BerkeleyLaw
UNIVERSITY OF CALIFORNIA

Kadish Center for Morality, Law & Public Affairs

GENERAL ASPECTS OF LAW

GALA

DEAN'S SEMINAR ROOM
(215 BOALT HALL)

THURSDAY
APRIL 5, 2012
4:10 – 6:00 PM

PART I: "EMOTION, NEUROSCIENCE, AND LAW: A COMMENT ON DARWIN AND GREENE"

PART II: "'ANY ANIMAL WHATEVER': HARMFUL BATTERY AND ITS ELEMENTS AS BUILDING BLOCKS OF HUMAN AND NONHUMAN MORAL COGNITION"

By

JOHN MIKHAIL

PROFESSOR OF LAW AND PHILOSOPHY

GEORGETOWN UNIVERSITY

NOTE

The GALA speaker will offer some brief introductory background remarks about the paper and the remainder of the time will be devoted to discussion. Participants are expected to read the paper in advance. Papers are distributed to those on our mailing list or electronically when available at the GALA website:
< http://www.law.berkeley.edu/9264.htm>.

Others may obtain copies from Ms. Amatullah Alaji-Sabrie at 510.642.3627 or asabrie@law.berkeley.edu. Copies can be made available in an alternate format upon request. GALA events are wheelchair accessible. For any disability-related accommodations advance notice is requested.

emotion review

Emotion Review
Vol. 3, No. 3 (July 2011) 293–295

ISSN 1754-0739
DOI: 10.1177/1754073911406150
er.sagepub.com

Emotion, Neuroscience, and Law: A Comment on Darwin and Greene

John Mikhail
Georgetown University, USA

Abstract

Darwin's (1871/1981) observation that evolution has produced in us certain emotions responding to right and wrong conduct that lack any obvious basis in individual utility is a useful springboard from which to clarify the role of emotion in moral judgment. The problem is whether a certain class of moral judgment is "constituted" or "driven by" emotion (Greene, 2008, p. 108) or merely correlated with emotion while being generated by unconscious computations (e.g., Huebner, Dwyer, & Hauser, 2008). With one exception, all of the "personal" vignettes devised by Greene and colleagues (2001, 2004) and subsequently used by other researchers (e.g., Koenigs et al., 2007), in their fMRI and behavioral studies of emotional engagement in moral judgment, involve violent crimes or torts. These studies thus do much more than highlight the role of emotion in moral judgment; they also support the classical rationalist thesis that moral rules are engraved in the mind.

Keywords
computation, emotion, law, neuroscience

In *The Descent of Man*, Darwin (1871/1981, p. 70) affirmed his belief in an innate moral faculty, explaining that he fully agreed with Kant and other writers "that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important." Darwin insisted that the moral sense is not a mysterious gift of unknown origin, however, but the natural result of evolution, with antecedents in the social instincts of other animals. He thus famously argued that "any animal whatever, endowed with well-marked social instincts, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or as nearly developed, as in man" (Darwin, 1871/1981, pp. 71–72). And he laid the foundation of subsequent research on the evolution of morality by examining a range of animal traits and behaviors, including their sociability, desire for companionship, and the misery they feel when they are abandoned; their love, sympathy, and compassion for one another; and their mutual willingness to sacrifice themselves and to render services to one another when hunting or defending against attack.

Darwin held that the social instincts of nonhuman animals developed "for the general good of the community," which he defined as "the means by which the greatest possible number of individuals can be reared in full vigor and health, with all their faculties perfect, under the conditions to which they are exposed" (Darwin, 1871/1981, pp. 97–98). The same was true of *homo sapiens*, he inferred; therefore, neither egoism nor a universalistic hedonism (the "Greatest Happiness Principle") was descriptively adequate: "When a man risks his life to save that of a fellow-creature, it seems more appropriate to say that he acts for the general good or welfare, rather than for the general happiness of mankind" (Darwin, 1871/1981, p. 98). Darwin endorsed Herbert Spencer's conclusion that "the experiences of utility organized and consolidated through all past generations of the human race, have been producing corresponding modifications, which, by continued transmission and accumulation, have become in us certain faculties of moral intuition—certain emotions responding to right and wrong conduct, which have no apparent basis in the individual experiences of utility" (Darwin, 1871/1981, pp. 101–102). Finally, Darwin held that this combination of social instincts, intellectual powers, and effects of habit would "naturally lead to the golden rule: 'As ye would that men should do to you, do ye to them likewise.'" This rule, he averred, "lies at the foundation of morality" (Darwin, 1871/1981, p. 106).

The idea that evolution has produced in us "certain emotions responding to right and wrong conduct" that lack any obvious basis in individual experiences of utility is a useful

Corresponding author: John Mikhail, 600 New Jersey Ave., NW Washington, DC, 20001, USA. *Email*: mikhail@law.georgetown.edu

springboard from which to clarify an important problem in the cognitive science of moral judgment. The problem is how to understand the role of emotion in moral judgment, and specifically whether a certain class of moral judgment is "constituted" or "driven by" emotion (Greene, 2008, p. 108; see also Greene et al., 2004; Koenigs et al., 2007) or merely correlated with emotion while being generated by unconscious computations (e.g., Huebner et al., 2008). In at least some interpretations, there are important differences between these formulations, although these differences may disappear at certain neurocognitive or neurobiological levels of scientific description. My claim is that the second formulation—the Darwin–Spencer thesis, according to which emotions "respond to" independent moral appraisals—is a better working model of moral cognition with respect to this class of judgments.

To see why, it is useful to look closely at the 25 "personal" dilemmas devised by Greene and colleagues in their original fMRI study (2001), and subsequently used by a number of other researchers (e.g., Greene et al., 2004; Koenigs et al., 2007; Moore, Clark, & Kaine, 2008). Greene found that these vignettes elicited increased activity in the medial prefrontal cortex (MPFC), posterior cingulated cortex (PCC), superior temporal sulcus (STS), and amygdala. Because these regions are associated with emotional processing, he concluded that these "characteristically deontological" judgments are driven by emotion. What seems to have escaped his notice and that of the scientific community generally, however, is that *all of the actions described by these vignettes are well-known crimes or torts* (Table 1). Specifically, 22 of the 25 scenarios satisfy a prima facie case for purposeful battery and/or intentional homicide (i.e., murder). Two other cases involve acts of rape and sexual battery, while the final case describes a negligent (i.e., unreasonable) failure to rescue.

Table 1. Standard legal analysis of Greene et al.'s (2001) "personal" dilemmas

Dilemma	Standard legal analysis
1. Transplant	Battery/Homicide
2. Footbridge	Battery/Homicide
3. Country Road	Negligent Failure to Rescue
4. Architect	Battery/Homicide
5. Lifeboat	Battery/Homicide
6. Hard Times	Battery/Rape/Sexual Assault
7. Smother for dollars	Battery/Homicide
8. Safari	Battery/Homicide
9. Crying Baby	Battery/Homicide
10. Plane Crash	Battery/Homicide
11. Hired Rapist	Battery/Rape/Sexual Assault
12. Grandson	Battery/Homicide
13. Infanticide	Battery/Homicide
14. Preventing the Spread	Battery/Homicide
15. Modified Lifeboat	Battery/Homicide
16. Modified Preventing the Spread	Homicide
17. Modified Safari	Battery/Homicide
18. Modified Bomb	Battery/Torture
19. Submarine	Battery/Homicide
20. Lawrence of Arabia	Battery/Homicide
21. Sophie's Choice	Battery/Homicide
22. Sacrifice	Homicide
23. Vitamins	Battery
24. Vaccine Test	Battery/Homicide
25. Euthanasia	Battery/Homicide

With one exception, then, what Greene actually did in the "personal" condition of his experiment was to put subjects in the scanner and ask them to respond to a series of violent crimes and torts. There are other relevant features of these scenarios, of course; some of them raise principal-agent problems and others involve duress or necessity, for example. Fundamentally, however, all of them describe acts that standard legal analysis would classify as serious wrongs, subject to conceivable, but ultimately weak, affirmative defenses. Moreover, all of them involve serious bodily injury and thus implicate the right to physical safety. By contrast, only five of the 19 cases in Greene's "impersonal" condition are batteries, and only one of these batteries is purposeful. The other four cases involve foreseeable but non-purposeful harms, at least two of which admit of an uncontroversial necessity defense. The remaining 14 "impersonal" scenarios are a hodgepodge of cases that raise a variety of ethical and legal issues, including fraud, tax evasion, insider trading, public corruption, theft, unjust enrichment, and necessity as a defense to trespass to chattels. Finally, five of these residual cases describe risk–risk tradeoffs in the context of vaccinations and environmental policy.

The upshot is that Greene's original experiments did not really test two patterns of moral judgment—one "deontological" and the other "utilitarian"—so much as different categories of potentially wrongful behavior. The basic cleavage he identified in the brain was not Kant versus Mill, but purposeful battery, rape, and murder, on the one hand, and a disorderly grab bag of theft crimes, regulatory crimes, torts against non-personal interests, and risk–risk tradeoffs, on the other. Moreover, his finding that the MPFC, PCC, STS, and amygdala are recruited for judgment tasks involving purposeful battery, rape, and murder does not undermine the traditional rationalist thesis that moral precepts are engraved in the mind (e.g., Grotius, 1625/1925; Kant, 1788/1993; Leibniz, 1704/1981). On the contrary, Greene's evidence largely supports that thesis. Crimes and torts have *elements*, and the relevant pattern of intuitions is best explained by assuming that humans possess implicit knowledge of moral and legal *rules*. Naturally, violent crimes and torts are more emotionally engaging than insider trading or environmental risk analysis, but it does not follow that emotion "constitutes" or "drives" the judgment that the former acts are wrong. Rather, what drive these intuitions are the unconscious computations that characterize these acts as battery, rape, or murder in the first place. By mischaracterizing their own stimuli, then, Greene and other neuroscientists (e.g., Koenigs et al., 2007) appear to have drawn specious conclusions and misconceived the nature of the problem.

Returning to Darwin, the main questions for cognitive science going forward include (1) how the brain computes

unconscious representations of purposeful battery, rape, murder, negligence, and other forms of harmful trespass, and (2) how these computations and the negative emotions they typically elicit are related to the complex cognitive and socio-emotional capacities that humans share with other animals (cf., Darwin, 1871/1981; Spencer, 1897/1978; see generally Mikhail, 2007, 2009, in press). Future research should focus more squarely on these topics and move beyond potentially misleading pseudo-problems such as how reason and emotion "duke it out" in the brain.

