

**Beyond Point-and-Shoot Morality:
Why Cognitive (Neuro)Science Matters for Ethics***

Joshua D. Greene
Department of Psychology
Harvard University

Forthcoming in *Ethics*

Abstract: In this article I explain why cognitive science (including some neuroscience) matters for normative ethics. First, I describe the dual-process theory of moral judgment and briefly summarize the evidence supporting it. Next I describe related experimental research examining influences on intuitive moral judgment. I then describe two ways in which research along these lines can have implications for ethics. I argue that a deeper understanding of moral psychology favors certain forms of consequentialism over other classes of normative moral theory. I close with some brief remarks concerning the bright future of ethics as an interdisciplinary enterprise.

Running header: Why Cognitive (Neuro)Science Matters for Ethics

* Many thanks to John Mikhail, Henry Richardson, and other participants in the symposium on Experiment and Intuition in Ethics held at Georgetown University, April 2011. Thanks to James Weinstein and other participants at the Origins of Morality Conference, Arizona State University, November, 2010. Thanks to Tommaso Bruni, John Doris, Steven Frankland, Geoff Holtzman, Dylan Murray, and Joe Paxton for comments on this manuscript.

The cognitive science of ethics is booming, thanks in no small part to philosophers.¹ Nevertheless, many philosophers wonder whether this, or any, empirical research could have implications for foundational questions in normative ethics. In this article I will explain why cognitive science matters for ethics. More specifically, I will defend the following thesis:

Science can advance ethics by revealing the hidden inner workings of our moral judgments, especially the ones we make intuitively. Once those inner workings are revealed we may have less confidence in some of our judgments and the ethical theories that are (explicitly or implicitly) based on them.

In Section I, I will describe our brains as *dual-process* systems and introduce the *camera analogy*, the organizing metaphor of this paper, and a central idea in my book.² In Section II, I will describe and present evidence for the dual-process theory of moral judgment. In Section III, I will describe related experimental research examining influences on our moral intuitions. In Sections IV and V, I'll describe two routes by which cognitive science can have implications for ethics, and with no illicit is/ought border crossings. I call these routes *direct* and *indirect*. In Section VI, I'll explain why a deeper understanding of moral psychology favors certain forms of consequentialism. I will then close with a few words about the bright future of ethics as an interdisciplinary enterprise.

¹ Sinnott-Armstrong, Walter, ed. 2007. *Moral Psychology*. Vol. 1-3. Cambridge, MA: MIT Press. Doris, J. M., ed. 2010. *The Moral Psychology Handbook*. Oxford: Oxford University Press.

² Greene, J. 2013. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin Press.

I. The dual-process brain

I own a digital SLR camera that, like many others, operates in two complementary modes. First, it has a variety of point-and-shoot automatic settings that are optimized for typical photographic situations (“portrait,” “action,” “landscape” etc.). I use these most of the time. Occasionally, I get ambitious and put the camera in manual mode, which allows me to adjust all of the camera’s settings by hand.

This dual-mode design exemplifies an elegant solution to a ubiquitous design problem, namely the trade-off between *efficiency* and *flexibility*. The automatic settings are highly efficient, but not very flexible, and the reverse is true of the manual mode. Put them together, however, and you get the best of both worlds, provided that you know when to use each mode.

The human brain has the same general design.³ First, we humans have a variety of automatic settings—reflexes and intuitions that guide our behavior, many of which are emotional. We may be conscious of such emotional responses, but we are generally not conscious of the processes that trigger them. We rely on our automatic settings most of the time,⁴ and they generally serve us well.

Our brains also have a manual mode. It is a general-purpose reasoning system, specialized for enabling behaviors that serve long(er)-term goals, i.e. goals that are not automatically activated by current environmental stimuli or endogenous somatic

³ See, e.g., Kahneman, D. 2003. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol* 58 (9):697-720.

⁴ See, e.g., Bargh, J. A., and T. L. Chartrand. 1999. The unbearable automaticity of being. *American Psychologist* 54:462-479.

states.⁵ The operations of this system are typically conscious, experienced as voluntary, and often experienced as effortful. Our manual mode allows us to formulate behavioral plans based on detailed and explicit knowledge of the situations we face, along with explicit general knowledge about the world and how it works. Manual mode allows us to guide our behavior using explicit rules and to think explicitly about how the world works. In short, manual mode thinking is the kind of thinking that we think of as “thinking.”⁶

The way our brains handle the efficiency / flexibility tradeoff is nicely illustrated by our navigation of the familiar tension between *now* and *later*. We have automatic settings that urge us to consume or acquire valuable resources, such as calorie-rich food and money, whenever they are available. This is very efficient because it is generally adaptive to acquire valuable resources. At the same time, we humans have the capacity

⁵ See, e.g., Miller, E. K., and J. D. Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24:167-202.

⁶ One might ask what, exactly, is “dual” in dual-process theories. Is it types of processing? Types of cognitive systems? Different brain regions? Different kinds of outputs? The answer is “all of the above,” but the core difference, in my view, concerns the type of processing. As Turing taught us, dual-process psychology can be implemented or simulated on a computer using a single processing system occupying a single physical location and using one set of computational principles (at low levels, at least). But, as it happens, that’s not how it generally works in the brain. Instead, distinct neural systems typically engage in distinct types of processing in distinct locations. Likewise, cognitive outputs typically reflect the kinds of processing that that produced them. As a result, a brain injury can alter behavior because it causes damage to a particular location, which houses a particular cognitive system, which processes information in a particular way, and which therefore tends to push behavior in a particular direction. Of course, not all dual-process dissociations are so clean, but sometimes they are. See Gazzaniga, M et al. 2008. *Cognitive Neuroscience: The Biology of the Mind*, 3rd Ed. New York: Norton.

to think explicitly about whether our long-term goals are served by immediate consumption/acquisition, along with the capacity to regulate our behavior accordingly. In other words, we can delay gratification, choosing, for example, a slimmer waistline in three months over chocolate cake right now. This is a dual-process phenomenon. If, for example, our manual mode capacity is occupied by a distracting memory task, our automatic settings gain the advantage, and we are more likely to choose chocolate cake over fruit salad.⁷

Recent brain imaging studies reveal the underlying neural dynamics.⁸ Brain regions such as the ventral striatum and the ventromedial prefrontal cortex (VMPFC) produce the automatic response favoring *now* and enable this response to influence behavior. Other brain regions, most notably the dorsolateral prefrontal cortex (DLPFC), enable the controlled response that sometimes favors *later*, depending on the situational details. We see the dual-process brain at work in other domains, for example, in the cognitive control of negative reactions to members of racial out-groups and sad scenes.⁹ In these cases among others, the amygdala, an ancient mammalian brain structure with direct connections to the VMPFC, plays a critical role in automatic responding, while the DLPFC plays a central role in coordinating manual mode thinking and responding.

⁷ Shiv, B, and A Fedorikhin. 2002. Spontaneous versus controlled influences of stimulus-based affect on choice behavior. *Organizational Behavior and Human Decision Processes* 87 (2):342-370.

⁸ McClure et al. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science* 306 (5695):503-7.

⁹ Cunningham et al. 2004. Separable neural components in the processing of black and white faces. *Psychological Science* 15 (12):806-13. Ochsner et al. 2002. Rethinking Feelings: An fMRI Study of the Cognitive Regulation of Emotion. *Journal of Cognitive Neuroscience* 14 (8):1215-1229.

Before moving on, it's worth highlighting three ways in which the camera analogy may mislead. First, while a camera must be in one mode or another, the waking human brain's automatic settings are always on. Second, a camera's dual modes can function independently of each other, but in animals there is an asymmetrical dependence. One can have automatic settings without a manual mode, as in most animals, but not the reverse. Finally, automatic settings need not be "innate" or "hard-wired." They can be acquired or modified through cultural learning (as in prejudicial responses to racial out-groups) and through individual experiences (as in classical conditioning).

II. Dual-Process morality

The dual-process theory of moral judgment

According to the dual-process theory of moral judgment¹⁰ (henceforth "dual-process theory"), moral psychology looks much like the rest of judgment and decision-making. Moral judgment is influenced by both automatic emotional responses (automatic settings) and controlled, conscious reasoning (manual mode). Moreover, these

¹⁰ Greene et al. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293 (5537):2105-8. Greene et al. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44 (2):389-400. Greene, et al. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107 (3):1144-54. Greene, J.D. 2007. The secret joke of Kant's soul. In *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*, edited by W. Sinnott-Armstrong. Cambridge, MA: MIT Press.

processes are enabled by the usual cast of neural characters in their characteristic roles. These tenets of the dual-process theory should be relatively unsurprising and uncontroversial. The more interesting, and correspondingly controversial, tenet of the dual-process theory is this:

The Central Tension Principle: Characteristically deontological judgments are preferentially supported by automatic emotional responses, while characteristically consequentialist judgments are preferentially supported by conscious reasoning and allied processes of cognitive control.

The name of this principle reflects a more general idea, which is that the central tension in ethics between deontology and consequentialism is a manifestation of the central tension in cognitive design between efficiency and flexibility.

