

## “Any Animal Whatever”: Harmful Battery and its Elements as Building Blocks of Moral Cognition

John Mikhail<sup>1</sup>

### I.

In *The Descent of Man*, Darwin (1981/1871, 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” (id. at 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

---

<sup>1</sup> Part I of this paper draws from Mikhail 2011b. The remainder is new. I am grateful to Henry Richardson and the participants in a Berkeley GALA seminar for comments on a previous draft.

conditions to which they are exposed” (id. at 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” (id. at 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” (id. at 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” (id. at 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 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 judgments is “constituted” or “driven by” emotion (Greene 2008, p. 108; see also Greene 2004, 2009; Koenigs et al. 2007) or merely correlated with emotion while being generated by unconscious computations (e.g., Huebner et al. 2008). On 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. The claim I wish to defend 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 begin by looking closely at the 25 “personal” dilemmas devised by Greene and colleagues (2001) and later used by them and a number of other researchers (e.g., Greene et al. 2004; Koenigs et al. 2007; Moore et al. 2008) in their fMRI and behavioral studies of emotional engagement in moral judgment. Greene found that these vignettes elicited increased activity in the medial prefrontal cortex (MPFC), posterior cingulate 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).<sup>2</sup> 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.

With one exception, then, what Greene and his colleagues (2001) actually did in the “personal” condition of their fMRI 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 these cases 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”

---

<sup>2</sup> In this table, “Battery” is meant to refer to purposeful harmful battery, a concept I will explicate in Part III.

<b>Dilemma</b>	<b>Standard Legal Analysis</b>
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

Table 1: Standard Legal Analysis of Greene's (2001) "Personal" Dilemmas

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 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 (2001, 2004) original imaging experiments did not really test two patterns of moral judgment—one “deontological” and the other “utilitarian”—as 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; Kant 1788; Leibniz 1705). 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 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) have drawn specious conclusions and misconceived the nature of the problem.

Returning to Darwin, considerations like these suggest that the main questions for a naturalistic moral psychology moving forward should 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 1981/1871; Spencer 1978/1897). 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.

## II.

In *Elements of Moral Cognition*, I argued in a broadly Darwinian vein that the best explanation of the properties of moral judgment implies the existence of an innate moral faculty. In this paper, I adopt the same basic theoretical framework, but I expand the argument by taking a closer look at one prong of the first question outlined at the close of Part I: how the brain computes representations of harmful battery. The hypothesis I ultimately wish to consider is whether one component of this process, which corresponds to the *prima facie* case of harmful battery, might turn out to be shared with some nonhuman animals, while a second component, which corresponds to justifications or affirmative defenses to harmful battery, may turn out to be uniquely human. My basic assumption is that this hypothesis seems plausible, and that the rule against harmful battery might therefore serve as a useful analytical tool with which to investigate Darwin’s famous observation that “any animal whatever” endowed with the necessary social instincts would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well-developed or as nearly well-developed as in humans.

Except for a few passing remarks, however, I do not engage directly with the literature on animal cognition in this paper. Instead, my primary objective is to take a close look at how harmful battery representations appear to operate in human moral psychology. In addition, I wish to make the novel and potentially controversial claim that the basic norm against harmful battery, when properly clarified, can be used to explain a wide range of moral intuitions that have been extensively discussed in the ethics and cognitive science literatures, the unified analysis of which is currently lacking. When it comes to moral intuitions about bodily harm, I argue, human

psychology may be simpler and more unified than we think. By focusing on this topic, we can perhaps play the role of Locke's under-laborer, clearing away some of the misconceptions that stand in the way of future research.

The examples from ethics I discuss in this paper 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), Young & Saxe (2008), Turiel (1983), and Hamlin et al. (2007). Many other illustrations could be offered, but these eight are adequate for my purposes. The claim I wish to defend is that all eight illustrations rely on fact patterns containing acts of harmful battery—the same wrong at issue in many standard trolley problems (Mikhail 2011a) and, as we have just seen, in virtually all the “personal” moral dilemmas devised by Greene and colleagues (2001, 2004). The point of convergence in all of these research endeavors 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 these justifications must be of a special character; otherwise the norm is difficult to dislodge. In this light, one can better appreciate the complex 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 2002, 2007b, 2011a; Nichols 2005).

The hypothesis of an innate moral faculty clearly plays a significant role in the traditional debates between natural law and legal positivism (compare, e.g., Grotius 1625, Hume 1751, Kant 1788, and Reid 1788, with Hobbes 1651, Bentham 1789, Austin 1832, and Kelsen 1925; see also

<b>Author</b>	<b>Definition of Natural Law</b>
1. Cicero	“something which is implanted in us, not by opinion, but by an innate instinct”
2. Ulpian	“what nature has taught all animals”
3. Isadore	“what is common to all nations and is set up by a natural instinct and not by any positive institution”
4. Gratian	“[what] is common to all nations by reason of its universal origin in a natural instinct and not in any (positive) constitution”
5. Aquinas	“a natural disposition of the human mind [that is] concerned the basic principles of behavior”
6. Suarez	“that form of law which dwells within the human mind, in order that the righteous may be distinguished from the evil”
7. Grotius	“the Dictate of Right Reason, indicating that any act, from its agreement or disagreement with the rational and social nature of man, has in it a moral turpitude or a moral necessity”

Table 2: Seven Classical Definitions of Natural Law

Table 2).<sup>3</sup> Although I have a significant interest in this subject (see, e.g., Mikhail 2007a, 2008), these debates are not my primary topic here. Instead, my main objective is to take a sustained and careful look at one type of intuitive moral perception and to consider what it might tell us about the rich internal structure of human—and perhaps also nonhuman—moral psychology.