References

Darwin, C. (1981). *The descent of man, and selection in relation to sex*. Princeton, NJ: Princeton University Press. (Original work published 1871)

Greene, J. (2008). The secret joke of Kant's soul; and Reply to Mikhail and Timmons. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 3: The neuroscience of morality: Emotion, disease, and development* (pp. 35–79, 105–117). Cambridge, MA: MIT Press.

Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural basis of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400.

Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.

Grotius, H. (1925). *On the law of war and peace* (F. W. Kelsey, Trans.). Oxford: Clarendon Press. (Original work published 1625)

Huebner, B., Dwyer, S., & Hauser, M. (2008). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*, 1–6.

Kant, I. (1993). *Critique of practical reason* (L.W. Beck, Trans.). New York: Macmillan. (Original work published 1788)

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to ventromedial prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*, 908–911.

Leibniz, G. (1981). *New essays on human understanding* (P. Remnant & J. Bennett, Eds.) Cambridge: Cambridge University Press. (Original work published 1704)

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences*, *11*, 143–152.

Mikhail, J. (2009). Moral grammar and intuitive jurisprudence: A formal model of unconscious moral and legal knowledge. In B. H. Ross (Series Ed.) & D. M. Bartels, C. W. Bauman, L. J. Skitka & D. L. Medin (Eds.), *Psychology of learning and motivation, Vol. 50: Moral judgment and decision making*. San Diego, CA: Academic Press.

Mikhail, J. (in press). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge: Cambridge University Press.

Moore, A., Clark, B., & Kaine, M. (2008). Who shall not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*, 549–557.

Spencer, H. (1978). *The principles of ethics*. Indianapolis: Liberty Classics. (Original work published 1897)

DRAFT – PLEASE DO NOT QUOTE; COMMENTS WELCOME

'Any Animal Whatever': Harmful Battery and its Elements as Building Blocks of Human and Nonhuman Moral Cognition

John Mikhail

Abstract: Recent work in experimental ethics suggests that its main questions for future research include (1) how the brain computes representations of battery, negligence, and other harmful trespasses and (2) how these computations and the negative emotions they elicit are related to the complex cognitive and socio-emotional capacities that humans share with other animals. In this paper, I draw on legal theory to examine the simplest part of the first question: how the brain computes representations of harmful battery. I argue that the key elements of the *prima facie* case of harmful battery, including voluntary act, harmful or offensive intention, and harmful contact, appear to form critical building blocks of moral cognition in both humans and nonhuman animals. The norm against harmful battery, therefore, appears likely to be the product of an animal instinct with deep evolutionary roots. By contrast, at least some of the more complex mental representations presupposed by familiar justifications to battery appear to be uniquely human. After outlining these ideas, the paper also argues that many famous thought experiments in ethics (e.g., Boorse & Sorenson 1988; Foot 1967; Nagel 1986; Rachels 1975) and many influential experiments in moral psychology (e.g., Cushman et al. 2006; Hamlin et al. 2007; Koenigs et al. 2007; Turiel 1983;) rely on harmful battery scenarios without clearly acknowledging this fact or considering what it implies for issues of experimental design, data interpretation, or theory construction. The significant generalization that has escaped notice is an acute sensitivity to battery as a property of the human mind.

I.

Recent work in experimental ethics suggests that its main questions for future research include (1) how the brain computes representations of battery, negligence, and other forms of harmful trespass, and (2) how these computations and the negative emotions they typically elicit are related to the complex cognitive and socio-emotional capacities humans share with other animals (Mikhail 2011a, 2011b; see also Cushman et al. 2006; Greene et al. 2009; Wood et al. 2007; Young et al. 2007; cf. Darwin 1981/1871). In this paper, I draw on legal theory to examine the simplest part of the first question: how the brain computes representations of harmful battery. The hypothesis I wish to explore is that one component of this process, which corresponds to the *prima facie* case of harmful battery, appears to be relatively simple, and thus may be shared with nonhuman animals, whereas a second component, which corresponds to affirmative defenses to harmful battery, is more complex, and thus may be uniquely human.

Further, I wish to show that the harmful battery norm can be used to explain and unify a wide range of familiar moral intuitions in ethics and cognitive science.

The examples from ethics include famous thought experiments by Foot (1967), Rachels (1975), Nagel (1986), and Boorse & Sorenson (1988). The examples from cognitive science include influential experiments by Cushman et al. (2006), Greene et al. (2009), Hamlin et al. (2007), and Turiel (1983). Many other illustrations could be offered, but these eight are adequate for my purposes. The claim I wish to defend is that all eight cases rely on fact patterns involving acts of harmful battery—the same wrong at issue in the footbridge problem and other standard trolley problems (Thomson 1976, 1985), as well as in the so-called "personal" moral dilemmas devised by Greene and colleagues (2001, 2004) and subsequently used by other researchers (e.g., Ciaramelli et al. 2007; Koenigs et al. 2007; Moore et al. 2008; Valdesolo & DeSteno 2006) in their fMRI and behavioral studies of emotional engagement in moral judgment (see Figure 1). The point of convergence in all of these experiments thus appears to be an acute sensitivity to *battery* as a property of the human mind. A decade of intensive research has revealed that this *prima facie* norm can sometimes be overcome by specific justifications, but only under special circumstances; otherwise the norm is difficult to dislodge. In this light, one can perhaps better appreciate the complex deontological structure of common morality, which, *pace* Sidgwick (1907: 453), appears to be "unconsciously utilitarian" only up to a point (roughly, the point captured by the joint operation of a prohibition of purposeful harm and a negative utilitarian justification of non-purposeful harm—on the latter, see Popper 1945; Smart 1958; on their joint operation, see, e.g., Mikhail 2011a; Nichols 2005). These claims will be elaborated, qualified, and defended in due course. First, it is necessary in this paper to examine the key elements of battery and what they reveal about the structure of the moral mind.

This remainder of this paper proceeds in four parts. Part II outlines the main elements of the *prima facie* case of harmful battery, including voluntary act, harmful or offensive intent, and harmful contact, and considers what these elements might tell us about the rich internal structure of both human and nonhuman moral psychology. Part III adds more texture to the inquiry by considering the primary defenses to battery in both morality and law, with special emphasis on different types of consent. The main argument of this part is that, unlike the *prima facie* case of harmful battery, which implicates aspects of moral judgment that might plausibly be shared with nonhuman animals, the mental operations presupposed by these affirmative defenses point us

toward elements of moral cognition that appear to be distinctively human. Part IV examines eight influential articles or research programs in ethics and cognitive science and argues that all of them rely on harmful battery scenarios without expressly recognizing this fact or considering its significance for issues of experimental design, data interpretation, or theory construction. Part V draws several lessons from these observations, including how they might bear on Darwin's (1981/1871: 71-72) famous claim that "any animal whatever" would acquire a moral sense or conscience once "its intellectual powers had become as well developed, or as nearly developed, as in man." Part VI concludes.

II.

Our first task is to consider what the elements of battery might tell us about the rich internal structure of human—and perhaps also nonhuman—moral psychology. In Anglo-American law, battery is usually divided into two classes: harmful battery and offensive battery. The touchstone of the former is harmful bodily contact. Offensive battery is usually conceived of as a distinct category that encompasses non-harmful bodily contacts that nevertheless offend a reasonable sense of personal dignity. A common example is spitting in another's face. Greene (2005: 345; see also Greene & Haidt 2002: 519) and other cognitive scientists sometimes refer to "assault" in the context of the trolley problems, but it is important to recognize that the technical legal concept of assault as it is used in the law of torts is distinct from both harmful and offensive battery, insofar as it turns on the apprehension of an imminent harmful or offensive bodily contact, but does not require that contact to occur. Assault thus implicates different mental representations than either harmful or offensive battery. Properly understood, however, all three of these acts appear to be useful stimuli for testing intuitive capacities for action comprehension and moral evaluation in adults, children, infants, and nonhuman animals.

Our focus here is harmful battery. We can construct a working definition of this concept by drawing on the Restatements (First) of Torts, which defines harmful battery as follows:

> 13. Battery: Harmful Contact
>
> An act which, directly or indirectly, is the legal cause of a harmful contact with another's person makes the actor liable to the other, if

a) the act is done with the intention of bringing about a harmful or offensive contact or an apprehension thereof to the other or a third person, and
b) the contact is not consented to by the other or the other's consent thereto is procured by fraud or duress, and
c) the contact is not otherwise privileged.

Prosser (1941: 43) summarizes the main elements of this definition as "unpermitted, unprivileged contact with a person." Prosser's formula is incomplete and misleading, however, in at least two respects: first, it omits the element of harmfulness; and second, it neglects to explain that the harmful contact required for battery must be produced by a voluntary act rather than a voluntary omission. Particularly given the empirical motivation of this paper, which includes the elaboration of an experimental paradigm that can be effectively used with a range of populations, including adults, children, infants, and nonhuman animals, it is important to make these additional elements explicit.

Setting aside the assault-related reference to the apprehension of an impending harmful or offensive contact in condition (a), and thus simplifying for purposes of this exposition, the Restatement definition implies that the mental representation of harmful battery is a complex mental construction, imposed on the stimulus by an appraisal system or action analysis of some sort, which consists of five elements: act, harmful or offensive intention, harmful contact, lack of consent, lack of privilege. Formally, this definition can be summarized as follows:

$$\text{Harmful Battery} =_{df} [\text{ACT}, \text{INTENT}_{H/O}, \text{CONTACT}_{H}, \sim\text{CONSENT}, \sim\text{PRIVILEGE}]$$

In this formula, the subscript "H" stands for harmful and the subscript "H/O" stands for harmful or offensive. These features will be explained as we proceed.

The Restatement (Second) of Torts (1965) provides a simpler definition of battery:

> 13. Battery: Harmful Contact
>
> An actor is subject to liability to another for battery if
>
> a) he acts intending to cause a harmful or offensive contact with the person of the other or a third person, or an imminent apprehension of such contact, and

b) a harmful contact with the person of the other directly or indirectly results

According to this definition, harmful battery consists of three primary elements: act, harmful or offensive intention, and harmful contact. Stated more formally, and still simplifying, the definition reads:

Harmful Battery $=_{df}$ [ACT, $\text{INTENT}_{H/O}$, CONTACT_{H}]

In other words, harmful battery is comprised of three main building blocks: action, intention, and contact (cf. Cushman et al. 2006). Significantly, lack of consent and lack of privilege have dropped out the picture. Does this mean that the legal understanding of harmful battery changed between 1935 and 1965? No: the difference between the two definitions stems from the fact that the drafters of the second Restatement (probably Prosser himself) opted to limit Section 13's definition, in effect, to the *prima facie* case for harmful battery, leaving the elements pertaining to possible defenses (i.e., consent or privilege) for other sections (see, e.g., Sections 49-62 on consent and Sections 63-76 on privileged use of force in defense of self or others). This decision is also reflected in the difference between the phrases "makes the actor liable" in the Restatement (First) and "is subject to liability" in the Restatement (Second).[1]

Apart from these differences, the foregoing definitions appear relatively simple and more or less consistent with one another. Yet, as philosophers surely would insist, complications lurk beneath the surface. What is an act? What is the proper analysis of causation and intention? These concepts have been the subject of intense philosophical scrutiny for decades (for act, see, e.g., Bennett 1995; Davidson 1980; Goldman 1970; and Holmes 1881; for causation, see, e.g., Donagan 1977; Hart & Honore 1959; and Mackie 1974; for intention, see, e.g., Anscombe 1957; Bratman 1987; Lawrence 1972; and Sidgwick 1907). It is therefore important to say a further word about these elements here, both to clarify how jurists have explicated them and to render them easier to apply in an experimental context.

