

REFLECTIVISM, SKEPTICISM, AND VALUES
forthcoming in *Social Theory and Practice* (2018)
Manuel R. Vargas
University of California, San Diego

John Doris's *Talking to Our Selves* is a gripping challenge to any view according to which reflective self-awareness is at the core of our agency. With enviable flair, Doris explores the philosophical implications of a wide range of findings from the behavioral and cognitive sciences and shows how they impugn the reliability of our reflective, self-aware agency. He argues that individual agency is best understood as a product not of accurate, reflective rationality, but instead, as a function of social and collaborative practices. The overarching aspiration is recast humans "not as little gods with big brains, but as animals that, alongside other animals, have evolved with a curious assortment of endowments for muddling through this world" (199).

The centerpiece of the book is a critique of *reflectivism* and Doris' proposal of an alternative. By *reflectivism*, Doris means the psychological doctrine "according to which *the exercise of human agency consists in judgment and behavior ordered by self-conscious reflection about what to think and do*" (x). The view has a traditional corollary, that *the exercise of human agency requires accurate reflection*. Doris argues that *reflectivism* and the accuracy corollary are false: agency does not require self-conscious reflection about what to think and do, and the exercise of human agency does not require accurate reflection.

In what follows, I focus on three ideas: *reflectivism*, the argument for skepticism, and the conception of values operative in Doris' account. Section one takes up the question of whether there is a more charitable way to construe the *reflectivist* project. I argue that there is a better way to understand the *reflectivist* position, one that makes *reflectivism* less vulnerable to Doris' criticisms. Section two explores some doubts about the force and need for the skeptical argument Doris raises against *reflectivists*. In section three, I conclude by asking whether collaborative reasoning in the way Doris emphasizes can deliver the sort of values that might plausibly underpin the kind of agency at stake in debates about responsibility.

I. REFLECTIVISMS

Doris understands *reflectivism* as the psychological doctrine "according to which *the exercise of human agency consists in judgment and behavior ordered by self-conscious reflection about what to think and do*." (x). The view also invokes an accuracy requirement—i.e., that one's reflection on beliefs, desires, and so on, needs to be accurate.

Although Doris can stipulate *reflectivism* to mean whatever he likes—it is his book and his term, after

all—I wonder if the philosophical syndrome Doris associates with reflectivism is better construed as evidence for a predominantly normative, less psychological thesis. Call the following thesis NR, or *normative reflectivism*:

NR: An ideal form of human agency consists in judgment and behavior ordered by accurate self-conscious reflection about what to think or do.

What NR insists on is the *ideality* of accurately reflective agency. It holds that accurately reflective agency is an (or perhaps “the”) appropriate governing aspiration or standard for various cognitive and affective practices. Normative reflectivism is, of course, compatible with psychological reflectivism. And, as a matter of the actual commitments of various reflectivists, many seem to have been psychological reflectivists—be they of Kantian, Platonic, or Cartesian provenance. Importantly, for the normative reflectivist, psychological reflectivism is a further and ultimately secondary thesis. What is central to normative reflectivism is the claim that there is something privileged about accurate self-conscious reflective agency. Whether that privilege takes the form of maintaining that such agency has special normative heft (“oughty-ness” as Doris says) or whether it is better understood as regulative ideal—a task or terminus for an activity—might vary by account.

Normative reflectivism is consistent with the ambitions of various paradigmatic reflectivists, and it provides a charitable way of reading reflectivist ambitions. For example, it is reasonable to understand the Delphic injunction to know oneself as aspirational, as specifying a goal of agency, even if we also acknowledge that we will always remain substantially or unavoidably mysterious to ourselves. Consider, too, the commonplace among some interpretations of Kant’s *Groundwork* that it is an open question whether or not anyone has ever acted in accord with the moral law. Finally, as Doris knows better than most, a familiar move in virtue theoretic traditions is to avert to the relative rarity of the sage or the person of genuine practical wisdom. On each of these approaches, reflectivism is first and foremost an ideal and only secondarily (if even) a psychological hypothesis.

