Empirical Work in Moral Psychology

Joshua May

[Penultimate draft; final version in the *Routledge Encyclopedia of Philosophy*.]

How do we form our moral judgments, and how do they influence behavior? What ultimately motivates kind versus malicious action? Moral psychology is the interdisciplinary study of such questions about the mental lives of moral agents, including moral thought, feeling, reasoning, and motivation. While these questions can be studied solely from the armchair or using only empirical tools, researchers in various disciplines, from biology to neuroscience to philosophy, can address them in tandem. Some key topics in this respect revolve around moral cognition and motivation, such as moral responsibility, altruism, the structure of moral motivation, weakness of will, and moral intuitions. Of course there are other important topics as well, including emotions, character, moral development, self-deception, addiction, well-being, and the evolution of moral capacities.

Some of the primary objects of study in moral psychology are the processes driving moral action. For example, we think of ourselves as possessing free will; as being responsible for what we do; as capable of self-control; and as capable of genuine concern for the welfare of others. Such claims can be tested by empirical methods to some extent in at least two ways. First, we can determine what in fact our ordinary thinking is. While many philosophers investigate this through rigorous reflection on concepts, we can also use the empirical methods of the social sciences. Second, we can investigate empirically whether our ordinary thinking is correct or illusory. For example, we can check the empirical adequacy of philosophical theories, assessing directly any claims made about how we think, feel, and behave.

Understanding the psychology of moral individuals is certainly interesting in its own right, but it also often has direct implications for other areas of ethics, such as metaethics and normative ethics. For instance, determining the role of reason versus sentiment in moral judgment and motivation can shed light on whether moral judgments are cognitive, and perhaps whether morality itself is in some sense objective. Similarly, evaluating moral theories, such as deontology and utilitarianism, often relies on intuitive judgments about what one ought to do in various hypothetical cases. Empirical research can again serve as a tool to determine what exactly our intuitions are and which psychological processes generate them, contributing to a rigorous evaluation of the warrant of moral intuitions.

Table of Contents

Updated and expanded in 2016 (originally published in 2012). Many thanks to Tim Crane, Aileen Harvey, Eddy Nahmias, Walter Sinnott-Armstrong, and Chandra Sripada for comments on drafts of this entry, and to Mark Schroeder for inviting me to write it. Unfortunately, I cannot acknowledge them in the published version, so I will here.

1. Moral Responsibility and Free Will

A famous challenge to our having free will and being morally responsible for what we do is (causal) determinism. If *determinism* is true, then the current state of the universe and the past together causally necessitate a unique future state. While *compatibilists* maintain that the truth of determinism does not preclude moral responsibility, *incompatibilists* insist that it does (see Free Will; Responsibility). One popular strategy among incompatibilists is to claim they have the intuitive, common sense, or default position (e.g. Kane 1999). This can then motivate incompatibilism, shift the burden of proof onto compatibilists, and so on.

The claim that one theory is a piece of common sense is subject to empirical investigation (see Experimental Philosophy). As such, some philosophers have presented ordinary people with hypothetical cases in order to see whether their natural inclination is toward incompatibilism. Some early studies, done primarily with undergraduate students in the U.S., have indicated that incompatibilism isn't more intuitive, since most participants count someone as morally responsible for a wrongdoing, such as stealing, in a deterministic universe (e.g. Nahmias, Morris, Nadelhoffer, and Turner 2006). However, subsequent studies suggest a more complicated picture. When presented with the abstract question of whether someone can be responsible in a world that operates as determinists maintain, the vast majority of people think *not* (about 86%). Yet, in line with previous results, most people will say the protagonist *is* morally responsible in a hypothetical case (about 72%), provided it is described in a concrete way that elicits emotional responses (Nichols & Knobe 2007).

Extant evidence suggests that both compatibilist and incompatibilist intuitions can be found in ordinary thinking (e.g. May 2014b; Nichols 2015). But that does not settle whether either is warranted. Some argue that allegedly incompatibilist intuitions result from a misunderstanding of descriptions of determinism. Ordinary people may well be comfortable with compatibilism upon understanding that deterministic processes needn't bypass the motivations, reasons, and decisions behind one's actions (Murray & Nahmias 2014). Others argue, however, that we do have genuinely incompatibilist intuitions, and they result from developing the concept of choice in what we perceive to be an indeterministic world. We just tacitly assume that choices cannot exist in a deterministic universe (Nichols 2015). Such debates reveal the complex interplay between normative questions about the status of certain intuitions and descriptive questions about the nature and source of such intuitions.

Determinism is the classical threat to free will and moral responsibility, but some have argued that empirical studies are likewise threatening. Social psychologists, for example, have demonstrated that arbitrary situational factors can affect what we do. In particular, helping behavior can change dramatically by slightly altering environmental factors, such as ambient fragrance, temperature, weather, noise levels, and lighting quality (see Miller 2009, §2). Presumably the phenomenon here is the familiar one of being less willing to help when in a bad mood. Still, the differences these factors can make are disconcerting. For example, in one provocative study, people in an area of a mall with pleasant smells, such as fresh baked cookies, helped more than twice as often as those near more neutral fragrances (Baron 1997). Presumably, most of us would not endorse such reasons for helping or not. One might worry that this undermines freedom and responsibility to at least some degree, insofar as it shows we rarely act on reasons we would endorse

upon reflection (Nahmias 2007). (Similar "situationist" results have led some to skepticism about robust character traits, and thus to criticize virtue ethics insofar as it relies on them—see Ethics and Psychology §2.)