Limiting our attention to the second Restatement's definition of harmful battery, at least seven concepts stand in need of clarification: *act*, *intent*, *contact*, *harmful*, *offensive*, *directly*, and *indirectly*. The following remarks are meant to clarify these elements, at least in a provisional manner that seems adequate for our purposes.

1. *Act*. Legal writers have often struggled to define the deceptively simple concept of an *act* in a way that is coherent, plausible, and operational. Some jurists, such as Austin (1879), restrict the concept to the voluntary movement of one's body, as distinct from all of the consequences which follow from this movement. In the example of shooting and killing another person, for instance, Austin identifies the act to be "the muscular motions" by which the agent raises the weapon, points it, and pulls the trigger. The act's *consequences* include "[t]he contact of the flint and steel, the ignition of the powder, the flight of the ball . . . [and] the wound and subsequent death" (1879: 427). At the opposite extreme is Salmond (1924), whose legal treatise is cited by Foot (1967) in her original trolley problem article. Salmond's broad definition of the "act" includes not only the agent's voluntary bodily movements (or deliberate omissions) but also their surrounding circumstances and consequences. Salmond (1924: 383) also illustrates his approach with an act of shooting, explaining that the circumstances include the fact that "the rifle is loaded and in working order, and that the person killed is in the line of fire," while the consequences include "the fall of the trigger, the explosion of the powder, the discharge of the bullet, its passage through the body of the man killed, and his death" (see Hall 1960: 172-173).

The Restatements of Torts and other leading authorities (e.g., Holmes 1881) hew closely to Austin's narrow definition of act. The first and second Restatements offer identical definitions of the concept, which read as follows:

> 2. Act
>
> The word "act" is used throughout the Restatement of this Subject to denote an external manifestation of the actor's will and does not include any of its results even the most direct, immediate and intended.

Comment c elaborates on this theme, albeit in a slightly inconsistent manner:

> c. *Act and its consequences*. The word "act" includes only the external manifestation of the actor's will. It does not include any of the effects of such manifestation no matter how direct, immediate and intended. Thus, if the actor, having pointed a pistol at another, pulls the trigger, the act is the pulling of the trigger and not the impingement of the bullet on the other's person. So too, if the actor intentionally strikes another, the act is only the movement of the actor's hand and not the contact with the other's body immediately established thereby.

The slight inconsistency here is the statement that "the pulling of the trigger" should be included within the scope of the act. Neither Austin nor Holmes, nor the Section 2 definition itself,

appears to warrant this usage, but instead would apparently restrict the act in this situation to the voluntary movement of the muscles of the hand and fingers, thereby locating the direct and immediate effects on the weapon ("the pulling of the trigger") in the domain of consequences.

Philosophers of action generally concur with this approach. They have developed a helpful terminology to distinguish voluntary bodily movements from even their immediate effects, calling the former "basic" acts (Danto 1965; Goldman 1970). Significantly, Goldman's original account of basic acts in *A Theory of Human Action* anticipates a range of subsequent cognitive science research on motor control, mirror neuron activation, action recognition, and the coordination of action comprehension and production in both human and nonhuman animals (see, e.g., de Pelligrino et al. 1992; Gallese & Goldman 1998; Rizzolatti et al. 2001; cf. Miller, Galanter & Pribram 1963). For example, Goldman (1970: 18) writes:

> A person's action often has far-reaching effects in the world, but whatever one does in the world at large must come, in one way or another, from one's body, especially from the movements of one's body. Thus, there is a central role that bodily acts play vis-à-vis our acts in general, and this special role is intended to be captured by the phrase 'basic acts' . . . . [T]he idea can be easily grasped with the aid of some examples. . . . [such as] extending one's arm, moving one's finger, bending one's knee, shrugging one's shoulder, opening one's eyes, turning one's head, puckering one's lips, wrinkling one's nose . . . .

Each of these action descriptions would likely fall under the orthodox legal definition of an act because it refers to a "voluntary muscular contraction, and nothing else" (Holmes 1881: 91). Depending on how it is interpreted, however, each description might still draw in elements that belong more properly to the category of consequences. We need not resolve this issue here, nor seek to achieve a more fine-grained analysis of the concept of an act. For our purposes, the important point is that the act element in the legal definition of harmful battery is an exceedingly narrow concept that denotes a voluntary movement of the human body, but nothing else. Formally, this understanding can be captured by substituting the concept of a voluntary bodily movement for the act element in our formula for harmful battery:

Harmful Battery $=_{df}$ [MOTION_{V}, $\text{INTENT}_{H\text{-}O}$, CONTACT_{H}]

In this formula, the subscript "V" refers to the idea of a "voluntary" motor event. Because none of the specific cases discussed in this paper turn on how voluntariness is analyzed, this concept can be left undefined for our immediate objectives. Our limited aim here is to sketch a theory of how the brain *represents* acts of harmful battery, not to make headway on complex philosophical problems of free will, determinism, compatabilism, and the like. For our purposes, then, we may assume that it is a property of the human brain to represent the movements of conspecifics and other animate agents *as if* they were freely chosen, and thus to distinguish the categories of voluntary and involuntary action. In this light, we may conclude that at least one such representation is a necessary component of the more complex representation of harmful battery.

2. *Intent*. "Intention" and related concepts ("intending," "intent," etc.) are often defined or used in a confusing way in American law to encompass what an actor knows as well as what she intends. For example, the Restatement (Second) of Torts offers the following definition:

> 8A. Intent
>
> The word "intent" is used throughout the Restatement of this Subject to denote that the actor desires to cause [the] consequences of his act, or that he believes that the consequences are substantially certain to result from it.

Likewise, the draft Restatement (Third) (2005) reads:

> 1. Intent
>
> A person acts with the intent to produce a consequence if:
>
> a) the person acts with the purpose of producing that consequence; or
> b) the person acts knowing that the consequence is substantially certain to result.

To see how far these definitions depart from ordinary usage, consider the example of a golfer who believes with substantial certainty that he cannot hit a long drive across the pond, but nonetheless tries with all his might to do so. According to the Restatements, he thereby *intends* to lose the ball in the water, even if he hits the shot of his life and clears the pond with twenty yards to spare (Finnis 1996: 243). Likewise, the trolley driver who turns his train onto a side

track, knowing that doing so will result in the death of one person on the side track, thereby has the *intention* to kill the person on the side track—even though most observers would balk at sentences such as "The man turned the train in order to kill the man on the side track," which uses an "in order to" test to trace the scope of intention (Mikhail 2011a). Finally, the definitions imply that the CEO of a company that begins a profitable new program, knowing that it will harm the environment, thereby intentionally harms the environment (Knobe 2003).

To avoid these potential anomalies, it is useful to distinguish and incorporate both prongs of the legal concept of intention into our formula for battery. For convenience, the labels "purpose-intent" and "knowledge-intent" (Jung & Levine 1987) can be used for this purpose. Consequently, two different versions of harmful battery can be formally distinguished,

$$\text{Harmful Battery}_{[\text{P-I}]} =_{df} [\text{ACT}, \text{PURPOSE}_{H/O}, \text{CONTACT}_{H}]$$

$$\text{Harmful Battery}_{[\text{K-I}]} =_{df} [\text{ACT}, \text{KNOWLEDGE}_{H/O}, \text{CONTACT}_{H}]$$

where "[P-I]" stands for purpose-intent and "[K-I]" stands for knowledge-intent.

Furthermore, if one substitutes the notion of a "goal" for purpose, the notion of "belief" for knowledge, and the notion of "voluntary bodily movement" for act, then these definitions can be rendered in a form even more amenable to psychological research:

$$\text{Harmful Battery}_{[\text{P-I}]} =_{df} [\text{MOTION}_{V}, \text{GOAL}_{H/O}, \text{CONTACT}_{H}]$$

$$\text{Harmful Battery}_{[\text{K-I}]} =_{df} [\text{MOTION}_{V}, \text{BELIEF}_{H/O}, \text{CONTACT}_{H}]$$

The second definition does not account for the "substantial certainty" aspect of knowledge-intent, and this omission raises complications that must be addressed in due course. These include actual versus constructive knowledge, subjective versus objective probabilities, probability thresholds, and a variety of other considerations. All of these topics, however, fall outside the scope of this paper. For our purposes, what these formulas usefully highlight is the basic difference between (i) goal-directed harmful or offensive contact and (ii) harmful or offensive contact that one merely believes (with some high, but unspecified, level of probability) is likely to occur. Both representations are encompassed by the legal definition of harmful

battery, but they are often conflated in the philosophical and psychological literatures and thus have been a source of needless controversy. Moreover, as I will endeavor to explain, the simplest case of the first definition, involving contact that is voluntary, goal-directed, and harmful rather than offensive, seems capable of unifying a wide range of familiar intuitions in ethics and cognitive science (see Part IV). It also seems like the best formula with which to begin considering the evolution of harmful battery intuitions in nonhuman animals (see Part V).

3. *Contact (Direct and Indirect).* Particularly because the contact element has recently become an important topic in moral psychology (see, e.g., Cushman et al. 2006; Greene et al. 2009), it is important to clarify the scope and meaning of the contact requirement for battery, and with it the meaning of "direct" and "indirect" in this context. First, the common law protection against unwanted physical contact encompasses all forms of direct (i.e., body-to-body) touching. Further, it extends not only to any part of the body but to anything that is closely attached to the body or otherwise practically identified with it (see, e.g., RST, Section 18, comment c; Prosser 1971: 34). For example, the contact requirement for battery can be met by grabbing, shoving, hitting, or kicking another person, but also by kicking the horse she is riding, striking the umbrella or cane she is holding, snatching a book or cup from her hand, flicking a towel or glove on her arm or pant leg, blowing smoke in her face, or pushing her with a pole (see, e.g., RST, Sections 18-19; Epstein 2008: 84-85).

