

# 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, 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 often think of ourselves as possessing free will, which undergirds our being responsible for what we do; we often believe we can be ultimately concerned for the welfare of another; and so on. Yet these 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, the empirical methods of the social sciences are at least an additional tool to bring to bear on the issue. Second, we can investigate empirically whether our ordinary thinking is correct. This typically involves checking the empirical adequacy of philosophical theories, assessing directly any claims made about moral cognition, motivation, and so forth.

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, moral theories, such as deontology and utilitarianism, often rely on intuitive judgments about what one ought to do in various hypothetical cases. Empirical research can again serve as an additional tool to determine what exactly the intuitions are and which psychological processes generate them.

## Table of Contents

1. Moral Responsibility and Free Will
2. Egoism and Altruism
3. Moral Judgment and Motivation
4. Weakness and Strength of Will
5. Moral Judgments and Intuitions
6. References and Further Reading

---

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. As such, some philosophers have presented non-philosophers 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). (For further discussion of variation among intuitions about moral responsibility, see Knobe & Doris 2010.)

Whether the incompatibilist intuitions are more reflective of ordinary thinking after all depends on whether we should take the affect associated with compatibilist intuitions as rendering them faulty in some way. If, for example, we assume emotion always distorts judgment, then the compatibilist intuitions might be construed as a “performance error.” In which case, reporting such intuitions would be akin to answering “Moses took two each” in response to the misleading question “How many of each animal did Moses take on the ark?”

While determinism poses a classical threat to free will and moral responsibility, some have argued that empirical research itself is likewise threatening. Social psychologists, for example, have long 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 free will or moral responsibility, assuming they require something like the capacity to act only on reasons one would endorse upon reflection (Nahmias 2007; Doris 2009). (Similar results have led some to skepticism about stable character traits, and thus to criticize virtue ethics insofar as it relies on them—see Ethics and Psychology §2.)

## 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 recently been drawn back 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).

Some discussion of egoism infuses 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 self-preservation as they can an ultimate desire to for the well-being of another. 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 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.

Moral philosophers 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 motivational 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 (VM) prefrontal cortex 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 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; Sinnott-Armstrong 2008, Vol. 3).

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 help my sister. 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 they 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 system,” 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—a claim which is often accepted on anti-Humean views (e.g. Darwall 1983). The crucial question for further empirical evidence to address is whether these desires must always precede normative *beliefs*.

## 4. Weakness and Strength of Will

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, while evaluative considerations do as well (May & Holton 2012).

However we construe weakness, its 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. Consider the phenomenon of *ego-depletion* in which self-control resources are used up over time. Social psychologists, especially Roy Baumeister and colleagues (Muraven, Tice, & Baumeister 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, guided by consciousness, 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. Tying the more ordinary phenomena of weakness and strength of will to categories in cognitive science in this way may help illuminate the philosophically interesting issues surrounding them (see Sripada 2010).

## 5. Moral Judgments and Intuitions

Ethical theories are often tested against our immediate, pre-theoretical 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. But 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 disgust. 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 more immoral (Schnall, Haidt, Clore, & Jordan 2008). In another study, participants highly susceptible to hypnotism were made to feel disgust upon hearing a morally-neutral word (e.g. “often”), and then presented with a set of hypothetical cases, some of which employed the term. Interestingly, on average participants rated the behavior in some of the cases with the disgust-inducing word as more morally wrong than those that didn’t include the word (Wheatley & Haidt 2005). At this early stage in the research, however, the exact extent of disgust’s role in ordinary moral judgment is unclear (see Mallon & Nichols 2010).

In addition to arguing that emotion drives moral judgment, some have added that reasoning’s role is merely in post hoc rationalization. In a series of studies, participants read cases designed to evoke moral outrage but that apparently lacked any paradigm moral transgressions (e.g. harm or violation of rights). One case, for example, describes a brother and sister who 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 by immediate emotional reactions (compare System 1), and reasoning primarily comes in after the fact to defend the intuitive judgment (compare System 2). (For some criticism, see Mallon & Nichols 2010, §2.)

The focus thus far has been on moral judgment generally, but some empirical research suggests that emotion drives only non-utilitarian moral judgments. 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 trolley, but at the cost of one death to a different person. Traditionally, there have been two key cases discussed. 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 change the trolley onto a track with only one person strapped to it. Most philosophers believe it is morally permissible to throw the switch, which saves the five but kills the one—only 8% believe it is impermissible (see Bourget & Chalmers 2009). In a key variant, the *Footbridge case*, while five workers are strapped to the tracks, one large man is on the footbridge, and the protagonist can stop the train to save the five only by pushing the man over to his death. Often, utilitarians and other

consequentialists will maintain that it is morally permissible to flip the switch in Side-Track and push the man in Footbridge. But deontologists and other non-utilitarians have often argued that the second case is impermissible, because (for example) it uses a person 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.)

Whether philosophers' intuitions are idiosyncratic, and what drives ordinary intuitions, can be examined empirically. One recent study with thousands of participants indicates that the vast majority believe flipping the switch in Side-Track is permissible, but pushing in Footbridge is not (Hauser et al. 2007). While such results comport well with deontological theories of morality and moral judgment, what appears to drive these intuitions may not. Brain imaging studies employing fMRI suggest that areas of the brain associated with emotion 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. VM patients) tend 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 affect, while consequentialist judgments rely more on our distinctive reasoning capabilities (for review, see Cushman, Young, and Greene 2010).