Some terminological clarification is in order. My use of “deontological” and “consequentialist” is not entirely consistent with standard philosophical usage, necessitating my use of the technical qualifier “characteristically.” This has been a source of some confusion.¹¹ I define “characteristically deontological” judgments as ones that are naturally justified in deontological terms (in terms of rights and duties, etc.) and that are more difficult to justify in consequentialist terms, such as judgments against killing one person to save five others. I define “characteristically

¹¹ Several authors disregard the “characteristically” qualifier (Greene, 2007, *Kant’s Soul, op. cit.*), insisting that the judgments I call “consequentialist” must (a) reflect a full commitment to consequentialism and (b) be inconsistent with alternative normative theories. Kahane, G., and N. Shackel. 2010. Methodological issues in the neuroscience of moral judgment. *Mind and Language* 25 (5):561-582. Kamm, F.M. 2009. Neuroscience and moral reasoning: A note on recent research. *Philosophy and Public Affairs* 37 (4):330-345.

consequentialist” judgments as ones that are naturally justified in consequentialist terms (i.e. by impartial cost-benefit reasoning) and that are more difficult to justify in deontological terms because they conflict with our sense of people’s rights, duties, etc. Approving of killing one to save five is a characteristically consequentialist judgment. (Note that I will sometimes drop the qualifier “characteristically” in the interest of brevity, using instead an asterisk to indicate that the qualifier still applies.)

Two further points about these terms: First, they imply nothing *a priori* about the psychology behind a given judgment, and therefore nothing about the judge’s *reasons*. Such psychological facts are to be ascertained empirically. Second, making “characteristically consequentialist” and “characteristically deontological” judgments requires no explicit or implicit commitment to consequentialist or deontological theories. For example, card-carrying deontologists, and people who carry no theoretical cards, can make characteristically consequentialist judgments. *Mutatis mutandis* for consequentialists.

My non-standard use of these terms will strike some as perverse, but there is a method to this madness. The dual-process theory aims to characterize the moral-philosophical tendencies of distinct cognitive systems.¹² This endeavor is complicated by the fact that these systems do not correspond precisely to distinct ethical schools as designated by contemporary ethicists. This is largely because sophisticated consequentialist and deontological theories are designed to capture the outputs of multiple cognitive systems. (Cf. “climbing the mountain from different sides.”¹³) Nevertheless, if I’m right, the psychological *essence* of deontology lies with the

¹² Here I’m referring to the Central Tension Principle and the more general idea that different moral philosophies may be supported to different extents by the outputs of different cognitive systems.

¹³ Parfit, D. 2011. *On what Matters*. New York: Oxford University Press.

automatic settings and the psychological *essence* of consequentialism lies with manual mode. To articulate and test this idea we need to modify both our vocabulary and corresponding concepts. To get sophisticated about moral psychology we must *temporarily* get unsophisticated about philosophical terminology. Please bear with me.

Evidence

The dual-process theory was inspired by the Trolley Problem,¹⁴ with which I assume most readers are familiar. For present purposes, the two key dilemmas are the *switch* (a.k.a. *bystander*) and *footbridge* cases. Very briefly for the uninitiated: In the *switch* case, one can hit a switch that will turn a runaway trolley away from five people and onto one. In the *footbridge* case one can push one person off a footbridge and into the path of a runaway trolley, saving five further down the track. People tend to give characteristically consequentialist responses to the *switch* case (“Yes, it’s permissible to hit the switch to save more lives,”) and characteristically deontological responses to the *footbridge* case (“No, it’s impermissible to push to save more lives,”).

Our first neuroimaging experiment¹⁵ compared “personal” moral dilemmas such as the *footbridge* dilemma to “impersonal” moral dilemmas such as the *switch* dilemma. Our hypothesis, based on the dual-process theory, was that the former would preferentially engage brain regions associated with emotion while the latter would preferentially engage brain regions associated with controlled cognition. This is indeed

¹⁴ Foot, P. 1967. The problem of abortion and the doctrine of double effect. *Oxford review* 5:5-15. Thomson, J.J. 1985. The Trolley Problem. *Yale Law Journal* 94 (6):1395-1415.

¹⁵ Greene et al., 2001, *Science*, *op. cit.*

what we found. The “personal” dilemmas elicited relatively greater activity in a large swath of the medial prefrontal cortex, including parts of the VMPFC, and a subsequent analysis with more data showed the same effect in the amygdala.¹⁶ Similar effects were observed in other brain regions previously associated with emotion. In contrast, the “impersonal” dilemmas elicited relatively greater activity in the DLPFC and allied brain regions. Such results were specifically predicted by the dual-process theory and thus lend it some support.¹⁷

¹⁶ Greene et al., 2004, *Neuron*, *op. cit.*

¹⁷ There has been much confusion about what this experiment shows and what it was intended to show. It was designed to test the dual-process theory, though not the Central Tension Principle specifically. Some critics have conflated the dual-process theory with the theory implicit in the 2001 personal/impersonal distinction. (See Greene, J.D. 2009. Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology* 45 (3):581-584.) Others have assumed that the 2001 experiment was intended to test the specific version of the personal/impersonal distinction that it used, and then faulted it for failing to do so adequately. (See Kamm (2009, *PPA*, *op. cit.*) and Berker, S. 2009. The normative insignificance of neuroscience. *Philosophy and Public Affairs* 37 (4):293-329. See also Mikhail, J. 2011. Emotion, neuroscience, and law: A comment on Darwin and Greene. *Emotion Review* 3 (3):293-295.) The testing and refining of the personal/impersonal distinction was accomplished by later experiments, discussed in Section III. (Greene, J.D. et al. 2009. Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition* 111 (3):364-71.)

The looseness of the personal/impersonal comparison in the 2001 experiment did not prevent it from supporting the dual process theory, but it did impose limitations. Subsequent studies have overcome these limitations by (a) focusing primarily or exclusively on “personal” dilemmas similar to *footbridge*, and (b) examining factors that cause or correlate with people’s *judgments*, rather than effects of different stimuli (personal vs. impersonal). These studies typically use one or more “impersonal”

Now, over a decade later, the dual-process theory is supported by a large and diverse body of evidence. We'll begin with evidence concerning the role of emotional intuition in characteristically deontological judgment—or, alternatively, characteristically non-consequentialist judgment. (Unless otherwise specified, the results below are based on (or driven by) responses to one or more “high-conflict”¹⁸ personal dilemmas.)

Patients with frontotemporal dementia, which typically involves emotional blunting, are about three times as likely as control subjects to give consequentialist* responses to the *footbridge* case.¹⁹

Patients with damage to the VMPFC are about five times as likely as others to give consequentialist* responses.²⁰ A research team in Italy produced similar

dilemmas, but only to establish the specificity of the main result, generated using one or more “personal” dilemmas.

Some have wondered why the 2001 experiment asked subjects to classify actions as “appropriate” or “inappropriate.” This was done to keep the response prompt constant across moral and non-moral trials. Subjects were explicitly instructed to judge based on moral considerations where applicable.

¹⁸ Koenigs, M., L. et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446 (7138):908-11.

¹⁹ Mendez, M. F. et al. 2005. An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology* 18 (4):193-7

²⁰ Koenigs et al., 2007, *Nature*, *op. cit.*. Kahane & Shackel criticized the Koenigs et al. study for employing dilemmas in which the utilitarian judgment is not, according to Kahane et al. and other philosophers surveyed, truly (i.e. uniquely) utilitarian. (Kahane, G., and N. Shackel. 2008. Do abnormal responses

results.²¹ A follow-up study shows that consequentialist* judgments are associated with absent skin-conductance responses (SCRs, which indicate affective arousal) in VMPFC patients and reduced SCRs in healthy subjects.²²

VMPFC patients give more consequentialist* responses to dilemmas in which familial duties are pitted against consequentialist* considerations (e.g., turning a trolley onto one's sister to save others).²³

People who exhibit greater physiological reactivity (constriction of peripheral blood vessels) to performing a stressful arithmetic task give less consequentialist* responses.²⁴

Low-anxiety psychopaths (known for their social-emotional deficits) are more likely than healthy people to give consequentialist* responses.²⁵

show utilitarian bias? *Nature* 452 (7185):E5; author reply E5-6) Koenigs et al. replied by analyzing the data from only those dilemmas approved by Kahane et al. and produced the same results.

²¹ Ciaramelli, E. et al. 2007. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience* 2 (2):84-92).

²² Moretto, G. et al. 2010. A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *J Cogn Neurosci* 22 (8):1888-99.

²³ Thomas, B. C. et al. 2011. Harming kin to save strangers: further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *J Cogn Neurosci* 23 (9):2186-96.

²⁴ Cushman, F.A. et al. in 2012. Simulating murder: the aversion to harmful action. *Emotion* 12(1):2-7.

²⁵ Koenigs, M. et al. 2012. Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience* 7(6):708-714

People who are more empathetic, or induced to be more empathetic, give more deontological* responses.²⁶

Individuals high in psychopathy exhibit lower amygdala responses and higher DLPFC responses to “personal” dilemmas.²⁷

Thinking about death reduces consequentialist* judgment.²⁸

Individuals with deficits in emotional awareness (due to alexithymia) make more consequentialist* judgments.²⁹

Amygdala activity correlates positively with ratings of negative emotion in response to *footbridge*-like cases and correlates negatively with consequentialist* judgments.³⁰

²⁶ Conway, P. & B. Garwronski. 2013. Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology* 104(2): 216-235. This very clever study used the “process dissociation” method to separate utilitarian* and deontological* components. This required a modified set of dilemmas.

