

single articles by Native American, Asian-American, and Hispanic authors and subjects, all of which clearly deserve more attention.

### References

- Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imagination*. New York: Oxford University Press.
- Sullivan, Shannon. 2006. *Revealing Whiteness: The Unconscious Habits of Racial Privilege*. Bloomington, IN: Indiana University Press.
- Yancy, George. 2012. "Looking at Whiteness: Tarrying with the Embedded and Opaque White Racist Self." In *Look, a White! Philosophical Essays on Whiteness*, 153–75. Philadelphia: Temple University Press.

*Julie K. Ward*

Loyola University

*Philosophical Review*, Vol. 128, No. 1, 2019

DOI 10.1215/00318108-7213372

Mark Fedyk, *The Social Turn in Moral Psychology*.  
Cambridge, MA: MIT Press, 2017. x + 246 pp.

Mark Fedyk argues persuasively for both the importance and the perils of interdisciplinarity in studies of ethical life. The book is dense with incisive argumentation and innovative proposals for integrating moral, social, and political philosophy with the psychological and social sciences. It will be of interest to aprioristically inclined normative and social theorists peeking over the fence at the empirical side of things, to experimentalists trying to operationalize or intervene on real-world ethical thought and action—and to everyone in between.

The early chapters aim to place some of the most common pitfalls of contemporary moral psychology into theoretical context. Fedyk takes issue, in particular, with the fervor for positing universal and innate moral-cognitive modules solely on the basis of speculations about what would have been adaptive for our ancestors. The problem with this inference, of course, is that numerous alternative, equally plausible psychological (and social) mechanisms could satisfy such evolutionary considerations. Relatedly, Fedyk argues that we cannot simply deduce, from an accurate account of the norms operative in a given community, precisely how those norms are psychologically represented by community members. The broader lesson, which Fedyk calls Mayr's lemma, is that different levels of explanation, such as the psychological and the structural, or the mechanistic and the teleological, stand in a many-to-many relation. Many

different psychological mechanisms could generate the behaviors that fit into a given structure, and many different social structures could arise from a given psychological mechanism. To bridge the gaps between explanatory levels, studies must be carefully designed to give all of the competing plausible explanations a fair shake.

This many-to-many challenge is, however, more general than Fedyk lets on. We need not posit different “levels” or types of explanation to generate it. Suppose I hypothesize that Coco intends to steal my cookies because she likes cookies. These hypotheses (about her intentions and her likes, and the causal-explanatory relation between them) are arguably operative at the very same psychological level but they, too, stand in a many-to-many relation. I may be right about Coco’s intentions but wrong about her reasons (maybe she just likes the thrill of stealing), or right about her reasons but wrong about her intentions (maybe she’s planning to ask me for a cookie politely). Theoretical underdetermination like this is simply ubiquitous.

And while Fedyk is right to emphasize that evolutionary scientists have been particularly likely to overlook Mayr’s lemma, the many-to-many problem is not quite as prone to generating fallacious inferences as he suggests. First, Fedyk argues that structural explanations provide literally no evidence whatsoever for the specific psychological mechanisms that might realize them. This is too strong. Given my prior hypothesis, when I gather evidence that Coco likes cookies, that should increase my confidence about her intent to steal them. There is a genuine evidential relation here, even though the evidence is consistent with many other plausible possibilities, such as her planning to ask me for them (my credence in both hypotheses should be raised!). The same is true for the problematic inferences targeted by Fedyk. The operative mistakes are better cast as overestimating a consideration’s unique evidential import rather than seeing evidence where none exists at all. Second, Fedyk argues that researchers posit innate moral modules on literally “nothing more” than evolutionary considerations (34). Now, I’ll be damned before I go to bat for E. O. Wilson, but lumping all, or even most, contemporary evolutionary-moral psychologists together as making fallacious *deductive* inferences about psychological mechanisms *solely* on the basis of adaptive just-so stories is uncharitable in the extreme. As if the experimentalists targeted by Fedyk (namely, those specializing in moral-dilemma vignettes) are not *also* keeping up with, and building, and revising their theories in light of evidence regarding moral development in children, moral behavior in nonhuman animals, and the manifold deliverances of neuroscience from studies using fMRI, lesions, implants, and so on. Minimally, one consideration (over and above the just-so stories) in favor of positing some novel dedicated cognitive module is that there is already independent reason to think that there are other dedicated modules (such as for language-learning).