One response to this challenge calls for reconceiving traditional constraints on agency, freedom, and responsibility. Perhaps, for example, we can only be held properly accountable for our actions if we abandon the idea that to be responsible we must always direct our actions through self-reflection (Doris 2015). Philosophical theories that would otherwise seem theoretically unsatisfactory in the abstract may deserve reconsideration in light of accumulating evidence about the nature of the human mind.

2. Egoism and Altruism

Morality sometimes requires beneficence, but it can seem morally problematic to do so for an ulterior purpose, such as self-interest. *Psychological egoism* maintains that we are always ultimately motivated by what we perceive to be in our own self-interest. While psychological egoists admit that one can care about the well-being of others, they maintain that such desires are not *ultimate* (or intrinsic)—they are merely instrumental to a desire for one's own benefit (see Egoism and Altruism). This theory has not been defended by many philosophers, but some have argued that empirical work lends it some credence (e.g. Slote 1964; Morillo 1990). Despite its lack of popularity, attention has been drawn to psychological egoism in light of work in social psychology, as well as the apparently weak philosophical foundation on which rejection of the view rests (Sober & Wilson 1998, ch. 9).

Much discussion of egoism involves evolutionary theory, especially given the proliferation of literature on "altruism" (see Units and Levels of Selection §2). One might think, for example, that we must be fundamentally self-interested because the evolution of our species via natural selection is governed by "selfish" genes that simply "seek" to replicate themselves; evolution makes altruism impossible. But this line of thought conflates evolutionary versus psychological senses of "altruism" and related terminology (Sober & Wilson 1998). Whether psychological egoism is true turns on whether all of one's ultimate desires concern one's own benefit. It would take more than the basic tenants of evolutionary theory to establish this, since "selfish" genes could, in principle, just as easily produce an ultimate desire for the well-being of others as they can an ultimate desire to for self-preservation. The question is whether it is more likely that human psychology evolved with altruistic ultimate desires in its repertoire. Philosopher Elliott Sober and biologist David Sloan Wilson (1998) have argued against psychological egoism precisely by appealing to the comparatively weak reliability of an egoistic mental mechanism in generating certain behavior, such as parental care (for criticism, see Stich, Doris, and Roedder 2010, §3).

Addressing a debate about motivation by appeal to evolutionary theory is rather tricky. An arguably more direct empirical approach is employed by those who study the mind more directly. Neuroscientists studying the brains of humans and other mammals, for example, may seem to have revealed that our actions are ultimately driven by pleasure and the avoidance of pain. After all, neuroscience thus far has identified a "reward center" of the brain, which regulates action, and it turns out to be intimately tied to pleasure (Morillo 1990). Yet recent research indicates that pleasure is dissociable from motivation. The behavior of rats, for instance, can be affected by increasing or

decreasing dopamine levels, independently of pleasure. When addicted to a substance, they can be motivated to obtain it even if they do not show normal signs of deriving pleasure from it. As the neuroscientist Kent Berridge and his collaborators have put it, different structures in the brain regulate "wanting" or motivation and "liking" or pleasure (Schroeder 2004, ch. 3; Holton 2009, ch. 5).

Another approach to altruism emerges in psychological research on empathy-induced helping behavior. The key starting point is the finding that higher levels of empathy felt for someone believed to be in need tend to increase rates of helping that person (the *empathy-helping relationship*). This well-established effect, however, does not prove that true altruism exists, since the ultimate motivation could be to benefit oneself. For example, one popular account among psychologists is that taking on another's perspective when empathizing causes one to blur the distinction between oneself and the other. Thus, concern for the well-being of the "other" isn't really altruistic (for criticism, see May 2011).

In any case, a series of experiments conducted over several decades seem to rule out many, if not all, of the relevant egoistic explanations. For example, in one experiment, subjects were asked to observe a fellow undergraduate, Elaine, receive some mild electric shocks. After several trials, the experimenter led participants to believe that Elaine is reacting badly to the shocks due to a traumatic past experience she had with an electric fence. They were then asked to help Elaine by taking the rest of the shocks in her stead. Some subjects, however, were experiencing higher levels of empathy, and some in that group were led to believe they would have to finish watching Elaine receive the rest of the shocks if they didn't help, as opposed to those who believed they could simply leave. According to one egoistic hypothesis, empathically aroused individuals tend to help more only because empathy makes watching another suffer especially unpleasant, and they would rather help than continue enduring this. If this is true, we should expect higher empathy to increase helping only in those who believe they must endure further empathic arousal upon choosing not to help. Yet this is not the case: several experiments have shown that those experiencing higher levels of empathy are still more likely to help whether or not they could easily escape the situation (Batson 2011, p. 96ff).

