Justin D'Arms and Daniel Jacobson (eds.), *Moral Psychology & Human Agency: Philosophical Essays on the Science of Ethics,* Oxford University Press, 2015, 282pp., $74.00, ISBN 9780198717812.

David Faraci, UNC Chapel Hill

Moral Psychology & Human Agency: Philosophical Essays on the Science of Ethics contains essays from the Workshop on Moral Psychology and Human Agency that took place at the University of Michigan in 2012, funded through co-editor Daniel Jacobson's Templeton grant, "The Science of Ethics." In the Introduction—which is primarily an overview of the book's contents—the editors describe the contributors as "philosophers who share our conviction that scientific inquiry is relevant to various classic philosophical questions, while also showing an appreciation for the difficulty of these questions that is not always evident in empirical moral psychology" (2). For the most part, the contributions bear this out. I'll examine each, critiquing them as stand-alone works in addition to discussing some of their connections to each other and the book's themes.

The first contribution is Guy Kahane's "Intuitive and Counterintuitive Morality" (Ch. 2). Kahane targets Joshua Greene, who argues for a "dual process model" on which moral judgements can be generated either by an "automatic" system of intuitions, or by a "deliberative" rational system. Greene also offers evidence purporting to show that the latter system tends to generate utilitarian judgments, the former deontological ones. Greene takes this to speak in favor of utilitarianism.

Kahane argues convincingly that judging that one should push the fat man in front of the trolley (e.g.) typically requires deliberation because it is *counterintuitive,* not because it is utilitarian. Yet Kahane also writes that Greene "has suggested to me that his dual process theory is really meant to essentially make this more general claim about intuitive and counterintuitive judgment" (17). Kahane

dismisses this quickly, pointing out that (a) this is not how Greene presents the theory in print; (b) this "banal" claim about intuitive vs. counterintuitive judgment is practically a truism; and (c) if Greene thinks deliberative judgements *favor* utilitarianism, he needs to provide more evidence.

Kahane then argues that even in cases where a counterintuitive result is in line with utilitarianism, the deliberation that occurs isn't utilitarian calculation, but rather a process of resolving conflict between competing intuitions—including non-utilitarian ones. He further provides evidence that there *are* people who at first blush appear to reason in a utilitarian fashion; but their "judgments actually turn out to be *non-deliberative*—and to reflect a strong antisocial, self-centered tendency" rather than a concern with the greater good (34).

If Greene really is comfortable with Kahane's "banal" dual process theory, the force of these further critiques may not be entirely apparent. Consider three types of people, the existence of whom is suggested by Kahane's discussion: those with (quasi-)utilitarian intuitions; those with deontological intuitions; and those who have to resolve conflicts between the two. Greene would presumably maintain that only the judgements of the third group have epistemic merit, and that these tend towards the utilitarian. Of course, he needs to gather further evidence for the latter claim, as Kahane suggests. But why should it trouble Greene that the deliberators aren't directly engaged in utilitarian calculation? Perhaps deliberators need to consider and overcome deontological intuitions in order to arrive at the truth! And what is the relevance of the "automatic pseudo-utilitarians?"

I think there are good answers to these questions, and that they point to a deeper critique of Greene. Help comes at the opposite end of the book, from the editors' (D&J) own contribution, "Sentimentalism and Scientism" (Ch. 11). D&J consider Peter Singer's contention that because we have good evidence that many of our moral judgements are rooted in emotionally laden intuitions, we face a choice: We can accept that *all* moral judgements are so rooted, leaving us with skepticism

or relativism. Or we can find a way to distinguish those judgements that are *not* so rooted, that have a rational basis.

As D&J note, Greene champions the latter option. For Greene, what ultimately matters is our ability to cordon off "rational" moral judgements from automatic/emotional/intuitive ones, lest we fall into skepticism or relativism, not just that the former favor utilitarianism. But Kahane has offered evidence that *no* process here is a purely rational one. Some people start with deontological intuitions, some start with (pseudo-)utilitarian ones, and some have to resolve conflicts between the two. Even if *everyone* who was conflicted ultimately leaned towards utilitarianism (which they don't), that would hardly prove the view more rational, especially since there is no evidence that the process of conflict-resolution is rational itself.