With Mayr’s lemma and other methodological strictures in hand, the middle chapters turn to a thoroughgoing attack on contemporary moral psy-

chology. Fedyk sets his sights on “adult deontic judgment and decision-making” (ADJDM), paradigmatically represented by studies testing American undergraduates’ trolleyological intuitions. Whereas many philosophers have simply palmed their foreheads in response to headlines advertising that human beings reason more like utilitarians in certain contexts and deontologists in others, Fedyk admirably attempts to systematize the mistakes behind such claims. Experimental trolleyology, he says, has nothing “to do with the most plausible consequentialist and deontological theories in philosophical ethics” (98). His objections here end up being somewhat idiosyncratic, however. Fedyk takes ADJDM to task for lumping all theories of a particular family (such as utilitarianism) together, because there are a variety of different plausible utilitarian theories, which are each constituted by a proprietary set of regulative norms. So it is, according to Fedyk, a full-blown category mistake to suggest that choosing to turn the trolley to kill the one and save the five is an instance of “utilitarian” over “deontological” reasoning; “they could just as meaningfully be called *pattern-a judgments* and *pattern-b judgments*” (98). (Fedyk’s argument also makes much of the fact that none of these normative theories are *realized* in human behavior, that is, there is no persisting social community actually living by all of, say, Kant’s or Bentham’s rules; more on this shortly.)

Although I am sympathetic with Fedyk’s overarching concerns here, the full implications of this argument are quite strong. They would entail that many moral philosophers have also been deeply confused about what they are up to. For example, Judith Jarvis Thomson explicitly frames her treatment of the trolley problem as an argument against *any* ethics of utility-maximization, much as Gettier cases are taken to speak against any justified-true-belief theory of knowledge (Thomson 1985). But if Fedyk is right, Thomson’s characterization represents a profound conceptual error, because, after all, utility-maximization theories make up a diverse family, and many of them might actually prohibit killing one to save five in such circumstances (perhaps variants of rule utilitarianism?). Now, Fedyk might just bite the bullet here and insist that Thomson was fundamentally confused about the *prima facie* normative-theoretic implications of such thought experiments. Alternatively, we might infer that Fedyk is making a somewhat persnickety point that can be addressed with a footnote (#NotAllUtilitarians) and then safely sidelined for many philosophical and psychological purposes. Moral psychologists will likely concede the argument, shrug, and carry on as before. (Listeners of the *Very Bad Wizards* podcast may recall that Dave Pizarro, one of Fedyk’s ADJDM targets, regularly concedes points like these in response to related criticisms from Tamler Sommers. In fact, Pizarro claims that the very point of one of the papers that Fedyk criticizes is to argue that motivated reasoning often prevents people from thinking like utilitarians or deontologists of any sort, that is, to demonstrate that in at least some conditions we don’t tend to reason in keeping with the prescriptions of any plausible normative theory.)

Although Fedyk is mainly oriented toward uncovering moral psychology's foundational flaws, he sprinkles in a variety of less profound criticisms of specific studies. For example, he circles back several times to the criticism that moral psychologists assume that utilitarianism and deontology are "mutually exclusive and mutually incompatible styles of reasoning" (98). This is indeed a regrettable tendency, but it is an odd one to harp on, as Fedyk himself cites an empirical paper that takes issue with the assumption of mutual exclusivity, and puts forward an alternative model to move past it—while still operating squarely within the ADJDM paradigm (Conway and Gawronski 2013). The broader question here is whether Fedyk discounts the extent to which moral psychology and experimental philosophy are, like all empirical communities, self-correcting collective enterprises. It's not always clear how the radical changes that Fedyk ultimately recommends map onto his criticisms of particular inferences and speculative conjectures in particular studies, which can be and in some cases have been addressed without overhauling the field.

But the proposals that Fedyk recommends are worth pursuing in their own right, regardless of how forceful his criticisms of current practice might be. The last third of the book contains its most novel and thought-provoking contributions, as Fedyk develops the "causal theory of ethics," a methodology for empirically identifying ethical behavior. His prescription, roughly, is for social scientists to identify the existing institutions that cause "good and valuable outcomes" (which include health, wealth, and happiness), and then to identify the specific norms and behaviors that help (cause) those social institutions to promote those valuable outcomes. According to Fedyk, norms and behaviors that fit this bill thereby meet a sufficient (but not necessary) condition for being ethical. Moral psychologists' role would then be to find the psychological mechanisms that underlie the ethical behaviors that promote the preferred outcomes. And the final empirical project would be how best to intervene in existing institutions to promote the now-identified ethical behaviors and valuable outcomes. This amounts to an extremely demanding form of naturalism, which asserts not just that "ought" implies "can," but that "ought" implies "is;" we are permitted to identify some existing pattern as ethical once we demonstrate that it leads to good social outcomes in the real world.