²⁷ Glenn, A.L. et al. 2009. The neural correlates of moral decision-making in psychopathy. *Mol Psychiatry* 14 (1):5-6. Glenn, A.L. et al. 2009. Increased DLPFC activity during moral decision-making in psychopathy. *Mol Psychiatry* 14:909-911.

²⁸ Trémolière, B. et al. 2012. Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition* 124(3): 379-384.

²⁹ Koven, N. 2011. Specificity of meta-emotion effects on moral decision-making. *Emotion* 11 (5):1255-1261.

³⁰ A.S. Shenhav & J.D. Greene, under review.

Citalopram—an SSRI that, in the short-term, increases emotional reactivity through its influence on the amygdala and VMPFC, among other regions—reduces consequentialist* responses.³¹

Inducing mirth (the positive emotion associated with humor, here thought to counter-act negative emotional responses) increases consequentialist* responses.³²

Visual thinking is generally more emotionally evocative than verbal thinking, and individuals with more visual cognitive styles tend to give less consequentialist* responses. Likewise, interfering with visual thinking (but not verbal thinking) makes judgments more consequentialist*.³³

Other evidence links characteristically consequentialist judgment to controlled cognition.

³¹ Crockett, M.J. et al. 2010. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc Natl Acad Sci U S A* 107 (40):17433-8.

³² Valdesolo, P., and D. DeSteno. 2006. Manipulations of emotional context shape moral judgment. *Psychological Science*, 17 (6):476-7. Strohminger, N. et al. 2011. Divergent effects of different positive emotions on moral judgment. *Cognition* 119 (2):295-300.

³³ Amit E. & Greene J.D. 2012. You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science* 23(8): 861-868.

Consequentialist* judgment is associated with increased DLPFC activity within individuals³⁴ and across individuals.³⁵

Performing a distracting secondary task (i.e. being under cognitive load) reduces consequentialist* responses³⁶ or slows consequentialist* responses, while having no effect on deontological* responses.³⁷

Removing time pressure and encouraging deliberation increases consequentialist* responses.³⁸

The experience of successfully solving tricky math problems (ones that require one to question one's intuitions) makes people subsequently more likely to give consequentialist* responses. Individuals who solve more tricky math problems tend to give more consequentialist* responses to a higher-stakes version of the *footbridge* case, independent of whether they solved math problems first.³⁹

³⁴ Greene et al., 2004, *Neuron*, *op. cit.*.

³⁵ Cushman et al. in prep.

³⁶ Trémolière et al., 2012, *Cognition*, *op. cit.* Conway and Gawronski, 2013, *JPSP*, *op. cit.*

³⁷ Greene et al, 2008, *Cognition*, *op. cit.*

³⁸ Suter, R.S., and R. Hertwig. 2011. Time and moral judgment. *Cognition* 119 (3):454-458.

³⁹ Paxton, J.M., L. Ungar, and J.D. Greene. 2012. Reflection and reasoning in moral judgment. *Cognitive Science* 36(1):163-177.

Individuals who generally favor effortful thinking over intuitive thinking are more likely to give consequentialist* responses.⁴⁰

People are invariably conscious of the consequentialist rationale for making consequentialist* judgments, but lack conscious access to the causes of their deontological* patterns of judgment (approving of some consequentialist* trade-offs but not others). People often question or rewrite dilemmas' assumptions so as to produce a coherent consequentialist justification for disapproval.⁴¹

That's a lot of evidence. All of it comes from trolleyology, but that's no reason to dismiss it. We philosophers have puzzled over trolley dilemmas for decades because they capture a central—if not *the* central—tension in normative ethics, and the myriad scientific results these dilemmas have generated implies that they tap something deep—revealing the hidden tectonics of the moral mind.⁴² That said, there is evidence for the dual-process theory beyond trolleyology:

⁴⁰ Bartels, D.M. 2008. Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition* 108:381-417. See also Conway and Gawronski (2013, *op. cit.*) and Moore, A.B. et al. 2008. Who shalt not kill?: Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science* 19 (6):549-557.

⁴¹ Cushman, F. et al. 2006. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science* 17 (12):1082-9. Hauser, M.D. et al., 2007. A dissociation between moral judgments and justifications. *Mind and Language* 22 (1):1-21.

⁴² Cushman, F, and J. D. Greene. 2012. Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience* 7(3):269-279.

Negative emotional responses predict characteristically non-consequentialist disapproval of harmless moral transgressions, including disapproval that is characteristically deontological (e.g., disapproval of breaking a promise without negative consequences).⁴³

Several experiments indicate that consequentialist considerations play a minimal role in people's judgments about punishment, though people readily appeal to consequentialist considerations when asked to explicitly justify punishments.⁴⁴ Instead, punishment judgments follow a pattern consistent with retributivism, a distinctively non-consequentialist, and specifically Kantian, philosophy of punishment. Punishment appears to be driven primarily by automatic negative emotional responses, and people who are more punitive tend to rely less on controlled cognition.⁴⁵

People's judgments and decisions about helping behavior follow a non-consequentialist* pattern and tend to be driven by automatic processes.⁴⁶

⁴³ Haidt, J. et al. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65 (4):613-28. Wheatley, T., and J. Haidt. 2005. Hypnotic disgust makes moral judgments more severe. *Psychological Science* 16 (10):780-4.

⁴⁴ See Carlsmith, K.M. et al. 2002. Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology* 83 (2):284-99. See further evidence and discussion in Greene 2007 and 2013 (*Moral Tribes, op. cit.*).

⁴⁵ Sargent, M.J. 2004. Less thought, more punishment: Need for cognition predicts support for punitive responses to crime. *Personality and Social Psychology Bulletin* 30:1485-1493.

⁴⁶ Small, Deborah A., and George Loewenstein. 2003. Helping a Victim or Helping the Victim. *Journal of Risk and Uncertainty* 26 (1):5-16. Kogut, T, and I Ritov. 2005. The singularity effect of identified victims in

Public health professionals (for whom the patient is the society at large) make more consequentialist* judgments than doctors and ordinary people in response to medical dilemmas.⁴⁷

In sum, the dual-process theory is supported by an exceptionally wide range of convergent and complementary evidence from many independent researchers.⁴⁸ No one piece of evidence is definitive, and each piece, taken in isolation, is open to alternative interpretations. But no theory of which I am aware comes anywhere close to explaining this body of evidence better than the dual-process theory, which explicitly

separate and joint evaluations. *Organizational Behavior and Human Decision Processes* 97 (2):106-116. Slovic, P. 2007. "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making* 2 (2):79-95.

⁴⁷ K.J. Ransohoff, D. Wikler, and J.D. Greene, in prep.

⁴⁸ Unfortunately, some philosophers have been given a different impression by Berker's (2009, *PPA*, *op. cit.*) discussion of the evidence supporting the dual-process theory, primarily in Sections II-III of his article. I have documented Berker's various errors, misleading statements, and misleading omissions in a set of notes that I prepared for a 2010 meeting. These notes are available on my webpage (<http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf>) or by request. Berker's errors include multiple false statements about statistical analyses and experimental confounds. More generally, he presents a misleading picture of the evidence supporting the dual-process theory by ignoring and/or erroneously dismissing most of the then-available evidence. For more details concerning these and other problems with Berker's article, readers are encouraged to consult the aforementioned online document.

predicted most of these results in advance and predicts (in the timeless sense) all of them.⁴⁹

Counter-evidence?

Is there any positive evidence against the dual-process theory? Kahane et al.⁵⁰ hypothesize that consequentialist judgments have so far been associated with controlled cognition simply because research has, so far, focused on cases in which consequentialist* judgments happen to be supported by controlled cognition. If you look elsewhere, they say, you'll find characteristically deontological judgments preferentially supported by controlled cognition.

I once speculated about the possibility of finding such cases:

This [dual-process theory] also makes sense of certain deontological anomalies... Consider, for example, Kant's infamous claim that it would be wrong to lie to a would-be murderer in order to protect a friend who has taken refuge in one's home... Kant sticks to his theory and rejects the intuitive response. (He "bites the

⁴⁹ It's worth noting that the experiments described above did not use philosophers as subjects, and they have focused on responses to specific cases rather than abstract principles. For this reason, one might wonder whether these results have any bearing on the psychology behind familiar philosophical theories. As I will explain in Section IV, I believe that the psychology of philosophers is, in some important ways, different from that of ordinary moralists, but that studies of ordinary people's judgments in response to specific cases can also illuminate the psychology behind familiar philosophical theories.