Moreover, the results of such experiments all conform to an altruistic theory, the *empathy-altruism hypothesis*, which states that empathy induces an altruistic ultimate desire for the welfare of the victim (Batson 2011). If this is correct, we have empirical evidence for the existence of altruism in humans, which entails that psychological egoism is false. While many agree the experiments have clearly ruled out a number of egoistic hypotheses, some believe there are plausible ones that remain unscathed (see e.g. Sober & Wilson 1998, ch. 8; Stich, Doris, and Roedder 2010, §4).

3. Moral Judgment and Motivation

Many of the issues dividing moral theorists rest on claims about how we come to judge things as right and wrong (see Moral Judgement), as well as what motivates us to act in accordance with such judgments. Two intimately related issues in this arena are (a) the connection between moral judgment and motivation, and (b) the role of "reason" in moral motivation.

Ethicists have long thought that there is an important connection between moral judgment and moral motivation. For example, if I believe I should accede to my friend's request to take her

to the airport, then I will at least typically have some motivation to do so. While perhaps I may lie in the end, claiming I have prior commitments, the "defeasible" motivation is still there. *Strong motivational internalists* believe this connection is necessary: making a moral judgment necessarily entails having some corresponding motivation to act in accordance with it, even if it is ultimately overridden by something else, like self-interest (see Moral Motivation §1).

This strong form of internalism, however, can be challenged by reference to empirical evidence on our motivational capacities (Roskies 2003). Consider the famous Phineas Gage and other *VM patients*—those with so-called "acquired sociopathy," studied at great length by Antonio Damasio (1994) and his collaborators. Often suffering from lesions in the ventromedial prefrontal cortex (vmPFC) of the brain, these patients have varying deficits in their ability to feel certain emotions and engage in pro-social behavior. Unlike psychopaths born with rather extreme anti-social tendencies (Nichols 2004, ch. 3), VM patients are arguably competent with moral terms and concepts, as evidenced by their typically high scores on Kohlberg's moral reasoning tests, for example. Yet various studies of their reactions to moral stimuli, such as low skin-conductance responses and self-reports, indicate that they often do not have the corresponding motivation to act in accordance with their moral judgments. If this is a correct description of their state of mind, VM patients are counter-examples to strong internalism: they make moral judgments but at least sometimes lack the corresponding motivation (see also Ethics and Psychology §3).

Replies to this empirical objection to internalism must cast doubt on the claim that VM patients make genuine moral judgments or the claim that they lack the corresponding motivation. One might argue, for example, that the deficit in acquired sociopathy is one of general decision-making and has little to do with moral or prosocial motivation (Kennett & Fine 2008). Moreover, much of the evidence comes from a few patients who were tested on their ability to make judgments about what other people should do. Yet the internalist might insist that the kinds of moral judgments that directly generate motivation are judgments about what *oneself* ought to do in some particular *circumstances* ("in situ"). Yet there is some evidence that VM patients struggle with exactly these kinds of judgments (Kennett & Fine 2008). So the internalist connection may remain in place: when patients do lack corresponding motivation it is due to a deficit in the relevant kind of decision-making about what one ought to do in particular circumstances.

A related, though distinct, issue is the role of "reason" in moral motivation—a la Hume's famous dictum that reason is the "slave of the passions" (see Hume, David §10). Assuming, in a rather stipulated manner, that the faculty of reason produces beliefs, contemporary philosophers address this perennial issue by focusing on what role beliefs can play in motivation. They focus in particular on normative or evaluative beliefs, such as beliefs about what one ought to do (see Moral Motivation §3 & §7). *Neo-Humean* philosophers maintain that the only role for normative beliefs is to determine how to satisfy our antecedent desires. For example, suppose I believe that I ought to loan my sister some money. According to the neo-Humean, the only role this belief can play in my motivation is to help satisfy an antecedent desire, and the only relevant desire seems to be this: the desire to do whatever I believe I should (e.g. Mele 2003, ch. 4). Those in the *rationalist* tradition, however, maintain that normative beliefs can generate a desire to act as the beliefs dictate, independent of any antecedent desire (e.g. Darwall 1983, esp. p. 39; Korsgaard 1986).

At least one relevant question here is causal: Can normative beliefs in humans produce a desire without this serving or furthering some antecedent desire? Empirical research can help us

answer such questions. One might suggest, for example, that the neo-Humean picture is best supported by what neuroscience tells us about the human brain (Schroeder, Roskies, and Nichols 2010). The brain's "reward center," after all, appears to be essential for normal motivation. Yet it also seems to be the seat of our ultimate desires, as it is involved in the kind of learning and pleasure associated with basic motivation (Morillo 1990; Schroeder 2004). Actions whose neural antecedents do bypass the reward center and originate in higher cognitive structures, however, are not exactly the paradigms of morality: habitual acts and tics involved in Tourette's syndrome, for example (Schroeder 2004, ch. 5.3).

While such research into the neurophysiological realization of mental states is promising and suggestive, granting the forgoing claims only establishes that normal, non-pathological *action* must be preceded by desires. This is often accepted by rationalists who can grant that all intentional action requires desire somewhere in the causal story (e.g. Darwall 1983). The crucial question for further empirical evidence to address is whether these desires must always precede normative *beliefs* that then serve or further the antecedent desires.