Reading Kahane alongside D&J, we find a critique of Greene's attempts to locate purely rational moral judgements. D&J critique the goal itself. They contend that their own view, rational sentimentalism, represents an option Singer, Greene and others miss: Moral judgements need be neither purely rational nor wholly non-rational; our sentiments may provide genuine guides to (an explicitly anthropogenic) morality, yet also be subject to rational assessment. On this view, the automatic processes Greene denigrates have epistemic merit after all, and not just as guides to something subjective or relativistic.

As a metaethicist with realist sympathies, I'm not yet on board with rational sentimentalism. But the goal of the D&J chapter isn't primarily to argue for the view; it is to expose flaws in various defenses of (often scientifically motivated) pessimism about the possibility of rationally assessable sentimental values. In this, they are overwhelmingly successful.

I think it unfortunate that the Kahane and D&J chapters appear at opposite ends of the collection. I suspect the reason is that the editors recognized that their chapter is closest in theme to Selim Berker's "Does Evolutionary Psychology Show that Normativity Is Mind-Dependent?" (Ch.

10), a critical discussion of Sharon Street's evolutionary debunking arguments in metaethics. But Berker's chapter has little to do with moral psychology or the "science of ethics," and frankly does not belong in this collection. The obvious defense would be that the evolutionary nature of Street's arguments makes it relevant. But (like Street herself) Berker is explicit that focusing on evolution is just one way of motivating Street's challenge for realism; it "is not essential for generating the puzzle" (244).

The next contribution is Brendan Dill and Stephen Darwall's (D&D) "Moral Psychology as Accountability" (Ch. 3). D&D defend a unified account of condemnation and conscience, arguing that both can be understood in terms of accountability. They are largely successful in undermining both experimental and philosophical defenses of various alternatives, and providing experimental evidence for their view.

My only significant worry about the empirical discussion concerned their claim that "the condemnation motive [is] satisfied when and only when the perpetrator has adequately held himself accountable for his wrongdoing" (50). They do not discuss what seem to me obvious foils, such as the death penalty. If the condemnation motive is sometimes satisfied by procuring the death of wrongdoers before they have repented, this threatens D&D's claim. And even if it does not satisfy the motive *fully,* the fact that many support the death penalty suggests that more than accountability is motivating them.

My larger concern is the chapter's framing as a defense of the claim that morality is to be understood, conceptually, in terms of accountability—a view Darwall is famous for championing elsewhere. D&D acknowledge that there are normative ideas that some consider moral, such as honor or purity, that don't necessarily involve accountability. But they deny that these really are moral. There are some hints that they take their position to be bolstered by empirical data—that the fact that, for instance, shame is psychologically quite different from guilt exposes the fact that the

latter, but not the former, is moral. But I see nothing here that would or should move someone who denies that morality is all about accountability—unless they did so only because they embraced Haidt and Kesebir's functional view, which D&D argue (convincingly, I think) is too broad.

The next two chapters concern the psychology of moral responsibility. David Shoemaker's "Remnants of Character" (Ch. 4) begins with a puzzle: Why do certain psychological abnormalities arouse in us a certain ambivalence regarding moral responsibility, a "feeling that while the agents are exempt in one sense, they may not be in another" (86)? (Full disclosure: Shoemaker and I have published together on some related issues.)

Shoemaker focuses on dementia, arguing that it is inappropriate to hold a demented agent *accountable* for her actions when she can neither remember what was done nor identify with the agent of those actions. Importantly, this does not preclude *attributing* actions to her in a way that legitimates assessment of her character. A good deal of the chapter is spent considering how to handle cases where traits exhibited by the demented agent seem inconsistent with those of her non-demented self. Shoemaker's solution is to maintain that our *deep selves* contain *clusters* of character traits, and that dementia may dampen some traits, thereby making others more apparent.

I think Shoemaker's position here is plausible and well-defended. But I worry a bit about drawing psychological conclusions from intuitions about cases, especially in a book on empirically informed moral psychology. Shoemaker presents his view on the nature of character as a theory that our intuitions about responsibility provide evidence for. I wish he had instead framed it as a psychological *hypothesis suggested* by our practices of holding responsible, one whose confirmation or refutation might either vindicate or undermine those practices. We should be prepared to find that Shoemaker is entirely right about how we *tend to think* about character and dementia, but that psychologically we are off base. This is not a major criticism; Shoemaker might even be happy to accept my reframing. But I do think it important that philosophers be wary of such issues, both for

dialectical clarity as well as for rhetorical reasons, especially for those expecting their work to be read by non-philosophers.