⁵⁰ Kahane, G., et al. 2011. The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*.

bullet,” as philosophers say.) But what is interesting about this bit of Kantian ethics is that it’s something of an embarrassment to contemporary Kantians, who are very keen to explain how Kant somehow misapplied his own theory in this case... If you want to know which bits of Kant contemporary Kantians will reject, follow the emotions.⁵¹

This suggests a more general formula for generating counter-intuitive deontological* judgments: Find a moral rule that reasonable adults rarely hold as absolute (“Don’t lie”) and then pit that rule against a serious harm that we are intuitively inclined to avoid and that we have consequentialist reason to avoid. This is precisely the strategy adopted by Kahane et al., who examine cases of “white lies” (Should you devastate your friend by telling him what his mean-spirited uncle really thought of him?), and “imprudent autonomy” (Should you buy an alcoholic beggar booze instead of food because that’s what he really wants?), etc. Kahane et al. conducted an fMRI experiment using dilemmas such as these, along with standard *footbridge*-like dilemmas. They predicted that deontological* judgments in response to their new dilemmas would show signs of preferentially engaging controlled cognition, providing evidence that these deontological* judgments are counter-intuitive, contrary to the dual-process theory and the Central Tension Principle, more specifically.

My collaborators and I did not find Kahane et al.’s fMRI and reaction time data convincing,⁵² but we wondered whether they might be right about some of their cases.

⁵¹ Greene, 2007, *Secret joke*, *op. cit.*, pg. 66.

⁵² They found no effect of “counter-intuitive” deontological* judgment in the DLPFC, the signature of controlled cognition, previously observed in association with consequentialist* judgment (Greene et al., 2004, *Neuron*, *op. cit.* Cushman et al. in prep). For the “white lie”-like dilemmas, the deontological*

With this in mind, Joe Paxton, Tommaso Bruni, and I conducted a more definitive behavioral test using one of our standard dilemmas (also used by Kahane et al.) and one of Kahane et al.'s "white lie" dilemmas.⁵³ We used the "tricky math problems" technique, mentioned above.⁵⁴ As expected, subjects who distrusted their intuitions on the tricky math problems were more likely to give consequentialist* responses to our standard case. And, most critically, subjects who distrusted their faulty intuitions on the math problems also gave more consequentialist* responses to Kahane et al.'s "white lie" dilemma—exactly what the original dual-process theory predicts, and exactly the opposite of what Kahane et al.'s theory predicts. Given that this dilemma was engineered to be a counter-example to the Central Tension Principle and the dual-process theory, it's hard to imagine this theory's receiving more striking confirmation. That said, I've no doubt that, somewhere, an exception to the dual-process theory's predicted pattern will be found. The point, however, is not that the dual-process theory predicts every case perfectly, but rather that it captures the general shape of philosophical moral psychology.

At this point, having read a short summary of the evidence supporting the dual-process theory, you may or may not be inclined to accept it as true, or approximately

judgments took slightly longer than the consequentialist* judgments and were rated as more difficult, consistent with Kahane et al.'s conclusions. However, these "counter-intuitive" deontological* judgments were actually *faster* than the intuitive deontological* judgments in the *footbridge*-like cases and were rated as no more difficult than them. This is not consistent with their theory.

⁵³ Paxton, J.M., T. Bruni, and J.D. Greene. 2013. Are "counter-intuitive" deontological judgments really counter-intuitive?: An empirical reply to Kahane et al. (2012). *Social, Cognitive, and Affective Neuroscience*.

⁵⁴ Paxton et al., 2011, *Cog Sci*, *op. cit.*

true. Nevertheless, for the remainder of this article, I will assume that it's correct, in order to explore its implications.

III. What pushes our moral buttons?

We have automatic emotional responses that support characteristically deontological judgments. But what triggers these emotional responses? What, for example, is it about pushing the man off the footbridge that makes us feel that it's wrong? Experiments are answering this question, among others.

It seems that there are two key factors that explain why most people say "yes" to the *switch* case and "no" to the *footbridge* case (henceforth, the *switch-footbridge* effect). The first is whether the victim is harmed as a means vs. side-effect,⁵⁵ a factor that has long been cited by ethicists as relevant here and elsewhere.⁵⁶ The second has to do with the "personalness" of the harm.⁵⁷ I and my collaborators have conducted a set of experiments⁵⁸ showing that these two factors interact to produce the *switch-footbridge*

⁵⁵ 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*. doctoral dissertation, Cornell University. Cushman et al., 2006, *Psych Sci*, *op. cit.*

⁵⁶ Foot, 1967, *Oxford Review*, *op. cit.*

⁵⁷ Royzman, Edward B., and J. Baron. 2002. The preference for indirect harm. *Social Justice Research* 15:165-84. Cushman et al. 2006, *Psych Sci*, *op. cit.* Waldmann, M. R., and J. H. Dieterich. 2007. Throwing a bomb on a person versus throwing a person on a bomb: intervention myopia in moral intuitions. *Psychological Science* 18 (3):247-5. Moore et al., 2008, *Psych Sci*, *op. cit.*

⁵⁸ Greene et al., 2009, *Cognition*, *op. cit.*

effect.⁵⁹ (I use term “interact” in the technical statistical sense, meaning that the influence of one factor is influenced by the presence/absence of another factor, as in an interaction between medications.) Here I will focus on the personalness factor because it is the most normatively relevant. We’ll return to the means/side-effect factor in Section VI.

The action in *footbridge* involves the application of *personal force*. That is, the agent directly impacts the victim with the force of his/her muscles. The effect of personal force is revealed by examining four footbridge variations. In the *footbridge pole* version, the agent pushes the victim with a pole. In the *footbridge switch* version, the agent drops the victim onto the tracks through a switch-operated trap door, while standing near the victim on the footbridge. In the *remote footbridge* version, the switch is located elsewhere, away from the victim. We asked separate groups of subjects to judge whether the action proposed is morally acceptable. The results are as follows:

<i>standard footbridge:</i>	31% Yes
<i>footbridge pole:</i>	33% Yes
<i>footbridge switch:</i>	59% Yes
<i>remote footbridge:</i>	63% Yes

The results for the first two cases do not differ significantly. That is, there is no evidence for an effect of physical contact *per se*. Likewise, comparing the last two cases, there is no evidence for an effect of spatial distance *per se*. However, the difference

⁵⁹ Other factors contribute to the *switch-footbridge* effect but these two factors account for much, if not most, of it.

between the first two cases and the last two, the difference between pushing and hitting a switch, is highly significant, doubling the number of “Yes” responses. (At this point, you may already feel a normative tingle in your extremities. That’s normal.)

I’ll briefly mention a parallel line of research,⁶⁰ led by Jay Musen, related to Peter Singer’s famous drowning child dilemma.⁶¹ It seems monstrously wrong to allow a child to drown in a shallow pond because one is concerned about muddying up one’s clothes. And yet it seems to most people morally acceptable, if not morally ideal, to spend one’s disposable income on luxuries for oneself, even if that money could be used to save the lives of desperately needy people. Why do we say both of these things? Evidence from Musen’s experiments suggests that our judgments are highly sensitive to *mere spatial distance*. This effect is nicely illustrated by contrasting two cases, which we’ll call *near* and *far*.⁶²

In the *near* case, you are vacationing in a developing country that has been hit by a devastating typhoon. You are safe and well-supplied in your temporary home on a coastal mountainside, but many people along the coast are dying. A relief effort led by an international organization is underway, and you can help by donating money. A relatively modest donation can save a life, and no money will go to waste. Are you morally obliged to donate?

In the *far* case it’s your friend, rather than you, who is there. Your friend uses a smartphone to capture audio and video from the scene, transmitting these to you live over the internet, while you sit comfortably at home. You can also donate to the relief

⁶⁰ Musen J.D. & J.D. Greene, in prep.

⁶¹ Singer, P. 1972. Famine, Affluence and Morality. *Philosophy and Public Affairs* 1:229-243.

⁶² These cases borrow from Unger, P.K. 1996. *Living high and letting die: our illusion of innocence*. New York, NY: Oxford University Press.

effort over the internet. In other words, you know everything your friend knows, see and hear everything your friend sees and hears, and you are in just as good a position to help. The only difference⁶³ is that your friend is physically near while you are physically far. It seems that this difference dramatically affects people's judgments.⁶⁴ (Tingle?⁶⁵)

IV. Moral implications: The direct route

It's time to explain and vindicate these *oughty* tinges, induced by purely *is-ish* experimental findings. I will do this with no is/ought sleight of hand.

Such experiments identify factors to which our moral judgments are sensitive. This information may be combined with independent normative assumptions⁶⁶ concerning the kinds of things to which our judgments ought to be sensitive. This combination can lead us to new, substantive moral conclusions. In other words,

⁶³ Is this the *only* difference? Strictly speaking, no. For example, in *far*, but not in *near*, you get your information over the internet. What matters for present purposes is whether there are other differences that are plausibly regarded as morally relevant. More on this shortly.

⁶⁴ Further details are withheld so as not to preempt the scientific publication of these results.

⁶⁵ For further normative tinges see Paharia, N., et al. 2009. Dirty work, clean hands: the moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes* 109:134-141. See also Nichols, S., and J. Knobe. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous* 41:663-685.

⁶⁶ Such assumptions are, ideally, shared by both philosophers and non-philosophers, but this may vary from case to case.

scientific information can allow us to trade in difficult “ought” questions for easier “ought” questions, and thus advance ethics.