4. Weakness and Strength of Will

Everyone agrees that people do not always do what they think they ought. We all sometimes succumb to temptation, exhibiting a kind of moral weakness when the action has moral significance (e.g. adultery). Some of us are characteristically weak-willed, while others are typically strong-willed, and each individual's willpower fluctuates depending on the circumstances (e.g. when intoxicated).

Interesting philosophical puzzles arise with such phenomena, but some have been concerned with a precise characterization of them in the first place, or whether they even exist at all. Some have defined "weakness of will" as *akrasia*—i.e. acting, or having a disposition to act, against one's *judgment* about what is best (see Akrasia). Others have focused on action that is contrary to what one *intends* to do (Holton 2009, ch. 4). But there is some empirical evidence that neither of these exhausts the ordinary notion of being weak-willed. Both factors seem to play some role, as do apparently non-psychological elements, such as the moral valance of the action (May & Holton 2012). The ordinary notion of self-control and its failure might be more expansive than traditional philosophical conceptions, making its reality more plausible, despite certain puzzling features.

However we construe weakness, it's opposite—strength of will—also deserves attention (see Self-Control). Focusing on intentions, we can inquire into what mental states and mechanisms underlie our ability to stick to what we've planned to do. The famous "marshmallow studies" conducted by the developmental psychologist, Walter Mischel, illustrate the ordinary phenomenon and the varied strategies for self-control. Mischel and his collaborators offered young children an opportunity to have one treat now or to wait and receive two. However, in some conditions, participants had to wait until an unspecified time while the tempting treats sit right in front of their eyes. Only some succeeded at "delaying gratification" and various strategies for exercising willpower have arisen, including self-distraction and reappraisal (e.g. imaging the tempting item as something unappetizing). Such skills are evidently integral to long-term success in one's personal and professional life, as early abilities to resist temptation predict higher test scores, lower drug use, better physical and mental health, less bullying, higher self-esteem, and more (Mischel et al 2010).

Willpower does not seem entirely fixed by one's genes either. Consider the phenomenon of *ego-depletion* in which self-control resources are used up over time. Social psychologists, especially Roy Baumeister and colleagues (Muraven et al 1998) have discovered that we are less likely to persist in activities that require self-regulation if we have recently already done so. For example, people cannot hold a handgrip exerciser for as long if they recently had to suppress emotional reactions while watching a sad movie clip. Strength of will, it seems, works like a muscle in that it can be strengthened, weakened, and has a limited store of energy on which to draw.

Importantly, the effects of ego-depletion can occur across a variety of domains, such as dieting and solving puzzles. A neo-Humean account would attempt to explain this only in terms of beliefs and desires. But such explanations might have difficulty accounting for the global effects of ego-depletion. Why, for example, would a desire to avoid eating some tempting food item affect one's desire to persist in holding a handgrip exerciser? Those parting with the Humean tradition may posit intentions as a distinct mental state, not reducible to beliefs and desires (see Intention §2). But one might go even further and posit a faculty of willpower that is distinct from these various states of mind (Holton 2009, ch. 6). This appears to have the advantage of explaining the systematic effects of ego-depletion.

Examining such research, one might conclude that an even more general phenomenon is occurring here. A scientifically fruitful categorization of cognitive processes divides them into two basic kinds, yielding a "dual-processing" approach. *System 1* processes are quick, automatic, relatively independent of conscious control, and so on. *System 2* processes are slow, effortful, controlled, etc. Weakness of will, then, may be encompassed in the more general category of actions that are predominantly the result of System 1 resources when those from System 2 have been recently exhausted (Levy 2011). This would nicely model the phenomenology of weakness: sticking to the plan of doing what's best is effortful and often giving in feels like letting a passion take over. Further connecting the more ordinary phenomena of weakness and strength of will to categories in cognitive science may help illuminate the philosophically interesting issues surrounding them, such as the differential roles of beliefs, desires, emotions, and attention (see Sripada 2010).

5. Moral Intuitions

Ethical theories are often tested against our immediate, pre-theoretical, and confident judgments about morally significant cases—what we might call "moral intuitions." Consider, for example, the widely shared judgment that slavery is immoral or that Hitler's campaign of genocide was evil. It counts against a theory to at least some extent if it conflicts with such clear intuitions, at least those arising in mature individuals (see Moral Development). If we are going to place weight on such intuitions, we may rightly ask: What drives them?

One recent line of empirical research focuses on the role of emotion as opposed to reasoning in moral judgment. In particular, Jonathan Haidt and his colleagues have conducted a number of experiments purporting to reveal a starring role for emotions. Disgust in particular has been the most extensively studied in moral psychology. In one experiment, participants recorded their moral judgments in response to various hypothetical scenarios either at a clean desk or a disgusting desk (with old food, sticky substances, etc.). Those who scored highly on their ability to

perceive changes in their bodily state tended to rate some of the actions as morally worse (Schnall, Haidt, Clore, & Jordan 2008). In another study, prior to rating the morality of a set of hypothetical cases, participants were randomly assigned to either drink some water, something bitter, or something disgusting. Participants tended to rate the behavior in some of the cases as morally worse when disgusted (Eskine et al 2011).

