the retaliatory response by hitting in response to being teased . . . was more serious and more deserving of punishment than teasing in retaliation for either teasing or hitting,"[181] suggesting that children's understanding of morality and punishment includes the view that physical injury is most egregious.

The sense of justice held by young children also incorporates a feature present in adult intuitions of justice: the effect of a mistaken belief of an offender. When children aged five and seven (but not three[182]) make judgments regarding blame, they take into account the fact that others might have incorrect beliefs.[183] Their judgments are quite nuanced. If a person's belief is different than the child's on matters of fact—that is, beliefs concerning what is true (as opposed to what is morally right)—then mitigation often is permitted.[184] However, if the different belief relates to what is right and wrong (for example, a teacher who thinks it is acceptable to discriminate against someone based on gender), then the person's mistake does not exculpate him.[185] This implies that children have a sophisticated understanding of others' beliefs and the role they play in the commission of moral offenses. Further research demonstrates similar results in older children.[186]

---

180. *Id.* at 209.

181. *Id.* at 223.

182. A vast research enterprise shows that children under four years of age cannot understand that others might have beliefs that are wrong or different from their own. That is one possible reason for the developmental difference, having nothing to do with a change in moral intuitions, but only more general abilities of understanding. *See generally* SIMON BARON-COHEN, MINDBLINDNESS: AN ESSAY ON AUTISM AND THEORY OF MIND (1995); Alan M. Leslie, *Pretending and Believing: Issues in the Theory of ToMM*, 50 COGNITION 211 (1994); Alan M. Leslie, *Pretense and Representation: The Origins of Theory of Mind*, 94 PSYCHOL. REV. 412 (1987).

183. Cecilia Wainryb & Sherrie Ford, *Young Children's Evaluations of Acts Based on Beliefs Different from Their Own*, 44 MERRILL-PALMER Q. 484, 484 (1998).

184. *Id.* at 90-92.

185. *Id.*

186. *See, e.g.*, Larry Nucci & Elsa K. Weber, *Social Interactions in the Home and the Development of Young Children's Concepts of the Personal*, 66 CHILD DEV. 1438, 1445 (1995) ("[C]hildren differentiate personal issues from matters of moral or conventional regulation . . . ."); Judith G. Smetana, *Toddlers' Social Interactions regarding Moral and Conventional Trasngressions*, 55 CHILD DEV. 1767, 1774 (1984) ("[T]oddlers initiate responses to moral transgressions . . . [and are] more likely to respond to moral than conventional transgressions with emotional reactions, physical retaliation, and increasingly with age, statements regarding the harm or injury caused.").

Smetana and her colleagues summarize this research and suggest that

> by 4 years of age, middle-class, primarily European American children, as well as African American preschoolers from lower socioeconomic backgrounds, reliably evaluate moral transgressions pertaining to physical or psychological harm and fairness as very serious, deserving of punishment, generalizably wrong, and wrong regardless of whether or not there is a rule or a teacher prohibits the act.[187]

In short, even young children seem to have textured and specific views regarding deserved punishment, and these views are not derived simply from the dictates of authority.[188]

### 3. Cross-Cultural Developmental Studies

Substantial research demonstrates that the developmental sequence for moral reasoning is not unique to the Western world. Nucci suggests that "there is considerable cross-cultural evidence that children and adults across a wide range of the world's cultures conceptualize prototypical moral issues pertaining to fairness and others' welfare in ways very similar to children and adults in Western contexts, and differentiate such issues from prototypical matters of convention."[189] In a recent review, Jenny Yau and Judith Smetana conclude that despite cultural differences, "[c]hildren as young as 3½ to 4 years of age have been found to treat moral transgressions as very serious, generalizably wrong, and wrong independent of rules and authority sanctions. In contrast, they treat conventional transgressions as less serious, contextually relative, and contingent on rules and authorities."[190]

This last point, that children believe moral rules to be universally applicable, bears expansion. As discussed above,[191] children distinguish between fundamental issues of justice (e.g., the belief that wrongdoing should be punished, property should not be forcibly taken, and so on) and issues of convention or collateral societal issues (e.g., objects over which individuals can be said to have

---

187. Smetana et al., *supra* note 179, at 210 (internal citations omitted).

188. *See, e.g.*, Marsha D. Walton & Andrea J. Sedlak, *Making Amends: A Grammar-Based Analysis of Children's Social Interaction*, 28 MERRILL-PALMER Q. 389 (1982) (suggesting sophisticated understanding of transgressions among five-to-ten year olds in classroom settings, including implicit knowledge of what others will think relevant to the way in which a transgression is construed).