The legal protection against unwanted contact, however, also extends to a variety of so-called indirect bodily contacts, including: (1) any touching of the person by any object or substance put in motion by the offending agent; (2) any touching of the person by an object or substance *not* put in motion by the agent, but nonetheless caused by the agent by moving another object or substance in such a way that it induces or causes the person to make contact with the first object or substance; and (3) any movement of the person's body by an indirect or mechanical means in such a way as to cause her body to come into contact with another object or substance. For example, throwing a rock at someone and intentionally hitting her for the purpose of causing harm or distress is clearly a battery, but the resulting contact is not "direct" in one obvious sense of that term. The same is true for throwing a bucket of dirty water on a pedestrian from a balcony forty feet off the ground. In each case, the harmful or offensive contact occurs at

a distance, after the object that has been propelled by the personal force of the agent travels through the air and makes contact with the recipient of the action (cf. Greene et al. 2009).

Likewise, and perhaps less intuitively, a harmful contact sufficient for battery occurs when a child pulls a chair out from under an elderly woman just as she is about to sit down, causing her to fall to the ground and suffer injuries (see, e.g., Epstein 2008: 7-8, discussing the case of *Garratt v. Dailey*). Here the agent does not actually touch the victim, nor make any contact with her. Nor does he exert any personal force upon any object that actually comes into contact with the victim. Nonetheless, a harmful or offensive contact sufficient for battery has occurred. Finally, in the trolley problems, the contact requirement is satisfied not only by grabbing and shoving a person in front of a moving train, causing the person to make harmful contact with the train, but also by (1) mechanically moving or dropping a person in front of the train, causing the train to make harmful contact with the person, or (2) mechanically moving the train, causing the train to come into harmful contact with the person.

Obviously, there must be some limit to the operation of the contact requirement. We understand, presumably as our ancestors did not and as nonhuman animals do not, that bodily movements will result in the motion of air molecules that may cause another person's body to come into contact with other these or other molecules. Even if one assumes a harmful or offensive intent, is this type of contact sufficient for battery? Without more, say the deliberate movement of a toxic substance in the direction of another person with the purpose or knowledge of causing a harmful or offensive contact, surely not. The law of battery generally presupposes a version of the intuitive physics that has driven a great deal of important research in developmental psychology in recent years (see, e.g., Spelke & Kinzler 2007).

Still, it is important to recognize that the contact requirement for battery is highly elastic and easily satisfied in a variety of trolley problems and other vignettes that have been widely discussed in the ethics and cognitive science literatures (see, e.g., Cushman et al. 2006; Fischer & Ravizza 1992; Greene et al. 2009; Kamm 1998; Unger 1996). Turning a train toward a person standing on a side track, for whatever purpose, satisfies the *prima facie* case for harmful battery, just in case the actor knows or believes with substantial certainty that this harmful contact is bound to occur. The distinguishing feature between this type of case (e.g., the Trolley or Bystander problem) and one that involves pushing or dropping a man in front of a train (e.g., the Footbridge or Drop Man problem), therefore, cannot be the occurrence of a harmful battery as

such. Rather, the distinction must be some *other* element of the agent's action plan, such as whether the battery occurs as a means (purpose-intent) or side effect (knowledge-intent).

4. *Harmful or Offensive.* The last requirements of the prima facie norm against harmful battery are (1) the intended contact must be harmful or offensive and (2) the resulting contact must be harmful. What do "harmful" and "offensive" mean in this context? Once again, for our purposes it is useful to consider how these concepts are defined in the Restatements of Torts.

Beginning with the former, the Restatement (Second) uses the word "harm" without further qualification "to denote the existence of loss or detriment in fact of any kind to a person resulting from any cause" (section 7 (b)). It uses the words "physical harm" to mean "the physical impairment of the human body, or of land or chattels [i.e., movable personal property]" (section 7(c)), and it uses the term "bodily harm" to mean "any physical impairment of the condition of another's body, or physical pain or illness" (section 15). Furthermore, bodily harm encompasses any physical impairment of a person's body, including not only physical pain or illness, but also any physical alteration of the body's normal structure or function to any extent.

Turning to the meaning of offensive, this concept is clearly less amenable to precise definition. The Restatement's approach is thus to appeal to an objective standard of reasonableness in characterizing it. Section 19 reads: "A bodily contact is offensive if it offends a reasonable sense of personal dignity." This (circular) definition is then elaborated as follows:

> In order that a contact be offensive to a reasonable sense of personal dignity, it must be one which would offend the ordinary person and as such one not unduly sensitive as to his personal dignity. It must, therefore, be a contact which is unwarranted by the social usages prevalent at the time and place at which it is inflicted (Section 19, Comment a).

Finally, the Restatement offers this caveat, which addresses the issue of abnormal sensitivity:

> The Institute expresses no opinion as to whether the actor is liable if he inflicts upon another a contact which he knows will be offensive to another's known but abnormally acute sense of personal dignity.

All of these explanations are somewhat vague and unsatisfying, but they are obviously intended to recognize and incorporate the existence of prevailing social norms concerning what constitutes acceptable and unacceptable touching into the legal definition of offensive contact, and thus into the definition of harmful battery itself. An actor who intends to cause an offensive contact, but

instead causes a harmful contact, with a person is subject to liability for harmful battery, despite lacking any intent to harm; this is the critical point of how offensive intent operates in the law of harmful battery. Once again, we need not go beyond these brief remarks here, because none of the cases examined here turn on how the concept of "offensive" is analyzed. Philosophers such as Feinberg (1985) have examined this topic at great length, showing that commonly held intuitions about offensive conduct can be surprisingly subtle and intricate. A more granular analysis along these lines is unnecessary for our purposes, however, and thus can be set aside.

III.

Thus far our focus has been the *prima facie* case of harmful battery—the basic norm in both morality and law that holds intentionally harmful contact to be presumptively wrong, that is, impermissible unless justified. Before turning to the widespread investigation of this norm in the ethics and cognitive science literatures, it is necessary to round out the foregoing analysis by briefly considering the primary defenses to battery in both morality and law. These defenses introduce more complexity into the analysis of harmful battery, for they call upon the application of moral imagination, perspective-taking, counterfactual and probabilistic reasoning, and norms of reasonableness to various stipulated or imagined circumstances. All of these mental operations are, of course, more difficult to formalize than the three-element tort of battery outlined above. Common defenses to battery that all normal adults and children beyond a certain age intuitively apply in everyday moral reasoning thus demand more complex cognitive capacities than are implied by a simple formula such as "voluntary, goal-directed harmful contact." Whereas the latter points us toward an aspect of human psychology that seems likely to be shared with nonhuman animals, these defenses reveal what is—or at least seems more likely to be—distinctively human (cf. Darwin 1981/1871).

One familiar defense to battery is consent. In the law of torts, this typically comes in at least three varieties: express consent, implied consent, and hypothetical consent. Express consent occurs when the recipient of a harmful bodily contact explicitly agrees to that contact or at least to the risk of incurring it. Boxing, football, hockey, and other high-contact sports are common examples. The participants in these athletic contests do not commit batteries merely by intentionally inflicting harmful contacts on one another. This conclusion follows, of course, from the fact that these acts of touching are consensual. The scope of consent, however, must be

respected, as must the established rules on what bodily contacts are permissible. Thus a fistfight or illegal tackle in a football or hockey game can, under some circumstances, constitute battery. This is true whether the context is a professional or scholastic sporting event or even something much less organized (e.g., "pick-up" basketball, or backyard game of kick-the-can).

The scope of the express consent in the context of medical procedures also must be respected, if these bodily contacts (which are usually harmful to some extent) are to avoid constituting batteries. This qualification immediately implies the capacity to express and recognize the scope of consent, a non-trivial matter in many circumstances. Moreover, whether the scope of consent has been exceeded can often turn decisively on facts about unobservable mental states. Thus, the exact same physical movements performed by a nurse or physician can be a routine physical examination or an act of sexual battery or molestation, depending on the state of mind of the actor and on the precise scope of the consent given by the recipient. Picking out an act of battery can thus sometimes require subtle acts of perspective-taking and mind-reading. And yet, if recent findings in experimental ethics and cognitive science are an accurate guide, it appears to be something that every "person beyond a certain age and possessed of the requisite intellectual capacity" (Rawls 1971: 46) learns to do as a matter of course. Not so with nonhuman animals, however; even after decades of intensive research, their capacities for mind-reading, at least at this level of complexity, remain in doubt (Penn & Povinelli 2007).

Implied consent is a more complex notion than express consent. It typically turns on whether a reasonable person would infer consent on the basis of the voluntary conduct (including speech acts) of the recipient of an alleged battery. Voluntarily standing in line and extending one's arm when the time comes to get vaccinated is the paradigmatic example in the law of torts; the doctor who subsequently pricks the owner's arm with a needle, thus knowingly inflicting a harmful bodily contact, does not thereby commit a battery, even if the recipient of this action privately does not want or agree to get immunized. Her overt bodily movements and what a reasonable person would infer on that basis are controlling as to the issue of consent (see, e.g., the case of *O'Brien v. Cunard Steamship Co*., with roughly this fact pattern).

Hypothetical consent is more complex still, depending not on any actual conduct or any express or implied agreement on the part of the recipient of an alleged battery, but merely on what a reasonable person in her circumstances *would* have agreed to, if she had been asked. The reason why an unconscious accident victim who is operated upon after being rushed into an

emergency room is not battered by the attending physicians is that her hypothetical consent to being touched and handled in this manner is assumed. She, or at least a reasonable person in her circumstances, would have agreed to these contacts, if asked—or so at any rate the law infers.

Hypothetical consent has been tested in the context of trolley problem experiments in a type of "reverse footbridge" condition, in which a man must be thrown out of the path of an onrushing train in order to save him from being crushed and killed by it. If the actor believes with substantial certainty that this act of "throwing the man" will cause him to come into harmful contact with the ground, is throwing him out of the path of the train a harmful battery? No: One can assume that the man would consent to be thrown and, if necessary, harmed in this manner in order to save his life or prevent a greater harm. Consistent with this analysis, the vast majority of experimental subjects find this action to be permissible, even though the action is described in the question eliciting this response in exactly the same manner as the standard footbridge problem (e.g., "Is it morally permissible for Luke to throw the man?"). Again, it seems doubtful whether any nonhuman animals engage in complex acts of moral reasoning like this, or even have the requisite mental capacities to do so. Thus, what is distinctively human about human moral cognition, at least with respect to harmful battery, appears to emerge at this level of analysis—the level of justification—rather than in relation to the *prima facie* case.