The role of incidental emotions in ordinary moral judgment may be limited, however. Disgust alone doesn't seem to alter the valence of moral judgments; at best it makes them slightly harsher (May 2014). Moreover, a meta-analysis of over 50 experiments suggests that the disgust effect is quite small among published studies and non-existent among unpublished ones (Landy & Goodwin 2015). From the extensive research on one emotion, we can perhaps conclude little about the role of emotions generally in all of moral cognition. Indeed, even Haidt (2012) now emphasizes emotion less, focusing instead on automatic intuitions, which only sometimes involve emotional reactions.

In addition to arguing that emotion drives moral judgment, some have added that reasoning's role is merely in *post hoc rationalization*. Haidt's (2012) slogan: "Intuitions come first, strategic reasoning second." In a series of studies, participants read cases of "harmless taboo violations" that evoke moral condemnation but that apparently lack harm. One case, for example, describes a brother and sister who only once engage in consensual incest with ample protection and without damaging their relationship. In interviews, people are in a state of *moral dumbfounding*—they are convinced the action is morally wrong, but are unable to find reasons for this judgment. Haidt (2001) suggests that this is largely mere rationalization: moral judgment is at least typically generated and sustained by automatic reactions (compare System 1), and conscious reasoning primarily comes in after the fact to defend the intuitive judgment (compare System 2), not to seek the truth.

On Haidt's *social intuitionist* theory of moral judgment: (a) moral beliefs are formed primary based on automatic intuitions that (b) are rather insensitive to good reasoning (one primarily invents reasons to convince others of one's views). Critics of Haidt's view tend to challenge one or both of these key features of the model.

Some commentators, for example, grant that moral beliefs are largely formed by automatic intuitions, but deny that these are insensitive to good reasoning. The research suggests only that, much like the rest of our mental lives, *conscious* reasoning is not as prominent as one might think and often involves confabulation or rationalization. Some studies probe further, though, and reveal that our automatic moral intuitions have a rich computational structure, informed by complex yet tacit reasoning with moral rules (e.g. Mikhail 2011; Mallon & Nichols 2010). A common comparison is with linguistic judgment: intuitive reactions about the meaning or grammaticality of a sentence are underwritten by tacit reasoning involving complicated linguistic rules that serve us well. The inability to articulate one's reasons is no sign of unreasoned belief. In fact, this is arguably a general feature of nearly all domains of cognition. For example, we automatically infer the mental state of another person (e.g. *He's sad*) in the absence of conscious reasoning, but not necessarily in the absence of good reasoning that amounts to more than rationalization.

Other critics of social intuitionism do not even concede that moral judgment is largely driven by automatic intuitions that are relatively insensitive to conscious, reflective reasoning. There is, after all, experimental evidence suggesting that automatic responses, such as implicit racial biases, can be corrected for based on conscious moral reasoning (Kennett & Fine 2009; Mallon & Nichols

2010, §2). Moreover, one might worry that a cognitive process cannot yield a genuine moral judgment unless reflective reasoning can play a role in shaping the mental states it generates. Haidt's model of moral cognition might look like the automatic social cognition of dogs or young children, who are arguably incapable of genuine moral judgment (see e.g. Kennett & Fine 2009).

6. Moral Knowledge

Uncovering the psychological origins of our moral judgments might vindicate or debunk them. But the long tradition of scrutinizing the genealogy of morality is undoubtedly aided by rigorous empirical evidence (see Experimental Philosophy, §3).

Take disgust, for example. Many think its influence on moral judgment is troubling. Kelly (2011) argues that disgust evolved to be overly sensitive to the threat of pathogens, and the co-opting of that mechanism for moral judgment leads to an unreliable influence. Haidt (2012), however, is untroubled by the apparent role of disgust in moral cognition. He argues that theorists need to accept that an important and respectable aspect of morality concerns sanctity and divinity, which disgust helps to identify. Haidt urges: "There's more to morality than harm and fairness."

Consider next the counter-intuitive implications of various brands of consequentialism (see Utilitarianism; Consequentialism). Utilitarianism, for example, seems committed to the moral acceptability of unfair actions and policies if they lead to more overall happiness. If framing an innocent man will placate a mob bent on great destruction, then the utilitarian counsels injustice. A wealth of empirical research attempts to pry apart characteristically utilitarian from deontological intuitions and probe their origins.

Most of the relevant studies involve presenting participants with variants on the famous trolley cases, in which (roughly) a protagonist attempts to save five people from being run over by a train, but at the cost of one death to a different person. Consider two key scenarios. In the *Side-Track case*, five workers are tied to the tracks on the trolley's path, but a switch next to the protagonist can divert the trolley onto a track with only one person stuck on it. Most philosophers believe it is morally permissible to throw the switch, which saves the five but kills the one. In the *Footbridge case*, while five workers are stuck on the tracks, one large man is on the footbridge, and the protagonist can only stop the train to save the five by pushing the man into the trolley's path. The characteristically utilitarian pair of responses is that it's morally permissible to flip the switch in Side-Track *and* push the man in Footbridge. But deontologists and other non-consequentialists have often tried to preserve the more intuitive verdict that pushing in Footbridge is impermissible. One rationale is that pushing uses the man as a mere means to an end, whereas the death in Side-Track is merely a foreseen but unintended side-effect of saving the five. (The cases are tied especially to debates about certain deontological principles; see Principle of Double Effect; Inviolability, §4.)