189. NUCCI, *supra* note 136, at 95-96.

190. Jenny Yau & Judith G. Smetana, *Conceptions of Moral, Social-Conventional, and Personal Events Among Chinese Preschoolers in Hong Kong*, 74 CHILD DEV. 647, 647 (2003) (internal citations omitted).

191. *See, e.g.*, supra notes 147-58 and accompanying text.

a property right). Children acknowledge that conventions could be other than as they are and that if another group of people had a different convention, it would not be "wrong."[192]

The capacity of children to make this distinction is relevant to the frequently cited argument that cultures differ in their morality, found in Richard Schweder et al.'s cross-cultural investigation of the relative rankings of wrongful acts as judged by Hindu Brahman children.[193] Of the thirty-nine acts listed, the children judged the following as the most serious wrong on the list: "[t]he day after his father's death, the eldest son had a haircut and ate chicken."[194] They judged as one of the least serious transgressions—ranked thirty-fifth of the thirty-nine—the case of a husband who beats a wife because she repeatedly goes alone to the movies. These results may appear to conflict with the notion of shared intuitions of justice; Western children would make no such ranking. However, the rankings must be understood in the context of the underlying beliefs held by the raters. Because the father's soul is put in jeopardy by the son's act, the son's act is judged as a most serious harm. Similarly, the raters' understanding that the wife's act is a serious breach of a clear contractual and social obligation helps explain why the husband's beating of her is not ranked as more wrongful. In short, while underlying intuitions regarding principles of justice may be similar, different social conventions create different perceptions about the wrongfulness of specific conduct.[195] As Cecelia Wainryb explains, "what appears to be moral variation may be due not to diversity in ethical concepts but rather to differences in informational assumptions."[196] This point is important when judging the locus of genuine variability in beliefs surrounding intuitions of justice.

---

192. *Cf.* Leigh A. Shaw & Cecilia Wainryb, *The Outsider's Perspective: Young Adults' Judgments of Social Practices of Other Cultures*, 17 BRIT. J. DEV. PSYCHOL. 451 (1999) (reaching similar conclusions with a Chinese sample).

193. Richard A. Schweder et al., *Culture and Moral Development*, *in* THE EMERGENCE OF MORALITY IN YOUNG CHILDREN, *supra* note 138, at 1, 36-71.

194. *Id.* at 40.

195. *See generally* Turiel, Killen & Helwig, *supra* note 152, *passim*; Wainryb, *supra* note 177, *passim*.

196. Wainryh, *supra* note 177, at 849. However, informational assumptions need not be determinative. *Id.*; *see also* Cecilia Wainryb, *Values and Truths: The Making and Judging of Moral Decisions*, *in* RIGHTS AND WRONGS: HOW CHILDREN EVALUATE THE WORLD, NEW DIRECTIONS FOR CHILD DEVELOPMENT 33, 33-46 (Marta Laupa ed., 2000) (arguing that the same underlying moral values in adults and children, or adults in different cultures, can lead to differing views on what is an immoral act because of different information about relevant facts).

### *D. Summary of Evidence*

Our hypothesis, that shared intuitions of justice derive in large measure from the relentless effects of evolutionary processes on human brains and consequent sentiments and behavioral predispositions, connects at a deep level with modern developments in biology and psychology. The hypothesis also appears to explain why these intuitions appear to be so stunningly consistent across our species, so subtle in their complexities, and so non-randomly focused on the harms to which their attention is particularly keen. We have not, in this Part, attempted a definitive proof of our hypothesis. Instead, we explored three different areas in which the data are consistent with our hypothesis: animal studies, brain science, and moral development.

The animal studies suggest that humans are not alone in punishing violations of norms of cooperation, reciprocity, and fairness. That so many social species exhibit these behaviors, and in such very specific ways, suggests that there is an evolutionary and adaptive root to these behavioral predispositions. This, in turn, suggests that similar human behavior may stem from the same root. The data further suggest that indifference to wrongdoing may in many species be evolutionarily unstable. And this, in turn, suggests that intuitions about punishment may frequently facilitate certain kinds of cooperation, such as social exchange. Such intuitions require an ability to perceive wrongful behavior, to remember transgressors, and to treat them differently than non-transgressors. We not only see this ability in many other social animals, but we also see it specifically in contexts of inflictions of physical harm, the taking of resources, and the violations of norms of reciprocity.