For example, suppose we want to know whether capital juries make good judgments⁶⁷—a normative question that is, at the outset, open. Next we get a bit of disturbing scientific information: Capital juries are sensitive to race. May we now say that capital juries, at least sometimes, make bad decisions? Not yet. For that we need an additional, non-scientific normative assumption, stating that the judgments of capital juries ought not be sensitive to race. However, if we’re willing to make this assumption, we may now draw a new and substantive moral conclusion: Capital juries, at least sometimes, make bad decisions.

As this example illustrates, we can reach interesting normative conclusions by combining interesting scientific facts with uninteresting normative assumptions. However obvious this may seem, some mistakenly assume that empirically based normative arguments are empty or question-begging if they rely on non-empirical normative assumptions.⁶⁸ The above example suggests a more general principle: An empirically-driven normative argument is non-question-begging if the normative assumptions it requires are less interesting (i.e. less controversial) than its normative conclusion. I am not claiming one can derive a moral “ought” from nothing but a scientific “is.” Rather, my point is that moral psychology matters for ethics, that it is

⁶⁷ I am here and elsewhere in this article remaining meta-ethically agnostic and therefore leaving it open as to what counts as a “good” judgment. “Good” judgments may be true, or may simply be good according to some other set of evaluative criteria..

⁶⁸ For example, Berker (2009, *PPA*, *op. cit.*) claims that non-empirical normative assumptions “do all the work” (pg. 326) in the above arguments, rendering the science “normatively insignificant.” That is not true. Normative assumptions do some work, but empirical evidence does essential work as well.

“normatively significant.” Moral psychology matters, not because it can generate interesting normative conclusions all by itself, but because it can play an essential role in generating interesting normative conclusions.

A natural objection to this example is that the work done by the science, while not insignificant, is *normatively* insignificant. The science does not challenge anyone’s values. Instead, it simply alerts us to an application of the values we already have.⁶⁹ With this objection in mind, let’s consider a second example, the case of consensual adult incest. Ought we condemn all incestuous behavior? This is a difficult “ought” question, at least for some people, and any answer one gives will be controversial. Suppose we learn (as we likely have already) that the inclination to condemn incest of all kinds is based on an emotional response whose function is to avoid producing offspring with genetic diseases.⁷⁰ As before, we need to answer a second, easier “ought” question before reaching a new moral destination: Ought we rely on such emotional responses in cases in which there is no special concern about genetic diseases? For example, we might imagine a brother-sister pair, Joe and Jane, who were separated in early childhood and who later met and fell in love.⁷¹ We might imagine that they become aware of their biological relation, but nonetheless choose to remain together, taking major precautions (e.g. vasectomy) to ensure that they are no more likely than typical couples to produce children with genetic diseases. With Joe and Jane in mind, we might make the following normative assumption: If our inclination to condemn Joe

⁶⁹ Thanks to T.M. Scanlon on this point.

⁷⁰ Lieberman, D, J Tooby, and L. Cosmides. 2007. The architecture of human kin-detection. *Nature* 445:727-731.

⁷¹ Cf. Weathers, H. 2008. How we fell in love, by the brother and sister who grew up apart and met in their 20s. *MailOnline.com*, Feb 17.

and Jane's behavior depends on an emotional response that makes their behavior seem wrong, and this emotional response evolved to prevent birth defects, and birth defects are not a special issue in their case, and we have no additional reason to condemn their behavior,⁷² then we ought not condemn their behavior. Having made this assumption, and having learned something from science, we may now conclude that we ought not condemn all incestuous behavior—an interesting normative conclusion. This example—a classic *debunking* explanation—is notable because it genuinely challenges some people's moral values. Indeed, such arguments can change people's minds, if you give them enough time to think.⁷³

We're now ready to make good on our tingles. Here, our initial, difficult "ought" question in this: Do people make good moral judgments when confronted with moral dilemmas? Next, science tells us that people's judgments are, at least sometimes, sensitive to things like mere personal force and mere spatial proximity. Next we face an easier "ought" question: Ought people's moral judgments be sensitive to such things? We all answer, "no," of course.

(Perhaps you're tempted to withhold your negation because personal force and spatial proximity may be reliable correlates of things that matter morally, such as the presence of harmful violence and binding social relationships. You should resist this temptation, bearing in mind the word "mere." The question is this: Were a friend to call

⁷² For example, we do not believe that their behavior will damage them or others psychologically, promote immoral behavior more generally, etc. One might object that this "no additional reason" clause is what's doing "all the work" in this argument. I disagree. This clause simply closes off alternative escape routes, forcing a confrontation between the emotional response that is driving the judgment and the empirically-based debunking argument that challenges it.

⁷³ Paxton et al., 2011, *Cog Sci*, *op. cit.*

you from a footbridge seeking moral advice, would you say, “Well, that depends... Will you be pushing or using a switch?” If questions such as this, including similar questions concerning mere spatial distance, are not on your list of relevant moral questions, then you, too, should say “no.”)

And thus we’ve earned an interesting normative conclusion: People, at least sometimes, do not make good moral judgments in response to moral dilemmas, for they are inappropriately sensitive to mere personal force and mere spatial distance. And thus we’ve shown that interesting scientific facts about moral psychology can, when combined with relatively uninteresting normative assumptions, lead us to relatively interesting normative conclusions. That’s progress, powered by science.

Limited progress, however. We’ve seen how one can get a substantively new “ought” by combining an old “ought” with a new “is,” but, still, one might hope for more. The above argument tells us that *something* is wrong with *some people’s* judgments, but it doesn’t tell us *what* or *who* is wrong.⁷⁴ (Are we over-sensitive to personal force in response to *footbridge*, or under-sensitive in response to *switch*?) Thus, we could use a more general theory that tells us when our judgments are likely to go astray.

V. Moral implications: The indirect route

According to the dual-process theory, some moral judgments are driven by automatic emotional responses, analogous to a camera’s automatic settings. Other judgments are

⁷⁴ See also Kumar, V. and R. Campbell. 2012. On the normative significance of experimental moral psychology. *Philosophical Psychology* 25(3):311-330.

made in manual mode, based on the controlled, conscious application of moral principles. So, what's better, automatic settings or manual mode?

Some readers seem to think that I think that emotion-based moral judgments are categorically bad.⁷⁵ I do not hold this view.⁷⁶ Rather, I believe, as suggested by the camera analogy, that automatic settings and manual mode are respectively better and worse at different things. Automatic settings are efficient, but inflexible, while manual mode is flexible, but inefficient.

So, when is it better to rely on automatic settings? Automatic settings can function well only when they have been shaped by *trial-and-error experience*. This may be the experience of our biological ancestors, as reflected in, for example, a genetic predisposition to fear snakes. Our automatic settings may be shaped by the experience of our cultural "ancestors," as reflected in a fear of guns, despite one's having never been harmed by one. Finally, our automatic settings are shaped by our own trial and error, as when one learns to fear hot stoves by touching them. These three mechanisms—genetic transmission, cultural transmission, and learning from personal experience—are the only mechanisms known to endow human automatic cognitive processes with the information they need to function well. For one of our automatic settings to function well, its design must be informed by *someone's* trial-and-error experience. Any other way, and it's a *cognitive miracle*.

(Note that being informed by trial-and-error experience is a necessary, but not sufficient, condition for good function. In addition to this, the selection mechanism that

⁷⁵ See Kamm, 2009, *PPA*, *op. cit.* and Woodward, J., and J. Allman. 2007. Moral intuition: its neural substrates and normative significance. *Journal of Physiology - Paris* 101:179-202. Woodward and Allman, by pressing me on this point, have helped me articulate my view more clearly.

⁷⁶ See Secret Joke, *op. cit.*, pp. 66-72.

classifies trials as successful or unsuccessful must employ an evaluative standard that is consistent with the standard we employ in calling a disposition “good” or “bad.” For example, some behaviors, such as aggression toward vulnerable out-groups, may be good from a biological/genetic point of view, but not from ours.)

Let us define *unfamiliar** problems as ones with which we have inadequate evolutionary, cultural, or personal experience. (Here, too, an asterisk indicates the specified technical meaning.) Driving a car is, at first, *unfamiliar.** Our hunter-gatherer ancestors didn’t drive, and cultural familiarity with driving is inadequate to produce skilled driving. And, of course, personal experience with driving is exactly what new drivers lack. Thus, learning to drive requires manual mode—effortful processing involving the conscious application of rules. If one could drive like an experienced driver from the outset, that would be a cognitive miracle.

Likewise, *it would be a cognitive miracle if we had reliably good moral instincts about unfamiliar* moral problems.* This suggests the following more general principle:

The No Cognitive Miracles Principle: When we are dealing with *unfamiliar** moral problems, we ought to rely less on automatic settings (automatic emotional responses) and more on manual mode (conscious, controlled reasoning), lest we bank on cognitive miracles.

This principle is powerful because it, when combined with empirical knowledge of moral psychology, offers moral guidance while presupposing nothing about what is morally good or bad. A corollary of the NCMP is that we should expect certain pathological individuals—VMPFC patients? Psychopaths? Alexithymics? —to make

better decisions than healthy people in some cases. (This is why such individuals are no embarrassment⁷⁷ to the view I will defend in the next section.)