Many studies now reveal that most people believe flipping the switch in Side-Track is permissible, but pushing in Footbridge is not (e.g. Mikhail 2011). While such results comport well with deontological theories of morality and moral judgment, what appears to drive these intuitions may not. Brain imaging studies suggest that areas associated with emotion—e.g. vmPFC, which projects to the amygdala—are more active in generating characteristically deontological judgments (e.g. in Footbridge) as opposed to consequentialist ones (e.g. in Side-Track). And the correlation

between affect and deontological judgments supports an inference to a causal relationship when conjoined with studies of patients with brain lesions. For example, those with emotional deficits (e.g. patients with frontotemporal dementia; patients with damage to the vmPFC) are more likely to report consequentialist intuitions about cases like Footbridge, which suggests that the missing affect plays a causal role in generating the deontological intuition in normal subjects. These and other data indicate, contrary to a traditional theme in the philosophical literature, that deontological intuitions are driven more by emotion, while consequentialist judgments rely more on our distinctive reasoning capabilities (for review, see Cushman, Young, and Greene 2010; Greene 2013).

It is difficult to argue that certain intuitions are unreliable just because they are emotionally-charged (Berker 2009). What's more troubling, however, is additional evidence suggesting that the relevant intuitions are based on whether the act in question is up-close and personal, rather than impersonal. In particular, we are more likely to think it's morally acceptable to sacrifice a man on a footbridge to save five if one simply has to flip a switch to open a trap door, rather than pushing. But it seems morally irrelevant whether an act involves such personal force or is prototypically violent (Greene 2013). We can't justifiably treat cases like Side-Track and Footbridge differently based on a morally irrelevant factor.

Based in part on the preceding body of research, Joshua Greene (2013) boldly argues in favor of utilitarianism. The key objections to utilitarianism rest on intuitions that Greene regards as utterly unwarranted. We shouldn't trust them because they are an automatic, emotionally-driven (System 1) response to morally irrelevant factors like personal force. Moreover, Greene argues that utilitarian intuitions can be trusted because they're driven by cognitive processes (in System 2) that are flexible, controlled, and well-suited to serve as a common moral currency to resolve moral debates across cultures.

Such conclusions are naturally controversial, generating replies from both philosophers and scientists. For example, some further brain imaging suggests that heightened activity in areas associated with controlled processing (System 2) correlates with *counter-intuitive* moral judgments generally, which are not always utilitarian (Kahane et al 2012). Other studies suggest that the dilemmas used to distinguish utilitarian from deontological intuitions fail to track these categories at all. Allegedly "utilitarian" intuitions fail to correlate with judgments, traits, or behaviors reflecting an impartial concern for the greater good (Kahane et al 2015). More theoretical replies take issue with Greene's claims about which brain areas correspond to automatic versus controlled processing (e.g. Klein 2011).

Thus, while there is growing evidence for a general dual-process approach to moral cognition, such empirical debunking arguments rely on tendentious claims requiring careful scrutiny. Clearly, though, empirical evidence can play an important role in uncovering the genealogy of our moral judgments, which can lead to conclusions about whether certain moral beliefs constitute knowledge.

7. References and Further Reading

Baron, R. (1997). "The Sweet Smell of… Helping: Effects of Pleasant Ambient Fragrance on Prosocial Behavior in Shopping Malls." *Personality and Social Psychology Bulletin* 23:498–503. (Provides evidence that pleasant smells can increase helping behavior; replicates similar results from previous studies.)

Batson, C. D. (2011). *Altruism in Humans*. New York: Oxford University Press. (Updated defense of the of the existence of altruism in humans; addresses new challenges to the empathy-altruism hypothesis; includes more than social-psychological data.)

Berker, Selim (2009). "The Normative Insignificance of Neuroscience." *Philosophy and Public Affairs* 37(4):293-329. (Argues the only way to keep the neuroscientific case against deontological intuitions from being fallacious is to recast it as independent of brain imaging data altogether.)

Cushman, F. Young, L. & J. Greene (2010). "Multi-System Moral Psychology." In *The Moral Psychology Handbook*, J. M. Doris & The Moral Psychology Research Group (eds.), Oxford University Press. (Develops a dual-process model of various data on moral judgments about physical harms, comparing cognitive and conscious processes with affective and intuitive processes.)

Darwall, S. (1983). *Impartial Reason*. Cornell University Press. (Argues clearly for an anti-Humean view of both motivation and reasons, along the lines of other rationalists, such as Nagel, Korsgaard, Wallace, and Scanlon.)

Doris, J. M. (2015). *Talking to Our Selves: Reflection, Ignorance, and Agency*. New York: Oxford University Press. (Argues that empirical research precludes the kind of reflective self-direction that many philosophers seem to think is crucial to agency and responsibility.)