Heidi Maibom's "Knowing What We Are Doing" (Ch. 5) aims to account for cases where "we are responsible for acting under situational influences that we are unaware of at the time of action, such as stereotyping or standing by when others are in need" (109). I share many of Maibom's intuitions about when agents are morally responsible. And it certainly seems true that a capacity for self-reflection is implicated in such cases, as she suggests. But Maibom is concerned not just with defending a view about what explains responsibility, but also with arguing that it undermines Deep Self views (such as Shoemaker's).

Unfortunately, the Deep Self view she considers strikes me as a straw man. Maibom apparently takes Deep Self views to require a focus on idiosyncratic aspects of an agent's psychology. In critiquing Deep Self views she writes: "There is nothing personal about being subject to the situational effects outlined by the social psychology literature . . . These are tendencies that we share with many or, in some cases, most people. They are part of who we are, of course, but not in a way that individuates us as persons" (116). Later, she writes that "the capacity for self-reflection is impersonal and universal. It has nothing to do with what kind of person someone is" (120).

I'm rather puzzled by these remarks. As I understand Deep Self views, nothing precludes sharing aspects of one's deep self with others—even all others. And it seems plausible that the capacity for self-reflection varies from agent to agent, anyway. So I don't see what Maibom has done to undermine the idea that one's capacity for self-reflection is part of one's deep self, or that this is relevant to one's level of responsibility in certain cases.

Next is Julia Driver's "Meta-Cognition, Mind-Reading, and Humean Moral Agency" (Ch. 6). Driver maintains that a number of conditions—especially affective disorders—that have been taken to eliminate moral agency merely lessen it. Driver's case studies are interesting, and there are some

exegetical issues that will likely be of interest to Hume scholars; but the discussion is so broad, and there are so many background assumptions that I found myself unclear what philosophical lessons could be drawn. (For example, Driver simply takes for granted the view that practices of holding accountable are pragmatically justified in terms of social utility.)

The remaining three chapters all defend theories in the psychology of moral agency. Shaun Nichols' "The Episodic Sense of Self" (Ch. 7) argues that humans have both a trait-based, thick conception of the self, as well as a much thinner "mere *I*," evidence for which can be found in the nature of episodic memory. Nichols does a particularly nice job of showcasing the potential benefits of applying philosophical acumen to empirical data in order to expose their relevance to philosophy without drawing hasty conclusions on either side.

Andrea Scarantino's "The Motivation Theory of Emotions" (Ch. 8) defends that very theory of emotions, which holds that emotions are goal-oriented action tendencies that are *prioritized*—they take precedence over other action tendencies and are prepared for execution. Scarantino argues that emotions have motivational features that his view accounts for better than standard alternatives. This seems true in many cases. But for the theory to be successful, *all* emotions must exhibit these features, and there must be no *other* features of emotions the theory fails to account for. I unfortunately do not have the space here to discuss Scarantino's attempts to accommodate apparently non-motivating emotions in his theory. But Scarantino is explicit that according to his view, "appraisals and feelings are no longer essential components of emotions" (180). I suspect some would consider these to be features a theory of emotions needs to accommodate. Partly because of (admittedly inchoate) worries along these lines, I find myself wondering why Scarantino's theory should be accepted as a theory of emotions themselves, rather than of the impulses emotions (typically) generate.

Finally, there is "The Reward Theory of Desire in Moral Psychology," (Ch. 9) in which Timothy Schroeder and Nomy Arpaly argue that the philosopher's notion of an intrinsic desire is a psychological natural kind: to intrinsically desire something is to constitute it as a reward in the technical sense associated with psychological theories of reward learning. Their defense of the view—drawn from their recent book, *In Praise of Desire*—is sophisticated and compelling.

I'll close with a small worry: I suspect the view will need to be further complicated if it is to properly distinguish intrinsic from instrumental desires. Imagine you intrinsically desire an end which you know to necessarily covary with a particular means. I suspect your reward system would respond to the two in the same way. Yet, surely, it does not follow that you intrinsically desire the means.