The foregoing remarks merely scratch the surface of how consent typically functions as a defense to battery in the law of torts. The subject is complex and occupies its own chapter of the first Restatement (Sections 49-62, comprising 21 pages with multiple illustrations). A similar characterization can be given of other common defenses to intentional battery, including the use of force in self-defense or defense of others (Sections 63-76), the use of force in defense of property (Sections 77-86 and 87), the use of force to effect an entry upon land (Sections 88-99), the forcible taking of chattels (Sections 100-111), and the privilege to use force or impose physical restraint in order to effect an arrest, prevent the commission of a crime, or serve legal process (Sections 111-145). The privilege to use reasonable force to enforce valid military orders (Section 146) or to discipline children (Sections 147-155) are also given special treatment, as are certain other force-based justifications. For its part, the American Law Institute's Model Penal Code (MPC) addresses many of the same issues in the context of the criminal law. For example, the MPC supplies detailed doctrines of justification and excuse under the categories of necessity/choice of evils (Section 3.02), execution of public duty or military orders (Section

3.03), self-protection (Section 3.04), protection of others (Section 3.05), protection of property (Section 3.06), law enforcement (Section 3.07), duress (Section 2.09), intoxication (Section 2.08), and mental disorder/diminished capacity (Section 4.01).

With some important exceptions, all of these criminal law and tort doctrines appear to track common moral intuitions, at least to a good first approximation. On this point, Holmes (1897: 459) was surely correct when he referred to the law as "the witness and external deposit of our moral life." By postulating unconscious knowledge of these legal rules, one can explain the intuitive moral data across an enormously wide and diverse range of cases. The precise character and extent of this knowledge are potentially difficult issues that warrant careful investigation, but the general thesis seems sound and could presumably be confirmed by controlled experiments. If so, then these bodies of law supply valuable evidence of the properties of human moral psychology, and what precisely may be distinctive about it.

IV.

Clarifying fundamental elements of moral cognition is not the only reason for analyzing how the brain computes representations of harmful battery. Another important objective is to explain and unify some familiar moral intuitions in ethics and cognitive science. For the past few decades, philosophers and psychologists alike have been relying on harmful battery cases and studying their effects without clearly acknowledging this fact or considering its implications for issues of experimental design, data interpretation, or theory construction. In this part, I first defend this observation by identifying some famous thought experiments in ethics that appear to rest on harmful battery intuitions. I then discuss some recent psychology experiments to which the same generalization seems to apply.

The examples I discuss have not been grouped together previously under this or any other legal rule. Part of the purpose in collecting them in this manner is thus to suggest that they can, in fact, be classified as members of a single, unified set. If so, then it follows that a range of superficially distinct cases drawn from diverse fields can be explained with reference to a single mental rule or algorithm, which different researchers appear to be stimulating for different objectives. The common denominator is intentional, harmful contact, inflicted without consent or other justification.

A. Ethics

1. Foot's cave explorer, capital punishment, transplant, and torture cases.

Foot (1967) famously conjured up the trolley problem and a class of similar dilemmas to illustrate the plausibility of the DDE and the type of non-utilitarian moral reasoning it warrants. Significantly, Foot did not explicitly refer to battery in her article, opting instead to consider these dilemmas in terms of "higher-order" concepts and distinctions: killing v. letting die, intending v. foreseeing, doing v. allowing, and negative v. positive rights. Many of her original examples, however, involve acts of harmful battery, and the resulting intuitions can be succinctly explained in these terms. Recall, for example, the memorable cases of the fat man stuck in the cave, who can be dislodged only by exploding a stick of dynamite; the judge who frames and executes an innocent man in order to prevent a bloody riot; the five patients who can be saved only by transplanting organs from one healthy individual; and the tyrant who threatens to torture five innocent persons if we do not agree to torture one (1967: 61-63). These cases form a varied lot, but one common denominator of the actions in this series is that each constitutes a purposeful harmful battery. That is, each case involves an intentional harmful contact, inflicted without consent or other justification. By contrast, the common element (or at least one common element) of the permissible actions in the companion series of cases Foot describes—turning a trolley away from five onto one, directing a scarce drug to five less seriously injured patients, and so on (1967: 62-64)—is that none of them involves a purposeful harmful battery. A parsimonious explanation could thus appeal to a rule against purposeful harmful battery, together with a negative utilitarian principle of justification, to explain common reactions to these cases.[2]

2. Rachels' bathtub cases.

Rachels (1975) denied the existence of a morally significant distinction between killing and letting die on the basis of another famous pair of cases. In the first case, Smith, who wants his young cousin dead in order to gain a large inheritance, deliberately drowns the child while he is taking a bath. In the second case, Jones, who likewise wants his cousin dead for the same reason, deliberately lets the child drown, even though he could have easily saved him. Rachels (1975: 114) argued that "the only difference" between the two cases is that "Smith killed the

child, whereas Jones 'merely' let the child die." Thus, the two actions are equivalent from a moral point of view, thereby undermining a distinction between active and passive euthanasia.

Rachels is surely correct to hold that in some cases the means to achieve a bad outcome are so insignificant in relation to that outcome that it seems absurd to draw a meaningful distinction on that basis. Still, he seems mistaken to assert that the only difference between his two cases is between killing and letting die. In light of our analysis, at least two further differences exist. First, Smith commits a harmful battery, whereas Jones does not. Second, only Smith's conduct would trigger criminal or civil liability. That this last statement seems shocking may owe partly to the fact that, as Rachels describes it, Jones' failure to rescue constitutes a type of *purposeful* homicidal omission, whereas most failures to rescue do not.[3] Further, there may be compelling reasons for finding the two actors equally reprehensible and concluding partly on this basis that active and passive euthanasia should be treated equivalently in law. My objective here is not to enter into these debates, but merely to note an apparent fallacy in Rachels' argument, as it relates to this paper's main thesis. Despite his assertion to the contrary, Rachels' bathtub cases *can* be distinguished on grounds other than the distinction between killing and letting die. The outcomes are the same, but the means are different, only one of which violates an independent norm against harmful battery.

3. Nagel's "accident on a lonely road"

Nagel (1986: 171) invites his reader to imagine an individual who is driving on a lonely road one evening and gets into a bad automobile accident. Because the other passengers are badly injured, the driver seeks help and manages to find an isolated house nearby that is occupied by a woman and her grandchild. There is no phone, however, and when the driver asks the woman if he can borrow her car, she becomes frightened and refuses. Terrified by the driver's evident desperation, the woman runs upstairs and locks herself in the bathroom, leaving the driver alone with the child. After pounding on the door and searching for the keys without success, the driver suddenly has an idea. Perhaps the woman might be persuaded to give him the keys if he were to twist the child's arm outside the bathroom door, causing her to cry in pain?

Nagel denies that it would be permissible to twist the child's arm in these circumstances, even though failing to do so might result in worse consequences for the accident victims. He assumes that common moral intuitions would support this conclusion. Whether or not his

prediction is accurate as a matter of experimental ethics, the important point for us is that the action Nagel invites his reader to evaluate is, once again, a deliberate and purposeful battery. Nagel's question can thus be put this way: Is it permissible to commit harmful battery as a means to achieve a good or worthwhile end? Understood in these terms, Nagel's case implicates the same underlying representation as the footbridge problem, Foot's cases, Rachels' first case (Smith), and the other cases of purposeful harmful battery discussed in this paper.

4. Boorse & Sorenson's "duck/shield" cases

Boorse & Sorenson (1988) contrast cases of what they call "ducking" with cases of what they call "sacrificing" or "shielding." In the former, an agent moves his own body (but no one else's body) in order to avoid an imminent harm to himself from an independent agent or instrument, knowing that by doing so he will expose a third party to the same harm. In the latter, the agent makes contact with or moves another person's body for the same purpose, again knowing that the same third-party harm will result.[4]

The authors illustrate their distinction with two examples. First, there is the comical story of two campers, Alex and Bruce, who come upon a ravenous bear:

> As Alex grabs his running shoes, Bruce points out that no one can outrun a bear. "I don't have to outrun him," Alex replies. "I only have to outrun you." Few contemporary Westerners will criticize Alex for running away at full tilt, or even for using his new Sauconys. But suppose Alex instead ties Bruce's ankles, or knocks Bruce unconscious and throws him to the bear. Alex is now blameworthy in ethics and law. The result is the same in both cases: Bruce's death. Further, Alex may know the result to be the same if he knows he can outrun Bruce. Nonetheless, most people sharply distinguish the two acts (1988: 79).

Second, there is the case of a gunman who shoots and kills his victim in one of two conditions:

> <u>Condition 1</u>
> Angela, at the end of a movie ticket line, sees [the gunman] about to shoot a .22 automatic at her. Angela knows that a .22 bullet will kill one person but not two. Angela leaps aside; the bullet kills Brenda, who is next in line.
>
> <u>Condition 2</u>
> Same as Condition 1, but Angela grabs Brenda and moves her in front as a shield; the bullet kills Brenda (1988: 79).

In an early venture into experimental philosophy, Boorse and Sorensen (1988: 88) tested their students' intuitions on these cases, reporting that 84 percent "said that ducking the bullet or

bear was not immoral and that the sacrificial alternative was immoral. If we exclude those who gave different verdicts on ducking bullets and bears, support for the distinction rises to 91 per cent." From our perspective, these results are better stated in terms of a norm against harmful battery. The studies by Boorse and Sorensen confirm that most people recognize an intuitive difference between committing harmful battery as a means to an end and achieving the same end without committing a purposeful harmful battery. Thus, their findings can be predicted and explained by the same formula that explains the other cases of purposeful harmful battery discussed in this paper. There is no need to invoke new concepts like "ducking" or "shielding."

B. Cognitive Science

5. Cushman's three principles of harm

Turning to similar examples from cognitive science, here, too, one finds a widespread reliance on harmful battery intuitions, which has gone unremarked in the literature. Expanding on early successes with trolley problems, many cognitive scientists use harmful battery cases in their experiments without clearly acknowledging this fact or considering what it might imply for issues of experimental design, data interpretation or theory construction. Simple and powerful generalizations are thus being neglected in favor of less adequate and more partial explanations.

A useful illustration of this phenomenon is the influential paper by Cushman et al. (2006) testing "three principles of harm." Building on prior research that sought to explain trolley intuitions in terms of the prohibition of intentional battery, Cushman and colleagues constructed a series of cases, each of which isolated and tested intuitive reactions to one of three key elements of harmful battery: action, intention, and contact (cf. Part II). Here is how these elements are described (2006: 1092):

- *The action principle:* Harm caused by action is morally worse than harm caused by omission.
- *The intention principle*: Harm intended as a means to a goal is morally worse than equivalent harm foreseen as the side effect of a goal.
- *The contact principle*: Using physical contact to cause harm to a victim is morally worse than causing equivalent harm to a victim without using physical contact.