The remainder of this paper proceeds in four parts. Part III outlines the main elements of the *prima facie* case of harmful battery, including voluntary act, harmful or offensive intent, and harmful contact, all of which are examined with an eye toward what these elements might teach us about properties of human moral psychology. Part IV 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 guiding idea of Parts III and IV 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

---

<sup>3</sup> For references to Table 2 as well as other classical formulations of natural law, see Mikhail 2007a: 754, n. 168.



may point us toward elements of moral cognition that are distinctively human. In Part V, I examine eight influential research endeavors in ethics and cognitive science and argue that all of them appear to rely on harmful battery scenarios without explicitly acknowledging this fact or considering what it might suggest for issues of experimental design, data interpretation, or theory construction. In Part VI, I briefly conclude by considering how the main results of the paper might bear on Darwin's (1981/1871: 71-72) conjecture that "any animal whatever" would acquire a moral sense or conscience, once its intellectual and social capacities became as well-developed as their human counterparts.

### III.

Our first task is to consider what the elements of battery might tell us about the properties of 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.” In my PhD dissertation and other early work (e.g., Mikhail 2000, 2002), I relied on Prosser's definition for the purpose of constructing a novel series of trolley problems designed to probe the intuitive operation of the battery prohibition. Since then, however, I have come to realize that Prosser's formula is incomplete and potentially misleading 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 objectives of this paper, which include the development of an empirical 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. From a formal perspective, which seeks to advance the goals of machine ethics (e.g., Marcus 2012) for instance, this definition can be summarized as follows:

Harmful Battery =<sub>df</sub> [ACT, INTENT<sub>H/O</sub>, CONTACT<sub>H</sub>, ~CONSENT, ~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 =<sub>df</sub> [ACT, INTENT<sub>H/O</sub>, CONTACT<sub>H</sub>]

In other words, harmful battery is comprised of three main building blocks: action, intention, and contact (cf. Cushman et al. 2006; Mikhail 2002). 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).<sup>4</sup>

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 the context of experimental moral psychology.

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

---

<sup>4</sup> These remarks are potentially misleading or inaccurate in several respects. For example, the Restatement (First) also discusses consent and its contours in Sections 49-62 and 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). I ignore these complications here.

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 simply to note 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 =<sub>df</sub> [MOTION<sub>V</sub>, INTENT<sub>H-O</sub>, CONTACT<sub>H</sub>]

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 human mind *represents* acts of harmful battery, not to make headway on complex philosophical problems of free will, determinism, compatibilism, and the like. For our purposes, then, we may assume that it is a property of the human mind or 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 tentatively infer 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 (Levine, Mikhail, and Leslie, forthcoming; Mikhail 2011a). Finally, these definitions of intent also 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, another counterintuitive implication that recent work in experimental philosophy has helped to illuminate (see, e.g., Knobe 2003, 2005, 2010).

To avoid these potential anomalies, it is necessary 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,

$$\begin{aligned} \text{Harmful Battery}_{[P-I]} &=_{df} [\text{ACT}, \text{PURPOSE}_{H/O}, \text{CONTACT}_H] \\ \text{Harmful Battery}_{[K-I]} &=_{df} [\text{ACT}, \text{KNOWLEDGE}_{H/O}, \text{CONTACT}_H] \end{aligned}$$

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 and computational research.

Harmful Battery <sub>[P-I]</sub> =<sub>df</sub> [MOTION<sub>V</sub>, GOAL<sub>H/O</sub>, CONTACT<sub>H</sub>]

Harmful Battery <sub>[K-I]</sub> =<sub>df</sub> [MOTION<sub>V</sub>, BELIEF<sub>H/O</sub>, CONTACT<sub>H</sub>]

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, the primary distinction these formulas help us to 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. Perhaps because both representations are encompassed by the legal definition of harmful battery, they often are conflated in the philosophical, psychological, and legal literatures, and thus can be a source of needless controversy. As we shall see in Part V, 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. As I have suggested, it also may ultimately prove to be a useful formula with which to investigate how certain aspects of moral cognition evolved in the species.

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

well-known 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 (Mikhail 2011a).

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; Mikhail 2011a; 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 we shall discuss turn on how the concept of "offensive" is analyzed. Philosophers such as Joel 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.

#### IV.

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 it is justified. Before turning to the widespread use of cases implicating 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 aspects of moral cognition that seem likely to be shared with nonhuman animals, these defenses may help us to comprehend which features of our moral psychology are distinctively human (cf. Darwin 1981/1871).