From the rapidly expanding field of brain science, we see how the combination of evolutionary processes and organismal development yields anatomically and functionally specialized human brains, which tend to function very similarly in individuals across the species. Damage to specific regions can result in dramatic social and moral behavioral aberrations, which suggests that those areas of the brain are sufficiently specialized to be necessary for social and moral activity. Imaging studies reveal widely shared patterns in neural activity, consistent with evolutionary hypotheses, when subjects are confronted with various moral scenarios. More personal and evolutionarily salient scenarios invoke various evolved behavior-biasing computations to a greater extent than do alternative scenarios. Far from responding with generic neural activity, human brains are intriguingly similar and differentially focused when

considering such things as bodily injury and violations of cooperation norms. This evidence suggests that human brains have evolved the specific capacity to process these kinds of transgressions in species-typical, characteristic ways.

In the child development literature, experimental evidence demonstrates that children understand that some acts are wrong and deserve punishment at a very early age. Indeed, the intuition appears to exist when children first are able to communicate these types of ideas. The notion of physical injury as a wrong is among the first elements to emerge in children, followed soon after by notions of fairness and reciprocity. These intuitions become increasingly textured as children grow older and change from simplistic notions of strict equality to more sophisticated notions such as equity, which take into account a person's capacities, contributions, and needs. Most strikingly, growing evidence suggests that the development of these rather sophisticated intuitions follow similar patterns cross-culturally, though of course there remains important debate about the extent of cross-cultural developmental similarities.[197]

While no single study or field of research conclusively proves our evolutionary hypothesis for the origins of shared intuitions of justice, the triangulation of the theoretical foundations from biology and psychology generally, alongside behavioral data in humans and other species, recent studies of human brain operations, and broad research into the characteristically human development of moral psychology, presents a strong case.

## IV. ALTERNATIVE EXPLANATIONS: GENERAL SOCIAL LEARNING AND EFFICIENT NORMS

Evolutionary explanations inevitably evoke the riposte that perhaps the phenomenon in question is simply due to "learning" or "culture." Could the consensus we have documented here be the result of social learning rather than evolution?[198] Perhaps the norms adopted

197. *See* Schweder et al., *supra* note 193.

198. Of course, any explanation for the very complex features of intuitions of justice must include evolution by natural selection as part of their origin, including explanations based upon social learning. The ability to learn is itself an evolved trait. *See* Tooby & Cosmides, *supra* note 29, at 119 (discussing the relationship between evolution and learning mechanisms). Thus, an argument that groups discover efficient norms regarding justice (see below) requires an evolutionary explanation for how the psychological mechanisms that perform tbis calculation—whether conscious or unconscious—came to exist. Similarly, an argument that people have a predisposition for social learning must also have an evolutionary explanation for this highly sophisticated, complex, functional capacity. In short, "learning" and/or "culture" do not obviate evolutionary explanations; they necessitate them.

simply reflect a human propensity to copy others' norms, hence the observed consensus without the need for our evolutionary explanation. Or, perhaps groups adopt the same norms because the norms are efficient for all groups.

In one sense, the social learning claim—that shared intuitions of justice are the product of social learning from the surrounding culture—does not conflict with our evolutionary argument in Part II because there we suggest that intuitions of justice are indeed "learned" in a sense. Babies do not come into the world with intuitions of justice intact; rather, they acquire them over time in a predictable and largely sequential manner through interaction with their environment.[199] In this sense, the evolutionary explanation envisions some involvement of the surrounding culture, just as the special human mechanism for acquiring language involves the surrounding culture.

But non-evolutionary explanations go a step further. They claim that intuitions of justice result *entirely* from general social learning with no reliance upon an evolutionarily developed special mechanism for acquiring intuitions of justice. In other words, the non-evolutionary explanations claim that people acquire views of justice in the same way that they learn most other knowledge or come to most other opinions.[200] That is, views of justice are learned, reasoned out, or chosen through rational decisionmaking.

### *A. Spontaneous Social Learning*

One non-evolutionary explanation, which might be called "spontaneous social learning," might proceed as follows: the views of justice on which there is so much agreement are those views that are

199. *See supra* Part III.C.1.