Significantly, the authors found that adults make moral judgments in conformity with these principles across a large database of superficially distinct fact patterns. For example, in the

following cases, adults judge the action constituting purposeful battery (Connor) to be morally worse than the companion case (Mike) in which an element of this *prima facie* case is missing:

> Aquarium (Conner)
> Conner is at a new aquarium exhibit when he sees a visitor slip on a wet floor, fall down, and break his neck. The visitor is still alive and can be safely evacuated by medics so long as he is not moved. He has fallen, however, on top of the oxygen supply line servicing five other visitors in an underwater observation pod. Without oxygen, the five visitors will soon die. If Connor does nothing the one visitor will be safely evacuated, but the five visitors in the pod will die. If Connor pushes the one visitor off the supply line this one visitor will die, but the five visitors in the pod will have their oxygen restored and will live. Pushing the one visitor is: [permissibility judged on a 1-7 scale]
>
> Aquarium (Mike)
> Mike is at a new aquarium exhibit when he sees a visitor slip on a wet floor, fall down, and break his neck. The visitor is still alive and can be safely evacuated by medics so long as he is not moved. He has fallen, however, on top of the oxygen supply line servicing five other visitors in an underwater observation pod. Without oxygen, the five visitors will soon die. If Mike does nothing the one visitor will be safely evacuated, but the five visitors in the pod will die. If Mike pulls the supply line out from under the one visitor this one visitor will die, but the five visitors in the pod will have their oxygen restored and will live. Pulling the supply line is: [permissibility judged on a 1-7 scale]

In these cases, Connor commits a harmful battery to achieve his good end, whereas Mike does not. The same distinction runs through many of the novel case-pairs devised by Cushman et al (2006). Although the term "battery" is not used in their study, their central findings thus directly support the main thesis of this paper. The key elements of the simplest case of harmful battery—voluntary, goal-directed, harmful contact—appear to function as critical building blocks of human moral cognition, a result that can easily be demonstrated experimentally. The rules and concepts implied by this norm thus appear to be psychologically "real," comprising significant components of moral-cognitive architecture.

6. Greene's "personal force" experiments

An influential paper by Greene and colleagues (2009) serves as another illustration of the same general phenomenon. The main aim of the paper is to integrate two lines of research on the pattern of moral judgment elicited by the trolley problems. The paper thus begins by identifying

two factors that may explain this pattern, "(1) the agent's intention . . . and (2) whether the agent harms the victim in a manner that is relatively 'direct' or 'personal,'" (Greene et al., 2009: abstract, 2). The paper then reports two experiments designed to explore the interaction of these factors with a novel factor, which the authors call "personal force" (2009: 2). The authors document the influence of this third factor on moral judgment, as distinct from that of physical contact and spatial proximity. In addition, they discover an interaction between personal force and intention, finding that "the personal force factor only affects moral judgments of intended harms, while the intention factor is enhanced in cases of personal force" (2009: 6).

Greene's definitions of physical contact, personal force, and other concepts are unusual and leave much to be desired from a modern legal or scientific standpoint. For example, Green defines "physical contact" in such a way that only immediate body-to-body contacts qualify. Shoving someone with a pole or hitting her over the head with a baseball bat, therefore, do not count as physical contacts on his account. Likewise, Greene distinguishes *muscular* and *mechanical* force and limits the scope of "personal force" to the former. Shooting someone with a gun, therefore, does not involve the application of personal force to the victim, on his analysis.[5]

Setting these considerations aside, the important point to recognize is that like virtually all of the personal dilemmas used by Greene in his original fMRI experiment (Fig. 1), all of the controlled case-pairs used in this study appear to involve acts of harmful battery. Some of these batteries are purposeful, while others are not; and this distinction can be used to explain much of the reported data. In the following variations of the Speedboat Dilemma, for example, Greene et al. (2009) report no significant difference in how the first two scenarios are evaluated. By contrast, the subjects in Greene's experiments evaluate the third scenario differently:

<u>Speedboat #1 (push with hands)</u>

> John is driving a small speedboat when he notices five swimmers drowning in the distance. If John does not drive toward them at top speed, he will not arrive in time, and all five will die. In order to drive at top speed, John must lighten the load on his boat. The only way to lighten the load is to push his passenger with his hands, causing the passenger to tumble off the back of the boat. This passenger cannot swim and will drown. If John pushes the passenger, the one passenger will drown, but John will save the five drowning swimmers. If John does not push the passenger, the one passenger will not drown, but the five swimmers will drown. Pushing the man is [permissibility judged on a 1-7 scale].
>
> <u>Speedboat #2 (push with oar)</u>
> John is driving a small speedboat when he notices five swimmers drowning in the distance. If John does not drive toward them at top speed, he will not arrive in time, and all five will die. In order to drive at top speed, John must lighten the load on his boat. The only way to lighten the load is to push his passenger with an oar, causing the passenger to tumble off the back of the boat. This passenger cannot swim and will drown. If John pushes the passenger, the one passenger will drown, but John will save the five drowning swimmers. If John does not push the passenger, the one passenger will not drown, but the five swimmers will drown. Pushing the man is [permissibility judged on a 1-7 scale].
>
> <u>Speedboat #3 (accelerate)</u>
> John is driving a small speedboat when he notices five swimmers drowning in the distance. If John does not drive toward them at top speed, he will not arrive in time, and all five will die. In order to drive at top speed, John must lighten the load on his boat. The only way to lighten the load is to accelerate quickly, causing the passenger to tumble off the back of the boat. This passenger cannot swim and will drown. If John pushes the passenger, the one passenger will drown, but John will save the five drowning swimmers. If John does not push the passenger, the one passenger will not drown, but the five swimmers will drown. Accelerating quickly is [permissibility judged on a 1-7 scale].

Because the law of battery does not recognize a distinction between pushing someone with one's hands and pushing him with a pole, it would not distinguish the first two cases. By contrast, the third case is distinguishable, insofar as it lacks the element of harmful contact and includes one fewer count of purposeful harmful battery. Thus, the law of battery can predict Greene's data in these cases and many other cases reported in his paper without relying on his idiosyncratic and counterintuitive definitions of physical contact, personal force, and the like. There is no need to reinvent the wheel.

7. Turiel's moral-conventional distinction

Turning to research in moral development, the moral-conventional distinction pioneered by Elliot Turiel and his colleagues (see, e.g., Nucci & Turiel 1978; Smetana 1981, 1984; Turiel 1983) also can be usefully understood from the perspective of harmful battery and its elements. As is well-known, Turiel and colleagues proposed a distinction between two types of violations, moral and conventional, which both children and adults reliably differentiate. They found that moral violations (e.g., one child hitting another) are widely held to be (1) more serious, (2) more deserving of punishment, (3) authority-independent and (4) universally applicable. By contrast, conventional transgressions (e.g., wearing pajamas to school) do not possess these properties.

Numerous follow-up studies have found that young children reliably draw a moral-convention distinction and do so in culturally diverse settings (see, e.g., Blair 1995; Hollos et al. 1986; Nucci et al. 1983; Smetana et al. 1984; Smetana et al. 1999; Turiel et al. 1987; Yao & Smetana 2003). The distinction has thus become a fixed feature of philosophically-sophisticated theorizing about moral development and moral cognition (see, e.g., Nichols 2004). What has largely escaped notice until now, however, is the fact that most of the violations classified as "moral" violations in the moral-conventional paradigm constitute purposeful harmful batteries. In one leading study, for example, these transgressions are described as acts of "hitting, kicking, biting, punching, pulling hair, or otherwise intentionally physically harming another child" (Smetana 1984: 1769). All of these actions fall under the simplest formula for harmful battery outlined in Part II—voluntary, goal-directed, harmful contact—and thus can be explained on this basis. By contrast, *none* of the cases that have been classified as conventional violations in the moral-conventional paradigm appear to involve purposeful harmful batteries (see, e.g., Nucci & Turiel 1978; Smetana 1981, 1984; Turiel 1983). The basic finding of the moral-conventional paradigm thus can be explained by assuming a keen sensitivity to battery as an innate property of the human mind, which emerges early in child development. One need not appeal to the higher-order (and more controversial) distinction between two different normative domains, moral and conventional, at least with respect to this data set.[6]

8. Hamlin et al.'s helper-hinderer experiments

Consider finally the celebrated finding by Hamlin and colleagues (2007) that 6- and 10-month-old infants appear to evaluate others based on the moral quality of their acts. The basic design of these experiments involved a red wooden circle with large eyes glued onto it (the 'climber'), who repeatedly attempts to climb a hill and is either aided up by a yellow circle (the 'helper'), who pushes it up from below, or a blue square (the 'hinderer'), who pushes it down from above. A third condition involved a neutral character who traces a path identical to that of the helper or hinderer. Given an opportunity to choose, infants are drawn towards helpers and inclined to avoid hinderers, choices that are not based on general perceptual preferences, which were carefully controlled. "Infants' preference for the helper and aversion to the hinderer," the authors (2007: 558) conclude, "are best explained as specifically social evaluations: a liking for those who act cooperatively to facilitate the goals of others, and a dislike of those who impede another's goals."

From the perspective of this paper, the interpretation of Hamlin's research should be modified to take into account the following considerations. First, the fact that the climber in these experiments is either bumped up the hill by the helper or bumped down the hill by the hinderer means that the occurrence of a harmful or offensive *contact* lies at the heart of these studies. What infants encounter in these experiments is thus not simply an actor who facilitates or impedes another individual's goals, but rather an agent who does so by means of physical contact with that individual that a harmful battery norm would, presumably, recognize as consensual or nonconsensual in each case. To put the matter differently, if one assumes that the simplest of the harmful battery formulas outlined in Part II is operative in the minds of these subjects, then it seems possible to predict and explain their behavior without assuming that such rich semantic notions as "helper" and "hinderer" are involved.

Second and along the same lines, as a general matter there are many ways in which one can help or hinder someone without touching her—for example, knowingly giving her accurate or false directions, respectively, when she is lost and asks for help. Thus, notions like "helper" and "hinderer" will range across a wide variety of cases, only some of which involve harmful or offensive contact and thus implicate the harmful battery norm. Furthermore, the fact that parents frequently hinder their young children's own navigation or otherwise frustrate their immediate goals (often for reasons of safety) without adverse social evaluations of the kind elicited by these experiments suggests that these reactions may not be tethered simply to the activities of helping

and hindering as such. Instead, however surprising it may be, it seems plausible to assume that some primitive notion of *consent* may be operative in these cases. In sum, like the other cases we have reviewed, Hamlin et al.'s (2007) helper-hinderer experiments appear to lend support to the main argument of this paper, insofar as they suggest that a harmful battery norm may emerge and help to structure the child's moral and social perceptions at a surprisingly early age.