One familiar and important 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 a 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 famous 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 is not battered by the emergency room attendants who operate on her is that her hypothetical consent to being touched and physically 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?”) (Mikhail 2011a). 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 barely 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).

Of course, codifications like these are merely a starting point. A great deal of insightful work has been done to refine, clarify, and challenge their core principles, concepts, and doctrines (see, e.g., Epstein 2008; Fletcher 1998; Gardner 2008; Robinson 1982). From our perspective, the principal issue not to lose sight of here is that, with only a few significant exceptions, all of these criminal law and tort doctrines appear to track common moral intuitions, at least to a good first approximation. In this connection, 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 predict and explain the intuitive behavioral data across a wide and diverse range of cases. The precise character and extent of this knowledge, of course,

are potentially difficult issues that warrant careful investigation. Nonetheless, the general thesis seems sound and should continue to be a reference point for future research in this area.

## V.

Let us now take up the question I raised earlier, which is whether the factor of purposeful battery can be used to explain a class of moral intuitions that has been widely discussed in the ethics and cognitive science literatures. I am inclined to believe that the answer is yes. For the past several decades, philosophers and psychologists alike have been relying on cases of purposeful battery and studying their effects without clearly acknowledging this fact or explicitly considering its theoretical and empirical implications. In what follows, I first elaborate on these claims by examining some famous thought experiments in moral philosophy that appear to rest on purposeful battery intuitions. I then discuss some influential experiments in cognitive science to which the same generalization applies.

To the best of my knowledge, the cases I discuss in this section have not been analyzed together previously under the heading of this or any other legal rule. Part of my aim in grouping them together here is to demonstrate that these cases can, in fact, be classified as members of a unified set, which can be explained by means of a single causal factor. If this conclusion is correct, then it suggests that broad generalizations are being neglected in favor of more partial explanations. The missing common denominator is goal-directed 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 Doctrine of Double Effect (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 (i) the fat man stuck in the cave, who can be dislodged only by exploding a stick of dynamite; (ii) the judge who frames and executes an innocent man in order to prevent a bloody riot; (iii) the five patients who can be saved only by transplanting organs from one healthy individual; and (iv) 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, all of which Foot assumes are impermissible, is that each of them encompasses a purposeful harmful battery. That is, each case involves an actor who generates a purposeful harmful contact, inflicted without consent or other justification. By contrast, the common element (or at least one common element) of the contrasting series of permissible actions 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 of all these cases could thus appeal to a rule against purposeful harmful battery, together with the DDE, to explain these common reactions.<sup>5</sup>

---

<sup>5</sup> Admittedly, Foot’s example of the fat man stuck in the cave is a more complicated case, and Foot herself does not clearly assert whether blasting him out and thus saving the others would be permissible in either of the two conditions she describes. Still, it seems plausible to assume that 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 also 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 nonstandard and may not be tortious insofar as 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 this condition, the fat man’s head is stuck inside the cave and

## 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, on reflection 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, under most circumstances only Smith's conduct would trigger criminal or civil liability. That this last statement seems shocking may derive from the fact that Jones' failure to rescue constitutes a type of *purposeful* homicidal omission, whereas most failures to rescue do not.<sup>6</sup> In

---

thus he will drown with the rest of the group. At the end of the day, therefore, 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 that *ex ante* the man would give his 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 Foot's second condition, in which the stuck man would be rescued in due course. In any event, the key point is that Foot's case of the fact man stuck in the cave 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, directly conflict with the main thesis of this paper, but if anything appear to lend indirect support to that thesis.

<sup>6</sup> 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

addition, it does not follow from the distinction drawn here that the two actors are not equally reprehensible or that active and passive euthanasia should not be treated equivalently in morality or law. My aim here is not to intervene in these debates, but merely to note an apparent fallacy in Rachels' argument, as it relates to our main topic. Despite his confident assertion to the contrary, Rachels' bathtub cases can, in fact, be distinguished on grounds other than the distinction between killing and letting die. The outcomes of these cases are the same, but the means are different, only one set of which violates an independent norm against harmful battery.

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

Nagel (1986: 171) invites us to consider 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 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 could 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. Should he do this?

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 also suggests that the moral intuitions of common sense would agree with him. Whether or not this assumption is correct, the important point for our purposes is that the action Nagel relies on

---

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.

in his thought experiment is, once again, a deliberate, harmful battery. The question he puts to his reader can be framed 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 mental representations as Thomson's footbridge problem, Foot's transplant problem, Greene's personal moral dilemmas, and a large number of other harmful battery cases discussed in this paper and throughout the philosophical literature. At first glance, Nagel's accident case may seem different, but on further reflection it appears to be yet another illustration of the broad reach of this single, powerful norm.

#### 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 category, an agent moves his own body (but not 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 category, 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.<sup>7</sup>

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).

---

<sup>7</sup> To be clear, several of Boorse and Sorenson's (1988) cases do not involve battery; hence, the generalization offered in the text does not apply to those 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.

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

Condition 1

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.