Fedyk's methodology has clear resonances with a variety of other approaches across an array of fields, including in applied economics, public policy, and philosophy. We have here a kind of pragmatist, experiment-in-living consequentialism, which departs most significantly from existing views in virtue of its laudable insistence on context-dependence. Fedyk cautions strongly against the assumption that an intervention that works well within one broader set of norms and institutions will achieve the same aims in other contexts. So, for example, maybe norms promoting smaller classroom sizes are "ethical" for Tennessee public schools but unethical for other educational institutions and contexts (166–67).

I am all for the overall project. The world would (eventually) become a much better place if more resources were devoted to this sort of interdisciplinary investigation, and Fedyk helpfully draws on a diverse range of examples from the sciences and the humanities to give the reader a sense of how this all might go. Nevertheless, Fedyk needs to say more about why to cast this project in the idiom of the ethical, as opposed to using any of a variety of other normatively salutary locutions. Perhaps the most significant conceptual obstacle, for me anyway, is that ethics *à la* Fedyk omits any distinctive role for the personal, *directed* character of many (of what we intuitively think of as) ethically relevant behaviors, such as related to promising, discrimination, and so on, when particular individuals do right or wrong by particular others (or themselves), somewhat independently from whether those behaviors promote broader goods.

Fedyk takes care to emphasize that his theory will help to identify only some rather than all ethical behaviors, but his case would certainly be strengthened by considering objections to the effect that certain classes of ethical behaviors are much more likely to be detected while others go neglected. Isn't his framework, basically by definition, biased toward uncovering only those norms and behaviors that lead to good consequences within existing social institutions? This concern is especially pertinent since he earlier takes experimentalists to task for designing studies biased in favor of supporting certain metaethical theories over others (105–110). In the book's penultimate chapter, Fedyk makes some gestures toward accommodating nonconsequentialist approaches, namely, virtue ethics, but even there the proposal is to identify traits as virtues when they both cohere with the norms that benefit the overall community and advance personal goods like happiness. Thus Fedyk-style virtue is when what's good for the greater gander also has positive externalities for the goose. But the more pressing question is where nonconsequentialist theories focused on rights and duties fit into the picture. Although Fedyk is clearly aiming for an ecumenical and inclusive methodology, many nonconsequentialists will object that their normative theories are being stipulated right off the stage. Why not opt for a broader theory that accommodates duty-based or interpersonal behaviors that treat people with respect and good will and are therefore ethical even if they do little to directly advance the common good? Perhaps Fedyk is concerned about how to empirically identify such rights and duties. But his all-out attack on ADJDM might lead him to discount the potential role that people's explicit judgments and overt reports could play in determining what's ethical and what's not. For example, nonideal theorists have long argued that we already have at least one tried and true empirical method for identifying rights: listen to oppressed people protesting that their rights are being violated.

## References

Conway, Paul, and Bertram Gawronski. 2013. "Deontological and Utilitarian Inclinations in Moral Decision Making: A Process Dissociation Approach." *Journal of Personality and Social Psychology* 104: 216–35.

Thomson, Judith Jarvis. 1985. "The Trolley Problem." *Yale Law Journal* 94: 1395–1415.

*Alex Madva*

California State Polytechnic University, Pomona

*Philosophical Review*, Vol. 128, No. 1, 2019

DOI 10.1215/00318108-7213385

Matti Eklund, *Choosing Normative Concepts*.

Oxford: Oxford University Press, 2017. 219 pp.

As Eklund might be the first to tell you, this book is not systematic advocacy of a particular position. As I might be the first to tell you, it is an extremely rewarding exploration of a number of issues central to metanormative debates.

At the heart of the book lies a metanormative theorist, which Eklund dubs *the ardent realist*, and their discomfort with a hypothetical case,<sup>1</sup> which Eklund dubs *Alternative*. The case goes like this (18): (i) there are two different linguistic communities, A and B, where A has the thin, all-things-considered, normative word 'ought' and B has the thin, all-things-considered, normative word 'ought\*'; (ii) 'ought' and 'ought\*' have the same normative role (as a matter of conventional semantics); (iii) 'ought' and 'ought\*' are not coextensive (so they do not ascribe the same property); and (iv) community A's claims about what ought to be done are true assertions about the ought facts, while community B's claims about what ought\* to be done are true assertions about the ought\* facts.

The setup here is similar to the moral twin earth cases from Horgan and Timmons, but whereas they were more interested to cause trouble for semantic and metasemantic theories offered by naturalist realists, Eklund wants to cause trouble for the ardent realist (32–37). We don't get a full characterization of ardent realism, but we do know that they want to deny that *Alternative* is possible. For if *Alternative* is possible, then there is no clear sense in which one community's normative term is privileged (I am tempted to add Sidgwick's phrase "from the point of view of the universe") compared to the other. That

1. I use the singular 'they' and cognates.