V.

The observations of Part IV suggest that many famous thought experiments in ethics and many influential experiments in moral psychology rely on harmful battery scenarios without clearly acknowledging this fact or considering what it implies for issues of experimental design, data interpretation, or theory construction. The significant generalization that has escaped notice is an acute sensitivity to battery as a property of the human mind (Figure 2).

There are many lessons to be drawn from these facts, and many reasons for clarifying the elements of battery and the precise representations they presuppose. Perhaps the most intriguing is that doing so enables us to consider in a new light Darwin's (1981/1871: 71-72) famous observation that "any animal whatever" would acquire a moral sense or conscience once "its intellectual powers had become as well developed, or as nearly developed, as in man." Battery is a core human moral wrong, which operates as a lesser-included offense of a wide range of more significant acts of aggression—rape, aggravated criminal assault, domestic violence, sexual molestation, genital mutilation, water-boarding and other paradigmatic acts of torture, and virtually all acts of murder and other unlawful homicides, to name a few examples. (Torture or murder can be committed without an underlying battery, but most cases of torture and murder do include battery as a lesser-included offense.) By clarifying the elements of ordinary harmful battery, one can begin to appreciate both how much and how little cognitive machinery is required for the brain to generate these representations in various contexts and circumstances.

What does a creature need to be able to perceive and recognize this species of wrongful conduct? For the *prima facie* case of harmful battery, only three main components: an action element, an intention element, and a harmful contact element (cf. Cushman et al. 2006; Mikhail 2000). In its simplest form, the first reduces to the concept of a voluntary motion detector that applies to the bodily movements of conspecifics (or other agents). The second reduces to the

ability to perceive and track goal-directed action (and thus may not even require "theory of mind" abilities at all, as they are conventionally understood; cf. Penn & Povinelli 2007). Finally, the third reduces to the ability to perceive and keep track of harmful bodily contacts with conspecifics (or other sentient creatures), including oneself, in what may be the standard case. From this perspective, one can easily imagine how "any animal whatever" possessing the requisite mental capacities might be capable of performing these relatively simple tasks.

The picture appears different, however, when one turns to the second half of the ability to compute battery representations: the capacity to generate justificatory norms and to apply them in complex and factually diverse circumstances. Here it seems reasonable to infer, at least on the basis of the available evidence, that a more accurate formula than Darwin's famous slogan would be "no other animal whatever." Only humans appear to engage in these complex acts of moral reasoning.

Of course, the experimental protocols currently in vogue in the field of animal cognition do not, as a general matter, seek to investigate the cognitive capacities or mental operations to which I refer. Unlike a "theory of mind," there is no literature on whether the chimpanzee has a "theory of battery," a "theory of negligence," or a "theory of transferred intent" and the like; nor, as far as I am aware, have researchers investigated whether chimpanzees or other nonhuman animals draw inferences about harmful contacts on the basis of a "reasonable chimpanzee" standard (compare Premack & Woodruff 1978 and the vast literature it has spawned). It seems possible, therefore, that once scientists begin focusing their attention on topics like these, their discoveries may force us to revise our understanding of the moral psychology of nonhuman animals in fundamental ways. If so, then the continuity thesis asserted or implied by Darwin and other prominent ethical naturalists, such as Hume, Proudhon, Kropotkin, and De Waal, might be reinforced and enhanced as a result of these investigations.

VI.

This paper has argued that any important research programs in ethics and cognitive science use harmful battery scenarios in their experiments without clearly acknowledging this fact or considering its implications. Significant generalizations are thus being overlooked in favor of less adequate and more partial explanations. The *prima facie* case of harmful battery is itself quite simple, and it seems reasonable to assume that its basic cognitive components—

voluntary act, harmful intention, harmful contact—are grasped by nonhuman animals. The case seems different with respect to justifications to battery, such as consent, self-defense, or (in the case of a standard set of trolley problems) the join operation of a means-constrained principle of necessity and a negative utilitarian justification of foreseen but unintended harm. More research is needed to sharpen and test these hypotheses, but they seem plausible in light of the available evidence.

We can draw at least two further lessons from the basic argument of this paper. First, researchers who study animals or who work on the evolution of morality often focus on complex, socio-emotional components of moral psychology, such as empathy, altruism, fairness, and reciprocity. Virtually no one studies simpler and more easily-defined capacities such as battery perception. Yet, as we have seen, the competence for battery perception appears to be stable, robust, and easily elicited across a wide range of populations. A simplified research strategy that rigorously investigates the operation of this single, well-defined norm in a variety of contexts thus might be the key to uncovering basic building blocks of moral cognition, thereby clarifying the significance and plausibility of Darwin's famous conjecture. Second, philosophers and psychologists who are interested in a naturalistic ethics often look to biology for inspiration. It is thought that biology provides an important missing source of information about human morality "because the process of organic evolution that gave rise to all forms of life has been left out of the discussion" (Alexander 1987: xvii, quoted in Machery & Mallon 2010: 3). This may be true; but it seems equally important, if not more so, to emphasize that what has frequently been left out of the discussion is the law.

References:

Alexander, R. D. (1987). *The Biology of Moral Systems*. Hawthorne, NY: Aldine de Gruyter.

Anscombe, G. E. M. (1957). *Intention.* Oxford: Basil Blackwell.

Austin, J. (1879/1861). *Lectures on Jurisprudence*. London.

Beale, J. (1920). The proximate consequences of an act. *Harvard Law Review*, 33, 633-xxx.

Bennett, J. (1995). *The Act Itself*. Oxford: Clarendon Press.

Blair, J. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57, 1–29.

Boorse, C., & Sorensen, R. (1988). Ducking harm. *Journal of Philosophy* __; reprinted in J. M. Fischer & M. Ravizza, (Eds.) (1992), *Ethics: Problems & Principles* (pp. 77-91).

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pelligrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive Affective Neuroscience*, 2(2), 84-92.

Cushman, F., Young, L. & Hauser, M. (2006). The role of reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science*, 17(2), 1082-1089.

Danto, A. (1965). Basic actions. *American Philosophical Quarterly, 2*, 141-48.

Darwin, C. (1981/1871). *The Descent of Man, and Selection in Relation to Sex*. Princeton, NJ: Princeton University Press.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.

De Waal, F. (1996). *Good Natured: The Origin of Right and Wrong in Humans and Other Animals.* Cambridge, MA: Harvard University Press.

------. (2006). *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.

di Pellegrino, G., Fadiga, L. Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res*., 91, 176-80.

Donagan, A. (1977). *The Theory of Morality*. Chicago: University of Chicago Press.

Epstein, R. (2008). *Cases and Materials on Torts* (9th edition). New York: Aspen.

Feinberg, J. (1985). *The Moral Limits of the Criminal Law, Vol. 2: Offense to Others*. New York: Oxford University Press.

Fischer, J.M. & Ravizza, M. (1992). *Ethics: Problems and Principles*. Fort Worth: Harcourt Brace Jovanovich.

Finnis, J. (1996). Intention in tort law. In *Philosophical Foundations of Tort Law* (D. Owen, ed.). Oxford: Clarendon Press.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. In J. M. Fischer & M. Ravizza (1992), *Ethics: Problems and Principles* (pp. 60-67).

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends. Cog. Sci.,* 2, 493-501.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165-93.

Goldman, A. (1970). *A Theory of Human Action*. Princeton: Princeton University Press.

Greene, J. (2005). Cognitive neuroscience and the structure of the moral mind. In S. Laurence, P. Carruthers, & S. Stich (Eds.), *The Innate Mind, Vol. 1: Structure and Contents.* New York: Oxford University Press.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517–523.

Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral Judgment. *Science,* 293, 2105–2108.

Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural basis of cognitive conflict and control in moral judgment. *Neuron*, 44, 389-400.

Greene, J., Cushman, F., Stewart, L., Lowenberg, K., Nystrom, L., & Cohen, J. (2009). Pushing moral buttons: The interaction of personal force and intention in moral judgment. *Cognition*, 111 (3), 364-371.

Hall, J. (1960). *General Principles of Criminal Law*. Indianapolis: Bobbs-Merrill.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450, 557-559.

Hart, H. L. A. & Honore, A. M. (1959). *Causation in the Law*. Oxford: Oxford University Press.

Holmes, O. W. (1881). *The Common Law* (1991 ed.). New York: Dover Publications

------. (1897). The path of the law. *Harvard Law Review*, __.

Hollos, M., Leis, P., & Turiel, E. (1986). Social reasoning in Ijo children and adolescents in Nigerian communities. *Journal of Cross-Cultural Psychology*, 17, 352-76.

Huebner, B., Hauser, M., and Petit, P. (forthcoming). How the source, inevitability, and means of bringing about harm interact in folk moral judgments.

Jung, D, & Levine, D. (1987.) Whence knowledge intent? Whither knowledge intent? *U.C. Davis Law Review*, 20, 551-xx.

Kamm, F. M. (1998). Moral intuitions, cognitive psychology, and the harming versus-not-aiding Distinction. *Ethics, 108,* 463-488.

Kelly, D., Stich, S. Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect, and the moral-conventional distinction. *Mind & Language,* 22, 117-131.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63,* 190-193.

------. (2005). Moral cognition and theory of mind: Exploring the connections. *Trends in Cognitive Sciences, 9,* 357-359.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to ventromedial prefrontal cortex increases utilitarian moral judgments. *Nature,* 446, 908–911.

Lawrence, R. (1972). *Motive and Intention: An Essay in the Appreciation of Action.* Evanston: Northwestern University Press.

Machery, E. & Mallon, R. The evolution of morality. In J. Doris, ed., *The Moral Psychology Handbook.* Oxford: Oxford University Press.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation.* Oxford: Clarendon Press.

Mikhail, J. (2000). Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in "A theory of justice." Unpublished PhD Dissertation, Cornell University.

——. (2011a). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.

——. (2011b). Emotion, neuroscience, and law: A comment on Darwin and Greene. *Emotion Review,* 3, 293-95.

Miller, G.A., Galanter, E., & Pribram, K. (1986/1963). *Plans and the Structure of Behavior*. Adams Bannister Cox Pubs.

Moore, A., Clark, B., & Kane, M. (2008). Who shall not kill? Individual differences in working memory capacity, executive control, and moral judgment. Psychological Science, 19(6), 549-557.

Nagel, T. (1986). *The View from Nowhere.* Oxford: Oxford University Press. In Fischer, J.M. & Ravizza, M., (Eds). (1992). *Ethics: Problems and Principles* (pp. 165-179).