Condition 2

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 & 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." These are striking findings, to be sure, but they do not need to be explained by novel concepts such as "ducking" or "shielding." A more parsimonious explanation can be given by invoking a rule against purposeful battery. Boorse & Sorensen's findings provide further support for the proposition that common moral cognition recognizes an intuitive difference between committing a deliberate battery as a means to an end and achieving the same end without committing a purposeful battery. In short, their findings can be explained by the same simple rule that explains the other cases discussed in this paper. There is no need to reinvent the wheel.

## B. Cognitive Science

Turning to cognitive science, here, too, one encounters a widespread reliance on harmful battery intuitions, which has gone almost entirely unremarked in the literature. Cognitive scientists frequently use cases of battery, necessity, and double effect in their experiments without clearly acknowledging this fact or explicitly considering what it might imply for issues of experimental design, data interpretation or theory construction. In what follows, I illustrate

these claims with examples from cognitive psychology, neuroscience, developmental psychology, and infant cognition. Again, the common factor unifies the illustrations appears to be purposeful battery, inflicted without consent or other justification.

#### 5. *Cushman's three principles of harm*

Building on some of my own previous research on the trolley problems (Mikhail 2002), Fiery Cushman and his colleagues (2006: 1092) investigated three “principles of harm,” which they held to be operative in moral judgment:

- *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, Cushman et al. found that subjects make moral judgments in conformity with these principles across a large database of superficially distinct fact patterns. For example, in the following pair of cases, subjects judge Conner's action to be morally worse than Mike's action:

##### 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; emphasis added].

##### 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; emphasis added].

In these cases, Connor commits an act that satisfies the *prima facie* case for purposeful battery as a means to achieve his good end, whereas Mike does not. The same critical distinction runs through many similar pairs devised by Cushman and his colleagues. By constructing clever examples like these, they discovered that ordinary adults appear to know the basic elements of battery intuitively (albeit somewhat inarticulately) and that they shape their moral intuitions accordingly.<sup>8</sup> The concept of “battery” is never clearly referenced in the study by Cushman and his colleagues, however, nor do the authors acknowledge that the three principles on which they rely combine to form a basic legal rule. As a result, one of the most robust findings of this important study remains needlessly obscure.

#### 6. *Young and Saxe’s poisoning cases*

Liane Young & Rebecca Saxe (2008) used fMRI techniques to identify a number of brain regions that appear to be selectively recruited for moral judgment tasks that require perception of an agent’s intentions. Their basic experimental design was to measure neural activity in cases of intended harm, attempted harm, accidental harm, and ordinary harmless conduct in a 2 x 2 design that crossed mental state information (the agent did/did not intend harm) and outcome information (the harm did/did not result). Young & Saxe discovered that the ventromedial prefrontal cortex (VMPFC), posterior cingulate cortex (PCC), and especially the right temporal parietal junction (RTPJ) were selectively more active in cases involving intended and attempted

---

<sup>8</sup> Cushman et al. found that subjects could generally explain their moral judgments in terms of the action and contact principles, but had more difficulty articulating the intention principle. This result tends to confirm my own early finding (Mikhail 2000, 2002) that subjects are generally incapable of articulating the principle of double effect, even though they appear to rely on that principle when generating a certain class of moral judgments.

harms than in the other two conditions, indicating that the task of ascribing intent to harm may be localized in these regions. Here are four of the cases Young & Saxe used in their research:

Case 1: Negative Belief / Negative Outcome

Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pour some coffee, Grace's friend asks for some sugar in hers. The white powder by the coffee is *not sugar but a toxic substance left behind by a scientist*. Because the substance is in a container marked "*toxic*", Grace thinks that it is *toxic*. Grace puts the substance in her friend's coffee. Her friend drinks the coffee and *dies*. Putting the substance in the coffee was [permissible or forbidden? Judged on a 1-3 scale; emphasis added].

Case 2: Negative Belief / Neutral Outcome

Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pour some coffee, Grace's friend asks for some sugar in hers. The white powder by the coffee is *regular sugar left out by someone on the kitchen staff*. Because the substance is in a container marked "*toxic*", Grace thinks that it is *toxic*. Grace puts the substance in her friend's coffee. Her friend drinks the coffee and *is fine*. Putting the substance in the coffee was [permissible or forbidden? 1-3 scale; emphasis added].

Case 3: Neutral Belief / Negative Outcome

Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pour some coffee, Grace's friend asks for some sugar in hers. The white powder by the coffee is *not sugar but a toxic substance left behind by a scientist*. Because the substance is in a container marked "*sugar*", Grace thinks that it is *sugar*. Grace puts the substance in her friend's coffee. Her friend drinks the coffee and *dies*. Putting the substance in the coffee was [permissible or forbidden? 1-3 scale; emphasis added].

Case 4: Neutral Belief / Neutral Outcome

Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pour some coffee, Grace's friend asks for some sugar in hers. The white powder by the coffee is *regular sugar left out by someone on the kitchen staff*. Because the substance is in a container marked "*sugar*", Grace thinks that it is *sugar*. Grace puts the substance in her friend's coffee. Her friend drinks the coffee and *is fine*. Putting the substance in the coffee was [permissible or forbidden? 1-3 scale; emphasis added].