Wielding the NCMP requires us to know (or make educated guesses about) two things: (1) which of our candidate judgments are preferentially supported by automatic settings vs. manual mode, and (2) which of the moral problems we face are unfamiliar*. We'll consider each of these epistemic demands in turn.

Knowledge of which judgments are preferentially supported by automatic settings vs. manual mode is precisely the kind of knowledge supplied by the research outlined in Section II in support of the dual-process theory. And thus, with the No Cognitive Miracles Principle before us, we are finally positioned to appreciate the (indirect) normative significance of this research: If we believe that we ought to rely on automatic settings vs. manual mode to different extents in different situations, and if cognitive science can tell us when we are relying on automatic settings vs. manual mode, then cognitive science gives us normatively significant information—information that can nudge us, if not propel us, toward new and interesting normative conclusions.

It's worth pausing here to correct a common misunderstanding. Some are under the mistaken impression that *neuroscience* plays a special, essential role in my normative

⁷⁷ See Pizarro, D., and D. Bartels. 2011. The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition* 121:154-161. These authors argue against utilitarianism as a normative theory on the grounds that antisocial personality traits predict utilitarian responses to standard dilemmas. However, it is unlikely that such individuals are especially concerned with maximizing happiness. Rather, it is more likely that they are especially unconcerned with causing harm to innocent people. In other words, they are really “un-deontological” rather than utilitarian. More recent research using “process dissociation” supports this interpretation by dissociating these two components. See Conway and Gawronski (2013, *JPSP*, *op. cit.*).

arguments.⁷⁸ It doesn't. Neuroscience is simply one source of evidence concerning which judgments are preferentially supported by automatic settings vs. manual mode. I've not claimed that one can derive moral "oughts" from scientific facts alone (See above); nor have I claimed that neuroscientific facts possess some special normative power. This view is a straw man.

Once again, to apply the No Cognitive Miracles Principle we must determine which of the problems we face are unfamiliar*. We can make educated guesses about unfamiliarity* in two ways. First, moral problems* that arise from recent cultural developments, most notably the rise of modern technology and the intersection of disparate cultures, are especially likely to be unfamiliar.* (Think climate change, global terrorism, global poverty, bioethics, etc.⁷⁹) As it happens, I strongly suspect that the *footbridge* dilemma is unfamiliar*, a bizarre case in which an act of personal violence against an innocent person is the one and only way to promote a much greater good.⁸⁰

Second, and perhaps more practically, we can use *disagreement* as a proxy for lack of familiarity*. If two parties have a practical moral disagreement—a disagreement about *what* to do, not about *why* to do it—it's probably because they have conflicting intuitions.⁸¹ This means that, from a moral perspective, if not from a biological

⁷⁸ See Berker 2009, *PPA*, *op. cit.*

⁷⁹ Such problems undoubtedly share features with familiar* problems, but this need not prevent us from identifying some problems as essentially unfamiliar*.

⁸⁰ I suspect that the *switch* case is an unfamiliarity* double-negative, with unfamiliar* impersonal violence unfamiliarly* promoting the greater good.

⁸¹ Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108:814-834. Haidt argues that moral controversies are primarily driven by conflicting moral intuitions.

perspective, at least one party's automatic settings are going astray. (Assuming that both parties have adequate access to the relevant non-moral facts.⁸²) Absent a reliable method for determining whose automatic settings are misfiring, both parties should distrust their intuitions. Thus, I propose that we distrust our automatic settings, and rely more heavily on manual mode, whenever we have practical moral disagreements that do not appear to be based on disagreements over non-moral facts. (And, for that matter, I make the same recommendation for when we think that our disagreements *are* over non-moral facts.)

I've said that we should not rely on our automatic settings when we're dealing with unfamiliar* problems. Two clarifications concerning the meaning of "rely": First, not relying on automatic settings doesn't mean assuming that the answers given by one's automatic settings are always wrong. This misguided policy is like that of a dieter who never eats because he's always hungry. Second, not relying on automatic settings doesn't mean discounting one's judgments when and only when one's automatic settings are actively engaged. Automatic settings can fail by being over-active (e.g., fearing a life-saving injection) or under-active (e.g., not fearing the subway's third rail). Relying on our automatic settings means allowing our judgments to follow their *ups and downs*.

On the one hand, the normative significance of dual-process moral psychology, by way of the NCMP, flies in the face of is/ought orthodoxy. On the other hand, our conclusion here shouldn't be too hard for most ethicists to embrace. The idea that we

⁸² Doris, J. M., and A Plakias. 2007. How to argue about disagreement: evaluative diversity and moral realism. In *Moral Psychology, Volume 2: The cognitive science of morality*, edited by W. Sinnott-Armstrong. Cambridge, MA: MIT Press. These authors argue many moral disagreements are not simply disagreements over non-moral facts.

should apply manual mode thinking to complex, controversial moral problems sounds more like a want ad for ethicists than a sheriff's "Wanted" sign. Things get a bit stickier, however, when we combine the NCMP with what we're learning about the cognitive underpinnings of competing moral philosophies.

VI. Tilting towards consequentialism

We should distrust our automatic settings and rely more on manual mode when attempting to resolve practical moral disagreements. So far, so palatable. But where does this lead? I believe it favors consequentialist approaches to moral problem-solving, ones aimed solely at promoting good consequences, rather than deontological approaches aimed at figuring out who has which rights and duties, where these are regarded as constraints on the promotion of good consequences. More specifically, I believe that reliance on manual mode favors act consequentialism at the level of first principles and something resembling rule consequentialism in everyday practice.⁸³ As private individuals, we should nearly always respect the conventional moral rules, but in establishing those rules (as voters and policy-makers) we should aim simply for the best long-term consequences. (For present purposes I will leave aside questions concerning the meta-ethical status of these claims.)

Why consequentialism? Recall from Section II:

⁸³ See Hare, R. M. 1981. *Moral thinking: Its levels, method, and point*. Oxford: Oxford University Press.

The Central Tension Principle: Characteristically deontological judgments are preferentially supported by automatic emotional responses, while characteristically consequentialist judgments are preferentially supported by conscious reasoning and allied processes of cognitive control.

Thus, shifting into manual mode means shifting toward consequentialism. Satisfied? Maybe not.

Why, you might ask, must manual mode thinking be consequentialist? Do not others—Kant, for example—engage in *moral reasoning*? It may seem that I’ve made a giant, unwarranted leap from empirical facts about the psychology behind ordinary people’s “characteristically” deontological judgments to claims about the psychology behind sophisticated philosophical theories.

There’s no denying that Kant et al. do plenty of work in manual mode. The critical question is: What are they doing in manual mode? I hypothesize that they are not, *for the most part*,⁸⁴ actually engaged in moral reasoning. By this I mean that they are not using reasoning to *figure out* what’s right or wrong. Instead, their reasoning serves primarily to justify and organize their preexisting intuitive conclusions about what’s right or wrong. In other words, what looks like moral rationalism is actually moral *rationalization*.⁸⁵

I’ve called this the Secret Joke of Kant’s Soul,⁸⁶ which is colorfully illustrated by Kant’s views on masturbation. Kant, being an uptight Eighteenth Century Prussian, is

⁸⁴ See below on biting “rubber bullets,” which involves genuine reasoning. Kant, however, bites at least one metal bullet. (See Section II.)

⁸⁵ Cf. Haidt, 2001, *Psych Review*, *op. cit.*

⁸⁶ Greene, 2007, *Secret joke*, *op. cit.*

uncomfortable with masturbation, but he's not content simply to voice his distaste. He wants to *prove from first principles* that masturbation is immoral, and he's got a pretty clever idea about how to do it: Masturbation is wrong because it involves *using oneself as a means*. We today find this bit of rationalization amusing because we no longer share Kant's sexual repression, but if I'm right, this passage is in fact representative of his general approach to ethics. Nietzsche agrees.

Kant's Joke--Kant wanted to prove in a way that would dumbfound the common man that the common man was right: that was the secret joke of this soul. He wrote against the scholars in favor of the popular prejudice, but for scholars and not for popularity.⁸⁷

If Nietzsche is right, this Kantian style of rationalizing goes well beyond the ethics of masturbation. A standard method—if not *the* standard method—for identifying the principles that define our rights and duties is rather like Kant's method in the above passage. One discerns, intuitively, the presence of rights and duties in particular cases, and then one searches for principles that might explain why those rights and duties are indeed present. I call this process *intuition-chasing*, conforming general principles to specific judgments that (mostly) follow the ups and downs of intuition. The opposite of intuition-chasing is *bullet-biting*, conforming judgments to principles, despite the ups and downs of intuition.

⁸⁷ Nietzsche, F. 1882/1974. *The Gay Science*. New York: Random House, pp. 205-206.

A nice empirical illustration of intuition-chasing and its pitfalls comes from an experiment by Eric Schwitzgebel and Fiery Cushman.⁸⁸ They presented both ordinary people and philosophers with cases like *footbridge* (harm as means, with personal force) and cases like *switch* (harm as side-effect, without personal force). We'll call these *footbridgesque* and *switchy* cases, respectively. Because these sets of cases straddle the means/side-effect distinction, one may appeal to the doctrine of double effect (DDE) to justify treating them differently.