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment.* Oxford: Oxford University Press.

——. (2005). Innateness and moral psychology. In S. Laurence, P. Carruthers, & S. Stich

(Eds.), *The Innate Mind, Vol. 1: Structure and Contents.* New York: Oxford University Press.

Nucci, L., Turiel, E., & Encarnacion-Gawrych, G. (1983). Children's social interactions and social concepts in the Virgin Islands. *Journal of Cross-Cultural Psychology*, 14, 469-87.

Penn, D., & Povinelli, D. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind.' *Phil. Trans. R. Soc. B*, 362, 731-744.

Popper, K. (1945). *The Open Society and Its Enemies*. London: Routledge.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Brain and Behavioral Sciences*, 59, 33-53.

Prosser, W. (1941). *Casebook on Torts*. Minneapolis: University of Minnesota Press.

------. (1971). *Casebook on Torts* (4th ed.). Minneapolis: University of Minnesota Press.

Rachels, J. (1975). Active and passive euthanasia. *New England Journal of Medicine*, __. In Fischer, J.M. & Ravizza, M., (Eds). *Ethics: Problems and Principles* (pp. 111-116).

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rizzolati, G., Fogassi, L. & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci*. 2, 661-70.

Salmond, J. (1924/1902). *Salmond on Jurisprudence*. London: Sweet & Maxwell.

Sidgwick, H. (1988/1907). *The Methods of Ethics*. (7th Ed.). Indianapolis: Hackett

Smart, R. N. (1958). Negative utilitarianism. *Mind*, 67, 542-543.

Smetana, J. (1983). Social cognitive development: Domain distinctions and coordinations. *Developmental Review*, 3**,** 131–147.

------. (1993). Understanding social rules. In M. Bennett (Ed.)., *The Development of Social Cognition: The Child as Psychologist*. New York: Guilford Press.

Smetana, J., Kelly, M., & Twentyman, C. (1984). Abused, neglected, and nonmaltreated children's conceptions of moral and social-conventional transgressions. *Child Development*, 55, 277-287.

Smetana, J., Toth, S., Cicchetti, D., Bruce, J., Kane, P., & Daddis, C. (1999). Maltreated and nonmaltreated preschoolers' conceptions of hypothetical and actual moral transgressions. *Developmental Psychology*, 35, 269-281.

Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, 10 (1), 89-96.

Terry, H. (1884). *Some Leading Principles of Anglo-American Law, Expounded with a View to Its Arrangement and Codification.* Philadelphia: T. & J. W. Johnson.

Thomson, J. (1976). Killing, letting die, and the trolley problem. *The Monist*, __. In Fischer, J.M. & Ravizza, M., (Eds)., *Ethics: Problems and Principles* (pp. 67-77).

------. (1985). The trolley problem. *Yale Law Journal*, __. In Fischer, J.M. & Ravizza, M., (Eds)., *Ethics: Problems and Principles* (pp. 279-292).

Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention.* Cambridge: Cambridge University Press.

Turiel, E., Killen, M., & Helwig, C. (1987). Morality: Its structure, function, and vagaries. In J. Kagan & S. Lamb (Eds.)., *The Emergence of Morality in Young Children*. Chicago: University of Chicago Press.

Unger, P. (1996). *Living High and Letting Die: Our Illusion of Innocence*. Oxford: Oxford University Press.

Valedesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17(6), 476-477.

Wood, J. N., Glynn, D. D., Philips, B., Hauser, M. D. (2007). The perception of rational, goal-directed action in non-human primates. Science. Vol. 317 (5843), pp. 1402-1405.

Yau, J. & Smetana, J. (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, 74, 1647-1658.

Young, L., Cushman, F.A., Hauser, M.D., & Saxe, R. (2007). Brain regions for belief attribution drive moral condemnation for crimes of attempt. *Proceedings of the National Academy of Science*, 104 (20), 8235-8240.

Cases:

Garratt v. Dailey, 279 P.2d 1091 (Wash. 1955)

O'Brien v. Cunard Steamship Co., 28 N.E. 266 (Mass. 1891)

Dilemma	**Standard Legal Analysis**
1. Transplant	Battery/Homicide
2. Footbridge	Battery/Homicide
3. Country Road	Negligence/Duty to Rescue
4. Architect	Battery/Homicide
5. Lifeboat	Battery/Homicide
6. Hard Times	Battery and Rape/Sexual Assault
7. Smother for dollars	Battery/Homicide
8. Safari	Battery/Homicide
9. Crying Baby	Battery/Homicide
10. Plane Crash	Battery/Homicide
11. Hired Rapist	Battery and Rape/Sexual Assault
12. Grandson	Battery/Homicide
13. Infanticide	Battery/Homicide
14. Preventing the Spread	Battery/Homicide
15. Modified Lifeboat	Battery/Homicide
16. Modified Preventing the Spread	Homicide
17. Modified Safari	Battery/Homicide
18. Modified Bomb	Battery/Torture
19. Submarine	Battery/Homicide
20. Lawrence of Arabia	Battery/Homicide
21. Sophie's Choice	Battery/Homicide
22. Sacrifice	Homicide
23. Vitamins	Battery
24. Vaccine Test	Battery/Homicide
25. Euthanasia	Battery/Homicide

Fig. 1: Standard Legal Analysis of Greene et al.'s (2001) "Personal" Dilemmas

Author	Case	Purposeful Harmful Battery?	
		Yes	No
Foot (1967)	Cave explorer	√	
	Capital punishment	√	
	Transplant	√	
	Torture	√	
	Trolley		√
	Scarce Drug		√
Rachels (1975)	Bathtub #1 (Smith)	√	
	Bathtub #2 (Jones)		√
Nagel (1986)	Accident on a lonely road	√	
Boorse & Sorenson (1988)	Shield #1 (ravenous bear)	√	
	Duck #1 (ravenous bear)		√
	Shield #2 (gunman)	√	
	Duck #2 (gunman)		√
Cushman et al. (2006)	Aquarium #1 (Connor)	√	
	Aquarium #2 (Mike)		√
Greene et al. (2009)	Speedboat #1 (hands)	√	
	Speedboat #2 (oar)	√	
	Speedboat #3 (accelerate)		√
Turiel et al. (1983)	Moral violations	√	
	Conventional violations		√
Hamlin et al. (2007)	Hinderer	√	
	Helper		√

Fig. 2: Harmful Battery Cases in Ethics and Cognitive Science Literatures

[1] These remarks are potentially misleading in several respects. For example, the Restatement (First) also discusses consent and its contours in Sections 49-62 and the privilege of using force in self-defense or defense of others in Sections 63-76. Moreover, the Restatement (Second) affirmatively resists equating "subject to liability" with *prima facie* liability, both in Section 5 and elsewhere (see, e.g., Section 13, comment d). For present purposes, I ignore these complications.

[2] Admittedly, the fat man stuck in the cave is a more complicated case than the others, and Foot herself does not affirm whether blasting him out and thus saving the others would be permissible in either of the two conditions she describes. Still, a fully elaborated concept of harmful battery, including possible defenses, can help to clarify why this case is more complex than the other cases Foot describes. The prohibitory norm against purposeful harmful contact is clearly implicated here; what remains open is whether blasting the man out of the cave could nonetheless be justified under these circumstances. The most plausible justification is distinct from possible defenses in Foot's other cases because it includes an element of self-preservation, which perhaps could be viewed as self-defense, along with an element of false imprisonment and the use of force in response thereto. (The "false imprisonment" at issue is unusual and may not even be tortious because it is unintentional, but otherwise the act seems to qualify.) Furthermore, there is a clear element of Pareto-optimality in the first of the two conditions Foot describes, in which the fat man's head is stuck inside the cave and thus he will drown with the rest of the group. At the end of the day, he would not be made worse off by being killed in order to save the others (cf. Huebner et al., forthcoming). For the same reason, one might plausibly infer his ex ante consent to be being blasted out of the cave in these circumstances. Finally, it seems possible to appeal to what Donagan (1977: 180) calls "the doctrine of the legitimacy of consent to possible sacrifice in a common [dangerous] enterprise" to explain the putative intuition that it would be permissible to use the dynamite in the second of Foot's two conditions, in which the stuck man would be rescued in due course. The key overarching point here is that Foot's case of the cave explorer implicates possible defenses to harmful battery that are not present in any of her other examples. The moral intuitions elicited by this case do not, therefore, conflict with the main thesis of this paper. Rather, they appear to lend further support to that thesis.

[3] Rachels (1976: 114) describes Jones as fully prepared to kill: "Like Smith, Jones sneaks in planning to drown the child in the bath. However, just as he enters the bathroom Jones sees the child slip and hit his head, and fall face down in the water. Jones is delighted; he stands by, ready to push the child's head back under if it is necessary, but it is not necessary. With only a little thrashing about, the child drowns all by himself, 'accidentally,' as Jones watches and does nothing." On these facts, it seems possible, although unlikely, that Jones' conduct would trigger criminal liability under the law of attempt. If so, then the contrary statement in the text would need to be qualified.

[4] Some of Boorse and Sorensen's (1988) cases do not involve battery; hence, the generalization offered in the text does not apply to these cases. My concern here is with the subset of their "duck/shield" examples that do appear to turn on the presence of a harmful battery.

[5] One likely result of these definitions is that some of the cases designed by Greene et al. (2009) appear to reflect the traditional distinction between *trespass* and *trespass on the case*, according to which consequences flowing directly and immediately from defendant's act (trespass) can be distinguished from those that are indirect or mediated (case) (see, e.g., Epstein 2008; Terry 1884). This distinction was abandoned long ago by modern legal analysis, and this development traces to the narrow definition of an *act* outlined in Part II (see, e.g., Holmes 1881).

[6] Critics of the moral-conventional distinction, such as Kelly and colleagues (2007), have recently sought to challenge Turiel's findings by adducing new experimental data in which the signature moral pattern of responses does not always correlate with harmful transgressions. These critics have done so, however, by using fact patterns that target the *justificatory* norms that apply to purposeful harmful battery rather than the *prohibitory* norm or prima facie case. For example, the authors' "whipping" case (Kelly et al. 2007: 123) involves the use of force to enforce valid military orders (cf. Restatement of Torts, section 146); their "spanking" case (Kelly et al. 2007: 124) involves the use of force to discipline children (cf. Restatement of Torts, Sections 147-55); and their "simulated interrogation" case (Kelly et al. 2007: 125) implicates the concept of consent (cf. Restatement of Torts, Sections 49-62). Their findings do not, therefore, appear to undermine a reformulated version of the moral-conventional distinction that is sensitive to the fundamental difference between these two parts of the battery prohibition.