In the first case, Grace knows the powder is toxic. She deliberately kills her friend, thereby committing both murder and (as an underlying offense) purposeful battery. In the second case, Grace mistakenly believes the powder is toxic, but it actually is sugar. Grace is

guilty of attempted murder, as well as the lesser-included criminal offense of attempted assault,<sup>9</sup> both of which are also impossible attempts in this case. In the third case, Grace believes the powder is sugar, but it actually is toxic; thus, her friend dies after drinking the coffee.

Depending on how reasonably Grace acted under the circumstances, this could be an accident for which she is blameless (reasonable mistake of fact), or a case of negligence. (Whoever had the bright idea to place a toxic substance in a container marked “sugar,” however, is clearly in trouble.) In the fourth case, Grace correctly believes the powder is sugar; thus, when her friend drinks the coffee, she is fine. This is merely ordinary harmless conduct, for which no particular moral judgment is required.

Although their main objective was to locate regions of interest (ROI) in the brain, Young & Saxe (2008) also collected behavioral data on these cases. As one might expect, the murder and attempted murder cases were judged more harshly than the negligent homicide and ordinary harmless conduct cases. More surprisingly, and more significantly for our purposes, the variance between the attempted murder and negligent homicide cases was practically as large as that between the murder and harmless conduct cases. This finding corroborates and reinforces the conclusion that one element of moral cognition is an acute sensitivity to goal-directed harm. Not only is this principle reflected in the law of intentional battery, but it also manifests itself in multiple and diverse ways in the intuitive verdicts of common moral cognition.

#### *7. Turiel’s moral-conventional distinction*

Consider next the large body of research on the moral-conventional distinction. Eliot Turiel and his colleagues (see, e.g., Nucci & Turiel 1978; Smetana 1981, 1984; Turiel 1983)

---

<sup>9</sup> Because of its focus on compensable harms, the law of torts generally does not recognize a tort of “attempted battery” that is distinct from a tortious assault. In these cases, there is no indication that Grace’s friend possessed the required conscious awareness to turn Grace’s action into a tortious assault. Nonetheless, Grace’s conduct in the second case would presumably qualify as an attempted assault in the criminal law, along with an attempted murder.

proposed this distinction to mark the difference between two normative domains, which they claimed children reliably differentiate. They found that moral transgressions (e.g., one child hitting another) are widely held by children to be (i) more serious, (ii) more deserving of punishment, (iii) authority-independent and (iv) universally applicable. By contrast, they found that conventional transgressions (e.g., wearing pajamas to school) do not possess these properties. Subsequently, a large literature has found that children as young as 3-4 years of age reliably draw this distinction in culturally diverse settings (see, e.g., Blair 1995; Hollos et al. 1986; Nucci et al. 1983; Smetana et al. 1999; Turiel et al. 1987; Yao & Smetana 2003). As a result, the moral-conventional distinction has become an important touchstone for philosophically sophisticated efforts to theorize about moral development (e.g., Dwyer 1999; Nichols 2004).<sup>10</sup>

For our purposes, what is important to note in connection with the moral-conventional distinction is that the vast majority of moral violations in this research program are acts of “hitting, kicking, biting, punching, pulling hair, or otherwise intentionally physically harming another child” (Smetana 1984: 1769)—in short, harmful batteries. Accordingly, the reactions of children in these studies can be accounted for by a rule against harmful battery and on this basis alone. By contrast, *none* of the cases that have been classified as conventional violations involve harmful batteries, at least as far as I have been able to discover (see, e.g., Nucci & Turiel 1978; Smetana 1981, 1984; Turiel 1983). A central finding of this influential body of research, therefore, can be explained once again by postulating an acute sensitivity to *battery* as a property

---

<sup>10</sup> As I note elsewhere (e.g., Mikhail 2007a, 2009b), the moral-conventional bears a strong resemblance to the traditional distinction between *mala in se* and *mala prohibita*. For some critical discussion of this topic in the context of empirical moral psychology, see generally Braman et al. 2010a, 2010b; Robinson et al. 2007, 2010.

of the human mind, which emerges early and reliably in child development.<sup>11</sup> One does not need to appeal to the more complex—and more controversial—moral-conventional distinction.<sup>12</sup>

#### 8. Hamlin's helper-hinderer experiments

Consider finally the celebrated research by Kiley Hamlin and her colleagues (2007) on moral cognition in infants, according to which 6- and 10-month-old infants appear to evaluate the conduct of other agents based on the social or moral quality of their acts. The basic design of these studies involved a red wooden circle with large eyes glued onto it (the 'climber'), who repeatedly attempts to climb a hill and either receives from a yellow circle (the 'helper'), who pushes the climber up the hill from below, or interference from a blue square (the 'hinderer'), who pushes the climber down the hill from above. A third condition involved a neutral character who traces a path identical to that of the helper or hinderer, but who does not interact with the climber. Given an opportunity to choose which object to reach out and grab, the infants who encountered these displays consistently preferred helpers over hinderers; indeed, they did so to a remarkable extent, virtually always preferring the former to the latter. These choices were not based on color, shape, or other perceptual factors, which were carefully controlled.