What might Doris say in response? He considers the possibility of reflectivism as a kind of conceptual thesis, but argues that it is “false, or at least dangerously overstated” (22). However, the form of conceptual reflectivism he seems to have in mind is one whereby agency *requires* reflection. **His strategy is to argue against the conceptual thesis on the basis of showing that exercising agency does not require reflection** (23). There are two difficulties for construing this reply as an adequate response to normative reflectivism as I have characterized it.

First, Doris construes the conceptual reflectivist as holding that agency requires accurate self-reflection. Notice, though, that normative reflectivism maintains only that *ideal* agency requires accurate self-reflection. Nonideal, or imperfectly ideal forms of agency might still be possible in the absence of accurate self-reflection.

Even if the ideals of Socratic wisdom, Kantian good will, or virtuous action require something we rarely or never have, we can better and worse approximate those ideals. Recognizable forms of agency lurk in the shadows of more unattainable ideals. Even where ideality is impossible or unlikely, other (presumably lesser) forms of agency may prove sufficient for philosophical, moral, and ethical evaluation. Unwise, self-interested, and unvirtuous agency is still agency—and typically the basis of our moral responsibility assessments.

Second, Doris' response to the conceptual reflectivist sits uneasily with Doris' pluralism about responsible agency. Doris holds that there are multiple ways to achieve the form of agency characteristic of moral responsibility (12, 175, 197). So, his positive view (discussed below) of dialogic, collaborative agency is proposed as one among several potential ways of attaining responsible agency. Given this pluralism, and given the formulation of normative reflectivism as an ideal of agency, it is unclear why there needs to be a conflict. That is, why not a pluralism that allows that some ideal forms of agency might invoke accuracy?

Doris could object that I've stacked the deck. Normative reflectivists, he might say, are committed to their ideal form of agency being *the* ideal form of agency. If so, it is a cheat to portray normative reflectivism as compatible with pluralism about agency. However, even if we strengthen NR to stipulate that it is *the* ideal form of human agency, Doris' pluralism is not obviously problematic for the antecedently committed normative reflectivist. Again, normative reflectivism is compatible with a diversity of non-ideal forms of agency. By the normative reflectivist's lights, Doris' positive proposal about agency might be an account of an important form of "good enough, but not ideal" agency.

To be sure, the normative reflectivist will still face the challenge at the core of Doris' book. Indeed, I suspect that the normative reflectivist will ultimately end up where Doris wants her, i.e., forced into a reconsideration of what views seem attractive once we acknowledge the unhappy fate of psychological reflectivism. However, if I am right that the animating thesis for reflectivists is reflectivism as an ideal or privileged form of agency, and not a psychological hypothesis, then these differences make a difference.

Consider the issue of morally responsible agency, or what Doris more simply calls *agency*. If you accept some version of normative reflectivism, you could accept Doris' case against psychological reflectivism, yet still insist that what would make someone *fully* morally responsible is their acting out of an accurately reflective agency. Such a view is compatible with thinking that the degrees of responsibility that are left—or that the flavors of blameworthiness rightly available to us—are nevertheless adequate for some core of the practice with which we are concerned.

An analogy is suggested by the case of romantic love. Suppose the complete acceptance of all aspects of a person's history and personality is an ideal for romantic love. The difficulty of achieving total acceptance of this sort in a human relationship does not seem to cut against its suitability as an ideal that structure our practices of romantic love. However, suboptimal love might not be scarce at all. Indeed, love's suitability as an

aim in some of our ordinary practices might well depend, in part, on suboptimal love's reasonably good availability. Yet love's normative appeal might still rest on its approximation to a regulative ideal for what acceptance in love should look like. So it goes for responsibility, says the normative reflectivist.

Doris makes three objections to the sort of view I've been offering, i.e., that accurate, self-aware rational agency is an ideal form of responsible agency.