What do these empirical data on moral judgment tell us about morality? Joshua Greene (2008) has argued that rationalist deontology is implausible, since the intuitions on which it rests are driven by an emotional response to transgressions that are “up-close and personal,” which is a “contingent, nonmoral feature of our evolutionary history” (p. 70). Similarly, one might argue that alleged counter-examples to consequentialism are not trustworthy since they are generated in response to morally irrelevant factors, such as disgust (see Sinnott-Armstrong et al. 2010, §3.3). While such conclusions are controversial (see e.g. Berker 2009), many currently agree that the data support the hypothesis that moral judgment arises from at least two distinct processes: affective/intuitive and conscious/cognitive (à la dual-process models). However, much remains unclear at this early stage in the research, including how these processes interact, what their origin is, and whether conclusions about their role in judgments about physical harm (as in the trolley cases) generalize to other facets of morality, such as justice and care (see Cushman, Young, & Greene 2010). Thus, any conclusions about morality proper based on the empirical data are rather tentative, though the future is promising.



## 6. 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 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.)
- Bourget, D. & D. Chalmers** (2009). "PhilPapers Survey: Results." Available at: <<http://philpapers.org/surveys/results.pl>>. (A systematic survey of around 900 professional philosophers concerning their beliefs on various core philosophical issues.)
- 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, pp. 47-71. (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.** (2009). "Skepticism About Persons." *Philosophical Issues* 19 (1):57-91. (Argues that empirical research precludes the kind reflective self-direction that many philosophers seem to think is crucial to personhood or agency.)
- 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.)
- Greene, J.** (2008). "The Secret Joke of Kant's Soul." In Sinnott-Armstrong (2008), Vol. 3 (Neuroscience of Morality), pp. 35-80. (Marshals a range of empirical evidence for the claim that deontological moral judgments are generated by processes that are less cognitive and more emotional than consequentialist intuitions.)
- 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.)
- Hauser, M. D., Cushman, F. A., Young, L., Jin, R., & Mikhail, J. M.** (2007). "A Dissociation Between Moral Judgments and Justifications." *Mind & Language* 22 (1): 1–21. (Reports data

on intuitions about various trolley cases supporting the hypothesis that deontological principles, such as the Doctrine of Double Effect, underwrite some moral judgments.)

- Holton, R.** (2009). *Willing, Wanting, Waiting*. Oxford: Clarendon Press. (Provides a generally anti-Humean account of various aspects of the will, including intention and choice; appeals to both philosophical and empirical evidence.)
- 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.)
- Knobe, J. & J. M. Doris** (2010). "Responsibility." In Doris et al. (2010), pp. 321-354. (Appealing to empirical research on ordinary intuitions, argues against the "invariantist assumption" that judgments about moral responsibility should all be determined by the same criteria.)
- Knobe, J. & S. Nichols**, ed. (2008). *Experimental Philosophy*. Oxford University Press. (Collection of articles that primarily probe ordinary intuitions about various philosophical topics from free will to epistemology; also includes theoretical papers addressing this methodology.)
- 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.)
- 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, pp. 297-320. (Defends the claim that moral judgment is in part 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* Vol. 49, Spindel Supplement: Empathy & Ethics, Remy Debes (ed.), pp. 25-39. (Critiques arguments for psychological egoism that appeal to the idea that we blur the distinction between ourselves and others, especially when we feel empathy for them.)
- 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.)
- 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 psychological 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* Vol. 87, No. 4, pp. 169–186. (A recent defense of a kind 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.)
- Nadelhoffer, T., Nahmias, E. & S. Nichols** (2010). *Moral Psychology: Historical and Contemporary Readings*. Wiley-Blackwell. (A recent 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), pp. 81-104. (Argues 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, pp. 169–185. (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 in an interdisciplinary way that moral judgments are formed by applying norms, which are based primarily on sentiments or emotional responses.)
- 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), pp. 105-126. (Provides 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. (Provides some 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, pp. 72-110. (An examination of the neurological basis of moral motivation in the brain; psychological hedonism 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, psychologists, and others on the evolution, cognitive science, and neuroscience of morality.)
- Sinnott-Armstrong, W., L. Young, and F. Cushman** (2010). "Moral Intuitions." In *The Moral Psychology Handbook*, J. M. Doris & The Moral Psychology Research Group (eds.). Oxford University Press, pp. 246-272. (Discusses the hypothesis that some moral intuitions are

formed by heuristics just as certain non-moral beliefs are, as in the work of psychologists such as Kahneman, Tversky, and Gigerenzer.)

- Slote, M. A.** (1964). "An Empirical Basis for Psychological Egoism." *Journal of Philosophy* Vol. 61, No. 18, pp. 530–537. (A 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. (Brief and informative 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, pp. 147–205. (An overview of the philosophical, biological, and psychological work relevant to the egoism-altruism debate. Much of the focus is on Sober & Wilson and Batson.)
- Wheatley, T. & J. Haidt** (2005). "Hypnotic Disgust Makes Moral Judgments More Severe." *Psychological Science* 16: 780-84. (Provides evidence that disgust can make those highly susceptible to hypnosis provide harsher moral judgments about some hypothetical scenarios.)