Hamlin et al. suggest that the infants' choices "are best explained as specifically social evaluations: a liking for those who act cooperatively to facilitate the goals of others, and a dislike

---

<sup>11</sup> To clarify, not all of the moral violations in the moral-conventional paradigm are batteries; for example, some are thefts or other property-based violations (e.g., stealing another child's pencil). My concern here is only with those cases that constitute battery. The general remarks made in the text should be understood in this light.

<sup>12</sup> An important paper by Kelly et al. (2007) challenges Turiel's findings by presenting new evidence which suggests that the signature moral pattern of responses Turiel identifies does not always correlate with harmful transgressions. The scenarios it utilizes, however, generally target the *justificatory* norms that apply to harmful battery rather than elements of the *prohibitory* norm or *prima facie* case of harmful battery. For example, Kelley and colleagues rely on a "whipping" case (2007: 123) that involves the use of force to enforce valid military orders (cf. Restatement of Torts, section 146); a "spanking" case (2007: 124) that involves the use of force to discipline children (cf. Restatement of Torts, Sections 147-55); and a "simulated interrogation" case (2007: 125) that implicates the concept of consent (cf. Restatement of Torts, Sections 49-62). Although Kelley et al. cast doubt on some aspects of Turiel's research program, therefore, their findings do not appear to undermine a reformulated version of the moral-conventional distinction that is sensitive to the critical difference between these two parts of the battery prohibition.

of those who impede another's goals" (2007: 558). Although this interpretation seems like a plausible starting point, it also seems likely to need further refinement. For example, there are strong reasons for thinking that "helper" and "hinderer" will turn out to be both over-inclusive and under-inclusive in relation to these studies. There are many ways in which one can help or hinder someone, only some of which involve the type of conduct that is germane here. Likewise, the fact that parents often obstruct the navigation of their children (often for reasons of safety) casts further doubt on whether these data are tethered to helping and hindering as such.

More broadly, the fact that the climber is bumped up or down the hill means that a bodily *contact* rests at the heart of these studies. What the infants encounter in these displays is not merely an agent who facilitates or impedes another's goals, therefore, but an agent who does so by means of a deliberate harmful or offensive contact, which a rule against purposeful battery would recognize and proscribe. This is how an informed adult would analyze this situation, so we should be open to considering whether the infants are effectively doing the same thing. If the subjects in these experiments can be assumed to know or possess this legal rule unconsciously, then one can predict and explain their behavior without recourse to notions such as "helper" and "hinderer." Of course, these arguments are highly speculative, and more research is needed to determine whether they are in fact plausible. Philosophically informed speculation, however, seems just what is required in cases like this (for a good discussion of this point, see Kar 2012).

## VI.

Let us take stock and consider what lessons can be drawn from these investigations. We have seen that virtually all of the moral judgments in the personal condition of Greene's (2001, 2004) original fMRI experiments involve acts of purposeful battery (Table 1). This is a significant conclusion, which also takes on added meaning because so many research programs

in neuroscience rely on the same or similar cases (e.g., Koenigs et al., 2007).<sup>13</sup> We have also seen that the philosophical thought experiments devised by Foot, Rachels, Nagel, and Boorse & Sorenson can likewise be usefully analyzed with reference to purposeful battery. Finally, we have discovered that the same act appears to be implicated by four prominent research programs in cognitive science: (i) Cushman’s work on adult moral judgments; (ii) Young & Saxe’s research on the RTPJ; (iii) Turiel’s moral-conventional distinction; and (iv) Hamlin’s studies of preverbal infants. Surprisingly, all eight of these research endeavors appear in one way or another to be investigating the *very same thing*: how the mind computes representations of battery (and other *prima facie* torts and crimes), and the precise justifications or excuses thereto recognized by common moral cognition (Table 3).<sup>14</sup> Despite this, none of these research programs relies on a technical legal vocabulary to describe the relevant mental operations, and to that extent their findings appear less focused and illuminating than they could be.

From a legal perspective, we know that battery is a core human wrong, which operates as a lesser-included offense of a wide range of more significant acts of crime, violence and aggression. Rape, murder, aggravated assault, domestic violence, sexual molestation, genital mutilation, water-boarding and other forms of torture, and virtually all other forms of unlawful violent conduct are, paradigmatically, also acts of battery. (Murder, torture, and other forms of inhumane treatment can be committed without an underlying battery, but in most cases they do include battery as a lesser-included offense.) By clarifying the key elements of battery, then, we

---

<sup>13</sup> For several reasons, this conclusion appears to pose significant challenges for some of Greene’s more extravagant normative and metaethical arguments in “The Secret Joke of Kant’s Soul” and other publications (see Greene 2005, 2008). Because this is a topic for another paper, however, I do not address it here.

<sup>14</sup> In this table, Case #2 in Young & Saxe (2008) is checked off as a purposeful battery. Strictly speaking, this is incorrect. Because no bodily contact actually occurs, Grace’s act does not constitute battery. Still, her action plan in this case does include the representation of a goal-directed, harmful bodily contact. For our purposes, the purposeful battery category should therefore be interpreted broadly to include attempted as well as completed acts.