200. The distinction we draw here is between an evolved "general" social learning ability, which may result in, among many other things, the acquisition of intuitions of justice and an evolved mechanism "specific" to acquiring intuitions of justice. We take it as uncontroversial that a psychological system that acquires many different kinds of information from the social world—a general social learning system—is the product of evolution hy natural selection. But the narrow question is whether the system for acquiring intuitions of justice was designed by natural selection for that relatively narrow function, as we argue is the best-reasoned conclusion, or was designed to acquire information across a wide range of content domains. Evolutionary psychologists refer to this distinction as between "domain specific" and "domain general" mental systems. *See, e.g.*, Tooby & Cosmides, *supra* note 29, at 97 (argning that framing psychological issues in terms of how "domain specific" cognitive mechanisms are has been and will continue to be productive in psychological research). For a discussion of tests for domain-specificity in the context of logical reasoning, see Laurence Fiddick et al., *No Interpretation Without Representation: The Role of Domain-Specific Representations in the Wason Selection Task*, 77 COGNITION 1, 2 (2002).

most efficient for a group to hold, perhaps because they permit productive social cooperation, much like the advantages we discuss in the evolutionary account given in Part II. Notably, this explanation relies on no special mechanism for the acquisition of intuitions of justice. But this explanation for the shared views of justice faces several difficulties.

1. Preferring Group Interests Over Individual Interests

This explanation depends on the individual's choice to adopt views of justice that are efficient *for the group*. But there is little reason to believe that people are predisposed toward preferring norms that are efficient for the group, rather than norms that are efficient for their own self interest. To make this explanation plausible, one would have to show that people prefer what is good for the group over what is good for themselves, a preference that the available data do not support.[201] Instead, evidence from both the laboratory and the real world support the view that when group interests are pitted against self interest, what has been referred to as "social dilemmas," it is more likely that self interest will prevail.[202] Acts of littering, pollution, overfishing, and limitless other examples are all cases in which people choose to accommodate their own interests over those of the group.[203]

In the lab, the question of self-interest-versus-group-interest has been addressed by using "public goods" games or "commons dilemma" games. In these games, people are put into groups and given an allotment of money. Subjects can keep some or all of the money or put some or all of it in a group pot. Money in the group pot is multiplied by the experimenter at some rate and subsequently divided among all the members of the group. Because money in the pot is multiplied and subsequently split among the group members, everyone in the group is made best off if everyone contributes all their money to the group pot, making the "pie" everyone shares larger. However, each person is individually better off by keeping his own allotment of money for himself. With a few exceptions,[204] when people play these games in the same group over time, they keep greater and

201. DAWKINS, *supra* note 21 (arguing that genes, rather than whole organisms, are the entities on which natural selection acts and so, everything else equal, selection will act to favor genes that cause their own replication as opposed to genes that cause individuals or groups or ecologies to prosper).

202. Robyn Dawes, *Social Dilemmas*, 31 ANN. REV. PSYCHOL. 169 (1980).

203. SAMUEL S. KOMORITA & CRAIG D. PARKS, SOCIAL DILEMMAS (1996).

204. *See, e.g.*, Robert Kurzban et al., *Incremental Commitment and Reciprocity in a Real-Time Public Goods Game*, 27 PERSONALITY & SOC. PSYCHOL. BULL. 1662 (2001).

greater amounts of their endowment, reaching nearly purely selfish behavior over time.[205] In short, when individual selfishness is pitted against group interest in the lab, selfishness eventually wins.

### 2. Complexity of Determining Efficient Norms as Beyond Individual Capacity

Even if people prefer a norm that is efficient for the group, though not necessarily good for them personally, another challenge for the spontaneous social learning explanation is that determining an efficient norm is an exceedingly complex matter, probably well beyond the capacity of an individual. Studies of human cognitive skills confirm their limitations. Substantial experimental evidence suggests that when people are presented with tasks that require the use of a general ability to reason using the rules of formal logic, they typically do not perform well.[206] Even in experiments in which the financial stakes are very high, in which one might assume that people would have an incentive to reason as logically and carefully as possible, behavior deviates substantially from wealth-maximizing rationality.[207] As Herbert Simon put it: "[A]ctual human rationality-striving can at best be an extremely crude and simplified approximation to the kind of global rationality that is implied . . . by game theoretical models."[208]

These limitations are illustrated vividly in Jared Diamond's recent account of the failure of civilizations.[209] In these cases, the human inability to foresee the consequences of the way that societies organize themselves and to regulate the use of resources played crucial roles in their eventual downfall. Among the Mayans and the Easter Islanders, for example, the mismanagement of resources, due to the inability to foresee how cultural practices would affect long-term growth and stability, played a central role in the collapse of those civilizations.[210] This differs little from modern times, in which

---

205. John Ledyard, *Public Goods: A Survey of Experimental Research*, *in* HANDBOOK OF EXPERIMENTAL ECONOMICS 111-94 (J. Kagel & A. Roth eds., 1995).