This experiment leverages people's tendency to shift around their judgments depending on the *order* in which cases are presented. This order effect is explicitly predicted by the dual-process theory as follows: Suppose you get *footbridge* first. Your automatic emotional response is "No!," and you go with it. Next comes *switch*. Here the emotional response is minimal, leaving the decision to manual mode. Your manual mode thinking is naturally drawn toward the characteristically consequentialist response, but here it has a further problem. You just said that it's wrong to trade one life for five in the *footbridge* case, and you're not confident that there is a morally relevant difference between the *switch* and *footbridge* cases. In an effort to be consistent, you judge that it's also wrong to hit the switch. That is, you give the same response to both cases, which is not the pattern of judgment encouraged (though not required) by the DDE.

What happens when *switch* comes first? Here, once again, the emotional response is minimal, allowing manual mode to hold sway. Manual mode is drawn to the characteristically consequentialist response, and in this case there is no consistency-

⁸⁸ Schwitzgebel, E. and F.A. Cushman. 2012. Expertise in moral reasoning?: Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language* 27(2):135-153.

based reason to say otherwise. Thus, you endorse hitting the switch. Next comes *footbridge*. The automatic setting kicks in, and the action feels terribly, horribly wrong. Manual mode recognizes the consequentialist rationale for pushing, along with the consistency-based rationale for endorsing pushing, but your automatic emotional response tells you that you simply cannot endorse such a horrific act. Thus, you say “no” to pushing. This response pattern—“yes” to *switch*, “no” to *footbridge*—is the one encouraged by the DDE.

To summarize, if you get *switch* first, your pattern of judgment is more likely to conform to the DDE. This happens because it’s a lot easier to say “no” to something that feels okay than to say “yes” to something that feels horribly wrong. In other words, saying “no” to *switch* involves biting a bullet, but it’s a *rubber* bullet—easy to bite because it’s actually a manual mode consequentialist* judgment that is jettisoned. To say “yes” to *footbridge* is to bite a *metal* bullet—much harder. The desire for consistency is sufficient to motivate biting a rubber bullet, but not a metal bullet—hence the order effect.

Results: Both ordinary people and professional ethicists exhibited the predicted order effect, and, remarkably, to the same extent. Later, Schwitzgebel and Cushman asked their participants whether they endorse (the critical bit of) the DDE. For the non-philosophers, the ordering of the dilemmas had no effect on their endorsement of the DDE. However, the ethicists were about 50% more likely to endorse the DDE if they got the *switchy* cases first.

Let us return to the motivating problem at hand, comparing lay moral psychology to professional moral psychology: Deontological philosophers, for understandable reasons, don’t want to be lumped in with lay moralists, who are undoubtedly much less thoughtful. Schwitzgebel and Cushman’s results help explain

what's right about this complaint, while also speaking to its limitations. Their experiment shows that philosophers *are* different from lay moralists and that they do indeed think harder. Both groups' judgments were affected by their intuitions, and both groups made a manual mode effort to be consistent. However, only professional philosophers felt compelled to adjust their theoretical commitments to make them consistent with the judgments they'd made. In other words, the folk are happy to let "popular prejudice" be "popular prejudice," but philosophers are motivated to translate that popular prejudice into principle. This is indeed a manual mode activity, but it's not one that is likely to free philosophers from the ups and downs of their automatic settings. On the contrary, it codifies those ups and downs.

I suspect that many readers will be tempted to draw a different lesson from this experiment, which is simply that some philosophers ought to have their licenses revoked. Suppose that's correct. Where *exactly* did our errant colleagues go wrong? The most popular answer, I think, will be this: The philosophers who said "no" to the *switchy* cases following their "no" to the *footbridgesque* cases messed up. Instead of being "consistent" by saying "no" to both, they should have recognized that these cases differ in ways that *justify* treating them differently. In other words, they were too quick to bite a bullet, albeit a rubber one.⁸⁹

Let's see where this leads. What critical difference did the errant ethicists miss? By design, the *switchy* and *footbridgesque* cases differ in two ways, in the presence of personal force and in the causation of harm as a means vs. side-effect. We've agreed that personal force is morally irrelevant, which leaves the means/side-effect factor, or perhaps a more subtle variant thereof. The accusation, then, is that the errant

⁸⁹ But see Thomson, 2008. Turning the trolley. *Philosophy and Public Affairs* 36 (4):359-374.

philosophers failed to remember the DDE. (Revoke their licenses!) But why *should* they have invoked the DDE? Wherein lies its justificatory power? It's been on the books for a long time, which gives it a dusty air of authority. But how did it get on the books? It seems that the DDE was codified because it was observed that certain intuitive patterns in moral judgment could be *summarized* by a set of principles now known as the DDE. We may infer this (inductively, not deductively) from the finding that lay moralists the world over make judgments consistent with the DDE while having no clue that they are doing so.⁹⁰ This suggests, in other words, that the Doctrine doesn't justify the judgments. Instead, *the judgments justify the Doctrine*.⁹¹ This evidence suggests that the justification for the DDE ultimately comes from nothing beyond the automatic settings that produce the pattern of judgment that it summarizes.⁹² (And if I'm right, the DDE is actually just a by-product of a morally irrelevant cognitive limitation.⁹³) Indeed, the DDE's lack of independent authority is evident in the willingness of philosophers to abandon it when it fails to get the intuitively right answers. The DDE famously choked

⁹⁰ Hauser et al., 2007, *M&L*, *op. cit.* The alternative explanation—far less likely in my opinion—is that ordinary people, including young children (Pellizzoni, Sandra, Michael Siegal, and Luca Surian. "The contact principle and utilitarian moral judgments in young children." *Developmental science* 13, no. 2 (2010): 265-270), make judgments consistent with the DDE due to unacknowledged philosophical influences.

⁹¹ Cushman, F.A., and J.D. Greene. 2011. The philosopher in the theater. In *The Social Psychology of Morality*, edited by M. Mikulincer and P. R. Shaver. Washington, D.C.: APA Press. See also Cushman, F.A. et al. 2012.. Judgment before principle: Engagement of the frontoparietal control network in condemning harms of omission. *Social, Cognitive, and Affective Neuroscience* 7(8):888-895.

⁹² One might note that it's been in use for a long time. (See [Mikhail in this volume].) True, but so have a lot of other moral and legal practices that few readers would defend.

⁹³ Greene, 2013, *Moral Tribes*, *op. cit.*

on the *loop* case,⁹⁴ for example, prompting a hunt for a better principle. But what is “better?” “Better” just means “better able to summarize our intuitive judgments.” Thus, to chide our colleagues for failing to invoke the DDE (or one of its more sophisticated successors) is simply to chide them for not abiding by the dictates of their—our—mysterious automatic settings.

I’ve said that characteristically deontological judgments are used to justify deontological principles, rather than the other way around, but that’s a bit too strong. What intuitions-chasing aims for is *reflective equilibrium*,⁹⁵ harmony between intuition and principle. This means that some intuitive judgments will be jettisoned to achieve a better fit with principle. In other words, even intuition-chasers do some bullet-biting. While this is undoubtedly true, I doubt it’s enough to save intuition-chasing from the unreliability of intuition. Suppose that a scientist unwittingly crafts a theory based on unreliable data. In so doing, he ignores some outliers, data points that don’t fit with his theory. He’s then informed that the data are unreliable. “Indeed!” he says, “Just as I suspected!” The reflective equilibrator’s analogous hope is that the judgments he’s jettisoned through the equilibration process, the outliers, are the bad ones. I think that’s too optimistic. We may think that we’re biting the right bullets, when in fact we’re just biting the *soft* bullets, the ones that are least emotionally offensive.

Let us consider this pitfall in its proper trolleyological context. What should a deontologically-minded trolleyologist make of the personal force effect and the doubts it raises about our intuitions? I’ll take this as an opportunity to respond to some

⁹⁴ Thomson, 1985, *YLJ*, *op. cit.*

⁹⁵ Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

remarks by Frances Kamm.⁹⁶ Here is Kamm’s response to the first iteration of the “personalness” hypothesis from my 2001 paper:

However, objections have been raised to this type of “personal versus impersonal factors” explanations of intuitions... the Lazy Susan Case is a counterexample to this explanation. In this case, five people are on one side of a lazy Susan turntable and the only way to save them from a trolley headed at them is to manually push the lazy Susan. However, if we do this, we will ram the lazy Susan into a bystander. In this case, a new threat (the lazy Susan) and something up close and personal is done to the bystander (pushing a lazy Susan into him)... Intuitively, even a nonconsequentialist might agree it is permissible to act in such cases. Cases like the Lazy Susan Case led me to say that “nonconsequentialists are not squeamish, they are downstreamish,” in the sense that they are willing to “up close” personally harm someone when (very roughly) the person’s being killed is causally downstream from the good of the five being saved.