Author	Case	Purposeful Battery?	
		Yes	No
1. Foot (1967)	Cave explorer	√	
	Capital punishment	√	
	Transplant	√	
	Torture	√	
	Trolley		√
	Scarce Drug		√
2. Rachels (1975)	Bathtub #1 (Smith)	√	
	Bathtub #2 (Jones)		√
3. Nagel (1986)	Accident on a lonely road	√	
4. Boorse & Sorenson (1988)	Shield #1 (ravenous bear)	√	
	Duck #1 (ravenous bear)		√
	Shield #2 (gunman)	√	
	Duck #2 (gunman)		√
5. Cushman et al. (2006)	Aquarium #1 (Connor)	√	
	Aquarium #2 (Mike)		√
6. Young & Saxe (200)	Case #1 (murder)	√	
	Case #2 (impossible attempt)	√	
	Case #3 (negligent homicide)		√
	Case #4 (harmless conduct)		√
7. Turiel et al. (1983)	Moral violations	√	
	Conventional violations		√
8. Hamlin et al. (2007)	Hinderer	√	
	Helper		√

Table 3: Purposeful Battery as an Operative Factor in Eight Case Studies

not only learn how a single, bedrock legal norm operates, but we also may be able to gain a better understanding of how more complex offenses are mentally represented and evaluated.

The basic psychological elements of the *prima facie* case of purposeful battery are themselves fairly simple. As we have seen, they can be decomposed into their basic cognitive components without too much difficulty. It is precisely the relative ease with which this can be done, in fact, that suggests that this norm might serve as a useful analytical tool with which to investigate Darwin's famous conjecture 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.”

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). From a naturalistic perspective, the first element in its simplest form may reduce to the concept of a voluntary motion detector that applies to the bodily movements of conspecifics (or other agents). The second element may reduce 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 element may reduce 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 readily see how “any animal whatever” having the necessary intellectual and social capacities might be capable of performing these relatively simple tasks.

The picture seems quite 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.” As far as one can tell, only humans appear to engage in these more complex acts of moral reasoning.

Of course, it is important to recognize that 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 we have devoted attention here. 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 have any researchers inquired whether chimpanzees draw inferences about unintentional harmful contacts on the basis of a “reasonable chimpanzee” standard (compare Premack & Woodruff 1978 and the vast literature it has spawned). It seems conceivable that if scientists were to focus their attention on topics like these, they might make surprising discoveries that would force us to revise our understanding of our own moral psychology and that of nonhuman animals. Perhaps what is required here is a leap of scientific imagination akin to what Newton managed when he came up with the crazy idea that apples fall to the earth *for the same reason* that heavenly objects move in elliptical orbits. By parallel reasoning, perhaps ordinary human moral cognition is not as special as we’d like to think. In any event, it seems plausible to assume on the basis of what we have uncovered here that at least some elements of the legal rules pertaining to harmful battery form essential building blocks of human moral cognition, at least some of which may in turn be operative in other animals as well. If one takes seriously Darwin’s insights on the natural history of moral cognition and seeks to build upon them, this seems like good place to start.

## References:

- Anscombe, G. E. M. (1957). *Intention*. Oxford: Basil Blackwell.
- Austin, J. (1995/1832). *The Province of Jurisprudence Determined*. Cambridge: Cambridge University Press.
- . (1879/1861). *Lectures on Jurisprudence*. London.
- Baird, J. (2001). Motivations and morality: Do children use mental state information to evaluate identical actions differently? Paper presented to Biennial Meeting, Society for Research in Child Development, Minneapolis, MN.
- Bennett, J. (1995). *The Act Itself*. Oxford: Clarendon Press.
- Bentham, J. (1948/1789). *An Introduction to the Principles of Morals and Legislation*. New York: Halfner Press.
- Berndt, T.J., and Berndt, E.G. (1975). Children's use of motives and intentionality in person perception and moral judgment. *Child Development*, 46, 904-912.
- Blair, J. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57, 1–29.
- . (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends Cog. Sci.*, 11, 387-392.
- 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).
- Braman, D., Kahan, D., & Hoffman, D. (2010a). Some realism about punishment naturalism. *University of Chicago Law Review* 77, 1531-1609.
- . (2010b). A core of agreement. *University of Chicago Law Review* 77, 1655-1660.
- 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.
- Costanza, P.R., Coie, J. D., Grumet, J. F., and Farnhill, D. (1973). A Reexamination of the effects of intent and consequence on children's moral judgments. *Child Development*, 44, 154-61.
- 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.
- Dancy, J. (1983). Ethical particularism and morally relevant properties. *Mind*, 92, 530-547.
- 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.
- Dwyer, S. (1999). Moral competence. In K. Murasugi & R. Stainton (Eds.), *Philosophy and Linguistics* (pp. 169-190). Boulder, CO: Westview Press.
- Epstein, R. (2008). *Cases and Materials on Torts* (9<sup>th</sup> 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.
- Fletcher, G. (1998). *Basic Concepts of Criminal Law*. Oxford: Oxford University 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.
- Gardner, J. (2008). *Offences and Defences: Selected Essays in the Philosophy of Criminal Law*. Oxford: Oxford University Press.
- Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165-93.
- Goldman, A. (1970a). *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.
- . (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*. Cambridge, MA: MIT Press, pp. 35-79, 105-117.
- 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.
- Grotius, H. (1925/1625). *On the Law of War and Peace*. (F. W. Kelsey, Trans.). Oxford: Clarendon Press.
- 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.
- Hauser, M. (2006). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: Harper Collins.
- Hauser, M., Cushman, F., Young, L., Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language*, 22, 1-22.
- Hilliard, F. (1859). *The Law of Torts*. 2 volumes. Boston: Little, Brown and Company.
- Hobbes, T. (1968/1651). *Leviathan*. (C.B. Macpherson, Ed.). New York: Penguin.
- 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., Dwyer, S. & Hauser M. (2008). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13, 1-6.
- Hume, D. (1983/1751). *An Enquiry Concerning the Principles of Morals*. (J.B. Schneewind, Ed.). Indianapolis: Hackett.
- 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.
- Kant, I. (1993/1788). *Critique of Practical Reason*. (L.W. Beck, Trans.). MacMillan: New York.