206. Cosmides & Tooby, *supra* note 62, *passim* (arguing and providing evidence for the view that humans do not perform well on tasks that require a general ability to use the rules of formal logic).

207. For a brief discussion, see COLIN F. CAMERER, BEHAVIORAL GAME THEORY: EXPERIMENTS IN STRATEGIC INTERACTION 60-62 (2003) ("[M]ultiplying the stake by two or ten makes little difference . . . .").

208. Herbert Simon, *A Behavioral Model of Rational Choice*, 69 Q. J. ECON. 99, 101 (1955); *see also* GERD GIGERENZER & REINHARD SELTON, BOUNDED RATIONALITY, THE ADAPTIVE TOOLBOX (2002) (arguing that there are important constraints on human decision making, making them only "boundedly rational").

209. JARED DIAMOND, COLLAPSE: HOW SOCIETIES CHOOSE TO FAIL OR SUCCEED (2005).

210. *Id.* at 177.

economists reach little consensus on policies that will improve economic growth because of the complexity of the system in question and the limitation of human cognitive abilities to understand and predict complex social systems. Diamond's analysis in many ways continues analyses by economists such as Adam Smith and Friedrich Hayek,[211] who argued that neither the aggregate decisions of a democratic society nor the guiding hand of a dictator (for example, among the Mayans) is sufficient.[212]

In short, the spontaneous social learning explanation for the agreement on views of justice is inconsistent with what is known about the limits of human calculation, including the sharp limits on the human cognitive capacity for deduction, the common human tendency to rely on simplifying heuristics, and other limitations on cognition. In contrast, the evolutionary explanation for intuitions of justice that we offer in Part II involves a process that "tests" alternative intuitional systems (or any other features of an organism) through the sieve of natural selection. The feedback loop on genes allows only a narrow range of developmental programs to be selected. The spontaneous social learning explanation has no such power.

### *B. Accumulated Social Learning*

One might try to save a social learning explanation from this last flaw—that it is beyond the cognitive ability of a single individual, or a single group—by altering the explanation in one respect. One might argue that such efficient norms are not learned spontaneously by each person or each group, but are the product of accumulated social learning across several generations. That is, groups may move toward efficient norms over many generations and may perpetuate those norms because they are efficient and lead to success.

In a sense, this argument borrows the refinement mechanism of evolutionary theory and makes it available to the general social learning explanation. In other words, the claim is that groups with efficient norms grow, do better, and spread their norms, and thus over time there comes to be a great deal of agreement because of the

211. F. A. HAYEK, THE FATAL CONCEIT: THE ERRORS OF SOCIALISM (1989); ADAM SMITH, AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS (Univ. Chi. Press 1976) (1776).

212. DIAMOND, *supra* note 209, at 177 ("[W]e have to wonder why the kings and nobles failed to recognize and solve these seemingly obvious problems undermining their society . . . . Like most leaders throughout human history, the Maya kings and nobles did not heed long-term problems, insofar as they perceived them.").

diffusion of the efficient norms.[213] The explanation is "evolutionary" in that it involves change over time, with less successful norms losing out to more successful ones. However, it differs from the evolutionary explanation in Part II, of course, because here the selection occurs through the operation of social learning, rather than through gene selection.[214] But this revised social learning explanation, which might be called "accumulated social learning," still faces serious difficulties.

### 1. Absence of Variation in Intuitions of Justice Among Groups Despite Large Differences in Situation and Culture

There exist vast differences in ecology, history, demographics, social structure, and many other variables from one group and culture to another. These differences are so striking and of such a nature that it seems odd that all groups would find the same norms to be the most efficient and, further, that each group would be equally effective in teaching these norms to each successive generation. It would be surprising if dramatic differences in social structure and social resources had no effect on how, when, and what a generation would

---

213. This type of view has its historical roots in the writings of anthropologists who developed ideas surrounding "functionalism," the notion that societal institutions keep the society working. The founders of the most relevant areas of this discipline are Emile Durkheim and Alfred Reginald Radcliffe-Brown. *See, e.g.*, EMILE DURKHEIM, THE DIVISION OF LABOR IN SOCIETY (Free Press 1984) (1893); ALFRED REGINALD RADCLIFFE-BROWN, STRUCTURE AND FUNCTION IN PRIMITIVE SOCIETY (1952).

In the more modern anthropological literature, this is referred to as "cultural group selection." *See, e.g.*, Joseph Soltis et al., *Can Group-functional Behaviors Evolve by Cultural Group Selection? An Empirical Test*, 63 CURRENT ANTHROPOLOGY 473, 474 (1995) (extending functionalist arguments with a formal analysis and ethnographic evidence). For a recent, comprehensive treatment of this and related issues, see PETER J. RICHERSON & ROBERT BOYD, NOT BY GENES ALONE: HOW CULTURE TRANSFORMED HUMAN EVOLUTION (2005).