Kamm’s response highlights the importance of doing the empirical work, as it’s all but impossible to know our minds from the armchair. As it happens, Kamm’s intuition about the Lazy Susan case and others like it⁹⁷ is widely shared. But it’s also the case that people—Kamm included, perhaps—are *both* “squeamish” and “downstreamish.” As noted in Section III such factors *interact*. In the Lazy Susan case and others like it, the

⁹⁶ Kamm, 2009, *PPA*, *op. cit.*

⁹⁷ 80% approval (unpublished data). See also *obstacle collide* in Greene et al., 2009, *Cognition*, *op. cit.*

effect of “squeamishness” disappears because the person is killed as a downstream side-effect. But when the harmful event is upstream, as in the footbridge variations, the effect of personal force (a kind of “squeamishness”) emerges. Once again, most people say “yes” to *footbridge switch*, but “no” to *footbridge pole* and *footbridge*. Kamm anticipates this contrast:

What if it were possible to press a switch that opens a trapdoor under the man on the bridge... ? Regardless of what fMRIs of the general population indicate, moral philosophers who object to pushing the man over the bridge would respond in the same way to this “impersonal” way of getting the man off the bridge.

Indeed, no self-respecting philosopher would explicitly give different answers to these two cases. But are we then licensed to ignore the experimental research, confident that our philosophical reflections will jettison any bad intuitions and corresponding judgments? Not so fast. We dare not give different answers to these two cases, but which single answer should we give? Kamm’s answer is clear: We should say that it’s wrong to act in both cases. But why? The psychological explanation is now familiar: We say “no” to both cases because saying “yes” to *footbridge* (metal bullet) feels worse than saying “no” to *remote footbridge* (rubber bullet). Thus, what appears to be a case of experts making better judgments than the folk (avoiding the pitfalls of untutored intuition through careful reflection) may in fact be a case of experts out-folking the folk (making the not-so-hard choices necessary to protect our strongest untutored intuitions from inconsistency). To put the point another way, unless you’re prepared to say “yes”

to the *footbridge* case, your automatic settings are still running the show, and any manual adjustments that you're willing to make are at their behest.

But, so what if our emotions are leading us around? Perhaps they're leading us to the right places (with a few bumps along the way). Kamm's hope is that some other distinction, a more sophisticated variation on the DDE perhaps, will save the day. But this just puts us back where we were in condemning the "errant" philosophers in Schwitzgebel and Cushman's experiment. The DDE and its more subtle successors have no independent justification. The most that can be said for them is that they appeal to factors that are not obviously morally irrelevant (unlike personal force). As noted above, I suspect that the means/side-effect distinction will seem irrelevant once we understand why we're sensitive to it. We can wait and see if I'm right, but we can't be indifferent to the outcome of this research, or assume its conclusion.

The upshot of the foregoing discussion is this: Ethicists need to worry about their intuitions, and not just the ones that they're willing to dump in order to save the ones they really want to keep. We can't assume that our manual mode thinking will scrub away the blemishes of moral intuition if we're relying on our moral intuitions to tell us where to scrub.

The next step is to agree, but to insist that this is a problem for everyone, consequentialists and deontologists alike. In the end, don't we all just have our intuitions? No. Not in the same way. Act consequentialism is not intuition-chasing. An act consequentialist's judgment may be *consistent* with the dictates of automatic settings, and in that sense there are "consequentialist intuitions," but an act consequentialist's judgment never *depends* on them. An act consequentialist can know what she thinks about a case without knowing anything other than the answer to this question: Which choice produces better consequences? Act consequentialism is WYSIWYG: What You

See Is What You Get. It doesn't rely on mysterious automatic settings, and thus its soul has no secrets.

But doesn't act consequentialism ultimately depend on some kind of intuition? After all, where do act consequentialists get their ideas about which consequences are worth promoting or preventing? Here Sidgwick is helpful in distinguishing among what he calls "perceptual," "dogmatic," and "philosophical" intuitions.⁹⁸ If Sidgwick is correct, consequentialism (and utilitarianism, more specifically) does ultimately depend on "intuition," i.e. on an affectively based evaluative premise. But intuition enters consequentialist theory at a very high level ("philosophical" intuition), and not as a reaction to particular actions ("perceptual" intuition) or action types ("dogmatic" intuition). This is consistent with the psychological evidence, for example, the fact that VMPCF patients, who cannot react emotionally to particular actions, tend to make characteristically consequentialist judgments. That act consequentialism is based on a "philosophical" intuition, rather than on "perceptual" or "dogmatic" ones, doesn't imply that it's correct, but it does shield it from the objection that it's too tightly yoked to the ups and downs of unreliable automatic settings.

Let's suppose, then, that we've forsaken all intuition-chasing. Why favor act consequentialism? Aren't there better alternatives? First, let's note that act consequentialism is a pretty good place to start. The idea that we should try to make things overall better makes moral sense to everyone. The objection to act consequentialism is not that it's based on a generally bad idea, but that it's too imperialistic, that it fails to leave room for other legitimate values, both moral and self-

⁹⁸ See [Brink in this volume], and Sidgwick, H. 1907. *The Methods of Ethics*, 7th ed. London: Macmillan, pp. 97-102.

interested. Like other act consequentialists, I believe that act consequentialism, properly understood, does surprisingly well in the real world, and that its failures “in principle,” are really failures to comport with intuitions that are not worth chasing. But I will not offer a positive defense of act consequentialism here.⁹⁹ For now, my point is simply that act consequentialism should get points for not chasing intuitions and that some of its competitors (including some forms of consequentialism) should lose points for doing so. Note that the present argument also casts doubt on theories that, rather than chasing intuitions with codifying principles, simply allow our intuitions roam free. Likewise, it casts doubt on theories that purport to derive from first principles, but that are in fact intuition-chasing—i.e., theories that are actually attempts to get from first principles *to the intuitively right answers* rather than attempts to get from first principles to wherever those principles happen to lead. (And, if you’re like me, you suspect that this covers most, if not all, of act consequentialism’s competition.)

VII. Conclusion

A great philosopher who despised the likes of me once wrote, “It is not profitable for us at present to do moral philosophy; that should be laid aside at any rate until we have an adequate philosophy of psychology, in which we are conspicuously lacking.”¹⁰⁰ I think Professor Anscombe went too far. I don’t regret the last half-century of ethics. Nor do I

⁹⁹ See Greene, 2013, *Moral Tribes*, *op. cit.*

¹⁰⁰ Anscombe, G.E.M. 1958. Modern moral philosophy. *Philosophy* 33 (124):1-19.

share her vision for our enterprise. But she did have a point. There's a lot going on beneath the hood of morality, and we ignore the details at our peril.

Our brains, like all complex functional systems, face a tradeoff between efficiency and flexibility. To promote efficiency, our brains have point-and-shoot automatic settings in the form of intuitive emotional responses. These are marvelous, but nonetheless limited in what they can do. In particular, we should not expect them to perform well in the face of peculiarly modern problems, ones with which we have inadequate genetic, cultural, and individual experience. Many of the most important moral problems we face may be of this kind. For example, it's only very recently that ordinary people have routinely had the opportunity to save the lives of distant strangers by making modest material sacrifices. Our automatic settings tell us that our exercise of this privilege is generally morally optional. Should we trust them?¹⁰¹

Philosophy happens in manual mode, and this gives us a choice about what to do with our reasoning capabilities. We can use them, as William James said, to "rearrange our prejudices." Alternatively, we can use our capacity for moral and scientific reasoning to transcend our prejudices. Today's ethicists undoubtedly do both, but we could do more of the latter and less of the former.

In the last half-century, ethics has gone in two different directions. Some ethicists have gotten very sophisticated about moral theory, devising principles that aspire to comport with both consequentialist thinking and incompatible moral intuitions. Others have given up on action-guiding moral theory altogether, and with it the hope of offering the world specific answers to specific moral questions. I don't think we should

¹⁰¹ Singer, P. 2009. *The life you can save: Acting now to end world poverty*. New York: Random House. Singer, P. 2005. Ethics and Intuitions. *Journal of Ethics* 9 (3-4):331-52.

abandon this ambition. But if I'm right, the way forward is not through the construction of increasingly sophisticated moral theories. Instead, we must get sophisticated about moral *data*. We need to understand the structure, origins, and limitations of our intuitive moral thinking, the better to know when our moral instincts are on target, and when they are giving us the right answers to the wrong questions.¹⁰²

Is this a new approach to ethics, or just a natural extension of current practice? It's a bit of both. On the one hand, this approach may be seen as nothing more than a further widening of our reflective equilibrium. Along with our "considered judgments" and organizing principles, we must add to the mix a scientific understanding of the psychological and biological processes that have produced them. (Call this *double-wide* reflective equilibrium.¹⁰³) I regard this as a natural extension of current practice, and one consistent with philosophy's historical commitment to active empiricism.

Nevertheless, from a professional perspective, what I am proposing may sound rather radical. Today, ethicists are not expected to know anything in particular about how the mind actually works, and are trained to dismiss anyone who laments this practice as confused about the relationship between "is" and "ought." I suggest that, in the future, we ethicists will require a detailed knowledge of moral psychology, and the more actively we participate in the generation of such knowledge, the better off we'll be. To do moral philosophy well we must understand the strengths and limitations of the tools we bring to our job. Why would anyone think otherwise?

¹⁰² See also Stich, S., 2006. Is morality and elegant machine or a kludge? *Journal of Cognition and Culture*, 6, 1-2, 181-189.

¹⁰³ The problem with reflective equilibrating is its susceptibility to the influence of inflexible automatic settings. If these ill effects can be neutralized with scientific self-knowledge, then reflective equilibrating is fine.