- Kar, R. (2012). The two faces of morality: How evolutionary theory can both vindicate and debunk morality (with a special nod to the growing importance of law). In James E. Fleming and Sanford Levinson, (Eds.), *Evolution and Morality* (pp. 31-99).
- Kelly, D., Stich, S. Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect, and the moral-conventional distinction. *Mind & Language*, 22, 117-131.
- Kelsen, H. (1945/1925). *General Theory of Law and State*. (Anders Weberg trans.). Cambridge: Harvard University Press.
- 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.
- . (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*. 33: 315-329.
- 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.
- Leibniz, G. (1981/1704). *New Essays on Human Understanding*. (P. Remnant & J. Bennett, Eds.). Cambridge: Cambridge University Press.
- Lilliard, A. S. and Flavell, J. H. 1990. Young children's preference for mental state versus behavioral descriptions of human action. *Child Development* 61, 731-xxx
- Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.
- Marcus, G. (2012). Moral machines. *The New Yorker*, November 12, 2012.
- Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838-50.
- Mendez, M., Anderson, E., & Shapira, J., (2005). An investigation of moral judgment in frontotemporal dementia. *Cogn. Behav. Neurol.* 18, 193-97.
- 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.
- (2002). Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery, the rescue principle, the first principle of practical reason, and the principle of double effect. Stanford Law School JD Thesis, available at <http://www.law.georgetown.edu/faculty/faculty-webpages/john-mikhail/index.cfm>

- . (2007a). “Plucking the mask of mystery from its face”: Jurisprudence and H.L.A. Hart, *Georgetown Law Journal*, 95, 733-779 (2007).
- . (2007b). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences*, 11, 143-152 (2007).
- . (2008). Scottish common sense and nineteenth-century American law: A critical appraisal. *Law and History Review*, 26, 167-175.
- . (2009a). 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.
- . (2009b). Is the prohibition of homicide universal? Evidence from comparative criminal law. *Brooklyn Law Review*, 75, 497-515.
- . (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.
- Mikhail, J., Sorrentino, C., & Spelke, E. (1998). Toward a universal moral grammar. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings, Twentieth Annual Conference of the Cognitive Science Society* (p. 1250). Mahwah, New Jersey: Lawrence Erlbaum Associates
- 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).
- Nelson, S. (1980). Factors influencing young children’s use of motives and outcomes as moral criteria. *Child Development*, 51, 823-829.
- 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. (2001). *Education in the Moral Domain*. Cambridge: Cambridge 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.
- Piaget, J. (1965/1932). *The Moral Judgment of the Child*. New York: Free Press.
- 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.
- Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Prosser, W. (1941). *Casebook on Torts*. Minneapolis: University of Minnesota Press.
- (1971). *Casebook on Torts* (4th ed.). Minneapolis: University of Minnesota Press.
- Quinn, W. (1989). Actions, intentions, and consequences: The doctrine of doing and allowing. *Philosophical Review*, \_\_. In Fischer, J.M. & Ravizza, M., (Eds.). *Ethics: Problems and Principles* (pp. 145-161).
- 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.
- Reid, T. (1969/1788). *Essays on the Active Powers of the Human Mind*. Cambridge: MIT Press.
- Rizzolatti, G., Fogassi, L. & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661-70.
- Robinson, P. (1982). Criminal law defenses: A systematic analysis. *Columbia Law Review*, 82, 199-291.
- Robinson, P., Kurzban, R., & Jones, O. (2007). The origins of shared intuitions of justice. *Vanderbilt Law Review*, 60, 1633-1688.
- (2010). Realism, punishment & reform. *University of Chicago Law Review*, 77, 1611-1632.
- Salmond, J. (1924/1902). *Salmond on Jurisprudence*. London: Sweet & Maxwell.
- Sidgwick, H. (1988/1907). *The Methods of Ethics*. (7<sup>th</sup> 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.
- Spencer, H. (1978/1897). *The Principles of Ethics*. (2 Vols). Indianapolis: Liberty Classics.
- Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, 10 (1), 89-96.
- 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.
- Valesdolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17(6), 476-477.
- 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. & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage*, 40, 1912-1920.

#### Cases:

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

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