214. Learning models are more sophisticated than we allude to here, but beyond the scope of the present discussion. Probably two of the most important transmission models of this type are based on 1) conformity to the most common norm and 2) imitation of those who are of high prestige. For a discussion of the former, see ROBERT BOYD & PETER J. RICHERSON, CULTURE AND THE EVOLUTIONARY PROCESS 205-40, 259-79 (1985) and Joseph Henrich & Robert Boyd, *The Evolution of Conformist Transmission and the Emergence of Between-group Differences*, 19 EVOLUTION & HUM. BEHAV. 215, 231-37 (1998). For a discussion of the latter, see Joseph Henrich & Francisco Gil-White, *The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission*, 22 EVOLUTION & HUM. BEHAV. 165, 180-92 (2001), which argues that there are benefits to deferring to people who are successful in an environment so that one can learn from these individuals. *See also* Robert Boyd & Peter J. Richerson, *Group Beneficial Norms Can Spread Rapidly in a Structured Population*, 212 J. THEORETICAL BIOLOGY 287, 288-89 (2002) (using evolutionary game theory to argue that beneficial social norms spread quickly through populations); Joseph Henrich, *Cultural Group Selection, Coevolutionary Processes and Large-scale Cooperation*, 53 J. ECON. BEHAV. & ORG. 3, 4-31 (2004) (presenting a formal model explaining how cultural group selection can lead to the spread of norms for prosocial behavior, including the punishment of norm violators).

learn about justice from each previous generation. One would expect some groups to be better than others at approximating the efficient norm and some groups to be better than others at teaching that approximation to the succeeding generation.

Yet, as we documented in Part I, there is a consensus on the core intuitions of justice even across demographics and cultures. How could one explain this consensus if those views simply reflect the efficient norm for each group? This seems a difficult task. It would require finding a set of conditions that are universal to all human groups, no matter their circumstances and culture, and showing that it is these universal conditions alone that set the efficient norms for the group's views of justice. If any factor, other than these universal conditions, had influence in shaping the core views of justice, then one would find differences between groups according to the differences of these non-universal factors. Perhaps others will take up this challenge—finding these universal conditions and showing how these alone determine core views of justice—but until such can be found, or even imagined, it would seem to make the spontaneous social learning explanation implausible. In contrast, the evolutionary explanation suffers no such problem. Evolutionary explanations, dependent as they are on evolutionarily evolved mechanisms that all humans share, fit naturally with human universals.[215]

### 2. Inconsistency with Developmental Data

Even more problematic for the accumulated social learning explanation is that it simply does not fit the developmental data summarized in Part III.C. For example, we know that children's views of morality emerge early and are nuanced from a young age. Recall that even very young children—around three years of age—are precocious universal moralists, believing that certain offenses are wrong, even if there were no rule about the act, and distinguish moral rules from rules of convention. In particular, children believe that unprovoked physical harm is wrong and, further, very young children distinguish among offenses (harm versus theft) and take intent and

215. *See* DONALD BROWN, HUMAN UNIVERSALS (1989) (discussing the relationship between evolutionary biology and human cultural universals and suggesting that "human biology *is* a key to understanding many human universals" and that "evolutionary psychology is a key to understanding many of the universals that are of greatest interest to anthropology"); *see also* John Tooby & Leda Cosmides, *On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation*, 58 J. PERSONALITY 17, 23-24 (1990) (arguing that the process of evolution by natural selection leads to a universal "design" for any given species, including humans, leading to many universal physiological and psychological features).

extenuating circumstances into account in evaluating blameworthiness for violations, suggesting nuanced views of justice.[216]

If views of justice were the result of social learning, one would expect that young children would have to be taught distinctions such as those between precautionary rules and moral rules.[217] But it seems implausible that these abstractions could be taught to such young children, who appear to come to these important distinctions precociously.[218] The development of these intuitions seems to parallel more closely, for example, the growth of teeth than learning to play chess. Like teeth, these intuitions simply develop according to a pre-programmed timetable for all children, as opposed to chess, which must be taught systematically and over time with explicit instruction. Indeed, chess has a small number of rules, as compared to the intricacies of moral judgment, and yet the learning of morality, somewhat paradoxically, seems to come much more easily and naturally to children than learning chess.