Doris, J. M. & **The Moral Psychology Research Group** (2010). *The Moral Psychology Handbook*. Oxford University Press. (Currently the most up-to-date and comprehensive source for discussion of empirically-informed moral psychology; includes many topics not discussed here.)

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). "A Bad Taste in the Mouth: Gustatory Disgust Influences Moral Judgment." *Psychological Science* 22(3):295-299.

Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin. (Provides a wide range of empirical evidence in favor of utilitarianism and suggesting that non-utilitarian intuitions should not be trusted.)

Haidt, J. (2001). "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108: 814-834. (Reviews empirical evidence that moral judgments are typically formed by quick and automatic reactions while reasoning plays a more secondary role after the fact as a post hoc construction.)

Haidt, J. (2012). *The Righteous Mind*. New York: Pantheon. (Argues that moral judgment arises primarily from six different types of intuitions, the foundations of which evolved to help groups succeed against other groups.)

Holton, R. (2009). *Willing, Wanting, Waiting*. Oxford: Clarendon Press. (A generally anti-Humean account of various aspects of the will, including intention and choice; appeals to both philosophical and empirical evidence.)

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). "The Neural Basis of Intuitive and Counterintuitive Moral Judgment." *Social Cognitive and Affective Neuroscience* 7(4):393-402. (Provides some brain imaging data challenging Greene's account of non-utilitarian moral intuitions.)

Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). "'Utilitarian' Judgments in Sacrificial Moral Dilemmas Do Not Reflect Impartial Concern for the Greater Good." *Cognition* 134:193-209. (Provides evidence that there is no connection between utilitarian attitudes and intuitions commonly alleged to be characteristically utilitarian.)

Kane, R. (1999). "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism." *Journal of Philosophy* 96:217–240. (Argues that free will is incompatible with determinism, starting from the claim that such a view is the default, common sense one.)

Kelly, D. (2011). *Yuck!: The Nature and Moral Significance of Disgust*. Cambridge, MA: MIT Press. (Argues that disgust evolved to be oversensitive to pathogens and is not a reliable guide to morality.)

Kennett, J. & Fine, C. (2008). "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths.'" In W. Sinnott-Armstrong (ed.) *Moral Psychology, Vol. 3*, MIT Press. (Argues that some evidence from psychopathology does not pose a problem for the thesis that moral judgment always issues in corresponding motivation.)

Kennett, J., & Fine, C. (2009). "Will the Real Moral Judgment Please Stand Up?" *Ethical Theory and Moral Practice* 12(1):77-96. (Critiques Haidt's social intuitionist model of moral judgment.)

Klein, C. (2011). "The Dual Track Theory of Moral Decision-making." *Neuroethics* 4(2):143-162. (Argues that the brain areas apparently correlated with utilitarian and deontological intuitions aren't specific to cognitive and emotional processing, respectively.)

Korsgaard, C. (1986). "Skepticism about Practical Reason." *The Journal of Philosophy* 83(1):5–25. (Argues that neo-Humean accounts of moral motivation cannot be supported without assuming a neo-Humean theory of the norms of practical reasoning.)

Landy, J. F. & Goodwin, G. P. (2015). "Does Incidental Disgust Amplify Moral Judgment? A Meta-analytic Review of Experimental Evidence." *Perspectives on Psychological Science* 10(4): 518-536. (Analysis of dozens of studies which suggests that merely feeling grossed out hardly influences moral judgment, if at all).

Levy, N. (2011). "Resisting 'Weakness of the Will.'" *Philosophy and Phenomenological Research* 82 (1):134-155. (Argues that weakness of will is not a psychological kind by appealing to empirical research on self-control, such as ego-depletion.)

Mallon, R. & S. Nichols (2010). "Rules." *The Moral Psychology Handbook*, J. M. Doris & The Moral Psychology Research Group (eds.). Oxford University Press. (Defends the claim that moral judgment is partly guided by the representation of moral rules, rather than solely emotion, for example.)

May, J. (2011). "Egoism, Empathy, and Self-Other Merging." *Southern Journal of Philosophy* 49(s1):25-39, Spindel Supplement: Empathy & Ethics, Remy Debes (ed.). (Critiques arguments for egoism that appeal to the idea that we blur the distinction between ourselves and others, especially when we feel empathy for them.)

May, J. (2014a). "Does Disgust Influence Moral Judgment?" *Australasian Journal of Philosophy* 92(1): 125–141. (Argues that research shows disgust only slightly influences moral judgment, which cannot support many philosophical arguments.)

May, J. (2014b). "On the Very Concept of Free Will." *Synthese* 191 (12):2849-2866. (Evidence of both compatibilist and incompatibilist elements in ordinary thinking, modeled on a prototype theory of concepts.)

May, J. & R. Holton (2012). "What in the World Is Weakness of Will?" *Philosophical Studies* 157(3):341–360. (A defense of the ordinary notion of weakness of will as involving multiple factors, including intention-violations, judgment-violations, and moral valence.)