It is of course not impossible that children's ideas about wrongs and punishment are transmitted to them by parents and peers while they are young. It is significant, however, that these intuitions come online across children in a reliably developing sequence at roughly the same time throughout the world, without differences reflecting the varied practical, social, and cultural conditions. That suggests these intuitions are part of a specific, developed learning mechanism, rather than the result of general social learning, which is necessarily dependent upon the child's environment, which in turn differs substantially from place to place.

### 3. Intuitional Knowledge as Having Distinct Characteristics from Learned Knowledge

Even if the social learning explanation were consistent with the developmental data, it nonetheless would be problematic. It is inconsistent with what we know about the nature of views of justice: their intuitional nature. If views of justice were learned from others, one would expect such views to have the character of other learned knowledge. That is, a social learning explanation assumes that our judgments about justice are like many other judgments that we make in our daily lives, such as how fast to drive on a stretch of road, how long to cook a sandwich in the microwave, or how carefully to take

216. *See supra* Part III.C.

217. *See supra* Part III.C.2.a.

218. *See supra* Part III.C.2.

notes on a lecture. These judgments, we assume, are the product of reasoning, promoted by life experience and education.

But social science evidence suggests that judgments about justice, especially for violations that might be called the core of criminal wrongdoing, are more the product of intuition than reasoning. Their intuitional nature means, among other things, that they are judgments quickly arrived at (even by people with little education or life experience), that they frequently are held with strong feelings of certainty, and that the reasons we hold such judgments are generally not consciously accessible to us.

In Jon Haidt's work on "moral dumbfounding," for example, people report strong intuitions about things that are morally wrong, such as consensual, non-reproductive incest, but are unable to provide a principled explanation for their judgments.[219] Similarly, Marc Hauser, Liane Young, and Fiery Cushman looked at judgments of morally permissible actions using the trolley problem,[220] in which a certain number of people can be saved from being killed by a runaway trolley if some action is taken (or not taken), resulting in the death of a smaller number of people, often a single individual. People are asked what choice should be made in these difficult situations and to explain their reasoning. While subjects commonly have strong and clear views on the proper result, they also commonly are unable to offer an explanation for their conclusions. For example, in one variation of the trolley problem, the subject can avoid the death of five people on the track by (a) pushing a bystander on the track whose body and backpack will jam the trolley's wheels, stopping it, or (b) throwing a switch to divert the trolley to a side track where it will kill one person. Despite the fact that the results of the two actions are identical, eighty-nine percent of the subjects considered the latter action moral but only eleven percent judged the former to be moral. More interesting for our purposes, seventy percent of the subjects could give no plausible explanation for their judgment.[221] This mirrors Haidt's results concerning incest.

---

219. Haidt, *The Emotional Dog and Its Rational Tail*, *supra* note 42, at 814.

220. Philippa Foot, *The Problem of Abortion and the Doctrine of Double Effect*, 5 OXFORD REV. 5, 8 (1967).

221. Hauser et al., *supra* note 131 (manuscript at 20-21). Even the 30 percent who gave what the authors classed as a "sufficient justification" may have been making purely ex post attempts at explanation rather than reporting reasoning they used in reaching their conclusion. *Id.* (manuscript at 19). The authors used what they called an "extremely liberal criterion": "A sufficient justification was one that correctly identified any factual difference between the two scenarios and claimed the difference to be the basis of moral judgment." *Id.*

There is at least some degree of consensus that many moral judgments are made by a deeply intuitive system. Based on his review of the existing social science literature, Haidt concluded that "moral judgments" derive from "quick automatic evaluations (intuitions)."[222] Similarly, Hauser concludes that "much of our knowledge of morality is . . . intuitive, based on unconscious and inaccessible principles. . . ."[223] If such judgments were the product of a set of principles of morality learned from others, it would be a straightforward matter to derive the "wrongness" of acts from these principles, just as mathematical inferences can be made from a set of axioms and subsequently explained by reference to them. "Moral dumbfounding" and related effects in the psychological literature suggest that this is not how these judgments are made.

4. Difficulties in Teaching Inarticulable Lessons

Further undermining the social learning explanation for shared views of justice is the fact that, to the extent that people are unable to articulate the principles that underlie their intuitions of justice, it is difficult to transmit those principles verbally to others. In other words, the intuitional nature of judgments about justice undermines the argument that they are socially learned through language. Of course, learning through non-linguistic means is possible, but the social transmission of information is vastly more difficult without natural language. The psychological literature is replete with examples of this general principle,[224] but the general idea is well illustrated by two relevant phenomena.