Mele, A. (2003). *Motivation and Agency*. New York: Oxford University Press. (Covers a wide range of philosophical topics related to motivation; provides a rich conceptual framework for discussing motivation.)

Mikhail, J. (2011). *Elements of Moral Cognition*. Cambridge University Press. (Argues for a rationalist and deontological conception of moral judgment as a capacity that's largely innate, modular, and automatic.)

Miller, C. (2009). "Social Psychology, Mood, and Helping: Mixed Results for Virtue Ethics." *The Journal of Ethics* (Special Issue on Situationism) 13:145–173. (Reviews effects of mood on helping; argues the research does not impugn the existence of some character traits.)

Morillo, C. (1990). "The Reward Event and Motivation." *The Journal of Philosophy* (87)4:169–186. (Defense of psychological hedonism based on work in neuroscience, especially experiments on rats and their "pleasure centers.")

Muraven, M., D. M. Tice, & R. F. Baumeister (1998). "Self-control as a Limited Resource: Regulatory Depletion Patterns." *Journal of Personality and Social Psychology* 74(3):774–89 (Provides evidence that self-control relies on a store of energy that can be depleted.)

Murray, D. & Nahmias, E. (2014). "Explaining Away Incompatibilist Intuitions." *Philosophy and Phenomenological Research* 88(2):434–467. (Experimental evidence that ordinary intuitions only seem incompatibilist because due to misunderstanding how determinism affects decisions.)

Nadelhoffer, T., Nahmias, E. & S. Nichols (2010). *Moral Psychology: Historical and Contemporary Readings*. Wiley-Blackwell. (A collection juxtaposing important empirical and non-empirical readings; includes useful introductions to sections.)

Nahmias, E., Morris, S., Nadelhoffer, T. & J. Turner. (2006). "Is Incompatibilism Intuitive?" *Philosophy and Phenomenological Research* 73: 28–53. Reprinted in Knobe & Nichols (2008). (Evidence that ordinary intuitions reflect the conviction that determinism is actually compatible with free will and moral responsibility.)

Nahmias, E. (2007). "Autonomous Agency and Social Psychology." In *Cartographies of the Mind*, eds. M. Maraffa, M. De Caro and F. Ferretti. Springer. (Argues that free will may be undermined to a certain degree based on situationist research in social psychology.)

Nichols, S. (2004). *Sentimental Rules*: *On the Natural Foundations of Moral Judgment*. Oxford University Press. (Argues that moral judgments are formed by applying norms, which are based primarily on emotional responses.)

Nichols, S. (2015). *Bound: Essays on Free Will and Responsibility*. Oxford University Press. (Argues that we should often treat each other as free and responsible even though we tend to think such a practice is incompatible with the plausible thesis that our choices are determined.)

Nichols, S. & J. Knobe (2007). "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." *Noûs* 41(4):663–685. Reprinted in Knobe & Nichols (2008). (Evidence that ordinary people tend to report compatibilist intuitions about moral responsibility and determinism only when the cases judged elicit more emotional reactions.)

Roskies, A. (2003). "Are Ethical Judgments Intrinsically Motivational? Lessons From 'Acquired Sociopathy.'" *Philosophical Psychology* 16(1):51–66. (Argues that patients with damage to the ventromedial prefrontal cortex are counter-examples to a strong form of motivational internalism.)

Schnall, S., J. Haidt, G. L. Clore & A. H. Jordan (2008). "Disgust as Embodied Moral Judgment." *Personality and Social Psychology Bulletin* 34:1096–1109. (Evidence that disgust makes some people provide harsher moral judgments about some hypothetical scenarios.)

Schroeder, T. (2004). *Three Faces of Desire*. New York: Oxford University Press. (Philosopher's defense of a reward-based theory of desire, grounded in empirical work largely from neuroscience.)

Schroeder, T., A. Roskies, & S. Nichols (2010). "Moral Motivation." *The Moral Psychology Handbook*, J. M. Doris & The Moral Psychology Research Group (eds.). Oxford University Press. (Examination of the neurological basis of moral motivation in the brain; egoism addressed briefly at the end.)

Sinnott-Armstrong, W. (2008). *Moral Psychology*, 3 Vols. MIT Press. (Massive collection of previously unpublished articles and replies from philosophers and scientists on the evolution, cognitive science, and neuroscience of morality.)

Slote, M. A. (1964). "An Empirical Basis for Psychological Egoism." *Journal of Philosophy* 61(18): 530–537. (Philosopher's defense of psychological egoism based on empirical work in psychology at the time, which was largely behavioristic in nature.)

Sober, E. & D. S. Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press. (Argues that philosophical arguments and social psychology don't provide sufficient evidence against psychological egoism, whereas evolutionary theory does.)

Sripada, C. (2010). "Philosophical Questions about the Nature of Willpower." *Philosophy Compass* 5 (9):793-805. (Overview and synthesis of recent philosophical and empirical work on weakness and strength of will.)

Stich, S., J. M. Doris, & E. Roedder (2010). "Altruism." In *The Moral Psychology Handbook*, J. M. Doris and the Moral Psychology Research Group (eds.). Oxford University Press. (Overview of the philosophical, biological, and psychological work relevant to egoism.)