First, consider that social transmission is rare in non-humans, restricted to narrow domains such as song-learning in birds. Without language, non-human animals are confined to very modest social learning, so much so that each case of social learning is considered extraordinary.[225] Second, consider some vivid examples from everyday

222. Haidt, *The Emotional Dog and Its Rational Tail*, *supra* note 42, at 814, 819-20 (discussing the automaticity of such judgments).

223. HAUSER, *supra* note 45, at 125. For a recent review of relevant literature, see Cass R. Sunnstein, *Moral Heuristics*, 28 BEHAV. & BRAIN SCI. 531, 531-42 (2005).

224. For a recent example, see Bjørn Sætrevik et al., *The Utility of Implicit Learning in the Teaching of Rules*, 16 LEARNING & INSTRUCTION 363, 363, 372-73 (2006) (showing that in teaching rules about chemistry to students, "learning was much more effective when more explicit ways of teaching were employed").

225. *See* Michael Tomasello, *Culture and Cognitive Development*, 9 CURRENT DIRECTIONS COGNITIVE SCI. 37 (2000) (discussing the human adaptation for cultural transmission of information, contrasting it with non-humans' capacities, in particular non-human primates); *see also* Bennett G. Galef, Jr., *The Question of Animal Culture*, 3 HUM. NATURE 157, 162 (1992)

life. The game of "charades" is fun precisely because, deprived of natural language, conveying simple ideas, even individual words, is difficult. Similar examples include the difficulty in understanding what pre-linguistic infants are trying to communicate and asking even rudimentary questions of someone with whom one does not share a common language. Social learning without language is not impossible, but it is certainly difficult. If individual words and simple questions pose a challenge, consider how much more difficult it would be to convey abstract, complex notions of deontic principles non-linguistically. If people do not have explicit access to these principles, it is unlikely that they are socially transmitted.

In addition to these many failings of the accumulated social learning explanation, recall that it also suffers the objection raised with regard to the original "spontaneous social learning" explanation: that such an explanation depends upon individuals putting group interests above their individual interests. As noted above, the available evidence suggests that such altruism has serious limits. Until it can be shown that there are special circumstances and reasons that prompt such altruistic behavior, the accumulated social learning explanation is on this ground as implausible as the spontaneous social learning explanation.

## CONCLUSION

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.

Whichever conclusion one prefers, important implications follow. Shared intuitions of justice are not easily altered, regardless of their source. Even if the source is general social learning, it must be social learning arising *only* from an aspect of human life experience that is so fundamental as to be essentially universal to all persons without regard to circumstances or culture. In other words, it is not a source that is open to easy change or manipulation. If it were not so fixed, if it were easily manipulable, then the natural variations in

(arguing that non-human animals do not imitate one another in a way that closely parallels human imitation, looking in detail at two of the most prominent cases of putative non-human cultural transmission).

circumstance and culture would have altered it and produced variations in intuitions of justice; there would not be the agreement that exists.

While a full account of these implications is another project,[226] consider the possibilities. For example, it may be unrealistic to expect the population to "rise above" its desire to punish wrongdoers and to expect the government to "re-educate" people away from their interest in punishing wrongdoers, as is urged by some reformers. It is unlikely that the shared intuition that serious wrongdoing should be punished can be changed through social engineering, at least not through methods short of the kind of coercive indoctrination that liberal democracies find unacceptable.

For another example, a criminal justice system that regularly fails to do justice or that regularly does injustice, as judged by shared intuitions of justice, will risk a great deal. It will inevitably be seen as failing in a mission thought important, even foundational, by the community—unless the system can hide its unjust operation. That would be hard to accomplish without breaching notions of press freedom and government transparency to which liberal democracies aspire.

As a final example, any realistic criminal justice system or program for its reform must acknowledge and engage the community's shared intuitions of justice. This does not mean that law can never deviate from those intuitions or try to change them. There is nothing sacred or immutable about our current intuitions of justice. But a criminal justice system must be realistic about how different kinds of changes may require different—and sometimes high—financial and social costs. The greatest success in shaping the perceived wrongfulness of particular conduct might not be to fight people's intuitions of justice, but to try to harness them, by providing information or arguments that strengthen (or weaken) the analogy between the target conduct and the core wrongdoing on which people have strong intuitions.

226. Paul H. Robinson & John M. Darley., *Intuitions of Justice: Implications for Criminal Law and Justice Policy*, 81 S. CAL. L. REV. (forthcoming 2007) (focusing on common intuitions of justice and injustice and examining their implications on the efficacy of criminal systems and reform movements).