First, he notes that (1) it is hard to show that one form of agency is better than another (as the normative reflectivist holds), and (2) it unclear why the pluralist should want to accept such a view (151). Fair enough. But agency with accuracy seems like something worth wanting, and in some recognizable sense, preferable if we can get it. Moreover, nothing in this thought is incompatible with thinking other forms of agency might also suffice to support our practices as we find them.

Second, Doris argues that his positive picture of socially embedded agency has a special normative appeal that alternatives—such as reasons-responsiveness accounts—lack. He does this by imagining an agent, Duke, who is in the space of reasons but is asocial. Doris argues that Duke is missing something, arguing that “he would be less richly an agent than chattier sorts, because many exercises of agency that make life worth living emerge in dialog, as we negotiate our selves with our intimates” (152).

Duke is missing something, but either what he is missing doesn't matter for responsible agency, or if it does, then Duke has it. That is, it is surely true that compared to richly dialogic agents, Duke comes up short. But recall that the kind of agency at stake is morally responsible agency. By stipulation, Duke already has access to all the reasons that matter to responsibility. If so, then whatever his dialogic shortcomings, he is not a less normatively attractive agent *with respect to responsible agency*. We might not want to party with Duke, but he navigates moral life as well as we can expect.

Of course, if Duke doesn't have access to all the reasons that matter for responsible agency, then we have an explanation of *why* Duke seems to be missing something. That is, Duke is missing the dialogic pathways necessary for tracking the considerations relevant for responsible agency. If that's the view, though, then notice that Duke does not count against a normatively ideal picture of agency that emphasizes the role of reasons responsiveness. In this case, the normative lifting is being done by the underlying fact of Duke's reasons-responsiveness, and the dialogic element is just a vehicle for it. One way to see this is to consider the affable, highly dialogic person who does very badly at recognizing and responding to moral considerations. Such a person might be the life of the party, but hardly someone we can count on when there are serious moral stakes in play.

Third, Doris concludes that where an account's normative appeal is arguable, it can't be used to decide among competitors. The thought is that reflectivism isn't, in the final analysis, an appealing view if its normative appeal is subject to dispute. In the context of evaluating normative reflectivism as an alternative

interpretive proposal of the reflectivist, this seems like a non-sequitur. Normative reflectivism is supposed to be charitable reconstruction of the view that propels Doris' opponents. Inasmuch as the first half of the book is supposed to be raising problems for the reflectivist, it isn't clear that the arguability of reflectivism's normative appeal cuts ice against the interpretive proposal I've offered. After all, what made reflectivism an interesting and important target was that it was supposed to be a kind of widespread, maybe even default view among philosophers. I don't see why the fact that reflectivism's normative appeal is arguable precludes normative reflectivism's putatively greater normative appeal for those drawn to reflectivism. So long as we see the normative reflectivist as appealing to people who think there is something valuable about responding to in the right way to the right reasons—a big group, plausibly—then the normative appeal of it seems both plausible on its own terms and an adequate basis for interpreting reflectivists as motivated by normative reflectivism.

2. SKEPTICISMS

I now turn to consider the skeptical challenge that animates the first half of the book, and argue that despite its centrality in the narrative, it is unconvincing against the strongest version of a plausible competitor view, and anyway, unnecessary.

Here's how Doris puts the skeptical challenge for the reflectivist: "Where the causes of her cognition or behavior would not be recognized by the actor as reasons for that cognition or behavior, were she aware of these causes at the time of performance, these causes are *defeaters*" (64-5). Doris' thought is that in any given case of putatively responsible agency, the involved bit of behavior might be subject to defeaters. Since defeaters can't be confidently ruled out in that case, the reflectivist is saddled with a picture chronically subject to skepticism. Doris' recommendation for overcoming the skeptical challenge is to jettison reflectivism, and in particular, the accuracy corollary. As we saw above, if we give up the idea that one's agency requires accurate reflection, the mere presence of defeaters—i.e., causes one would not recognize as reasons for a cognition or behavior—is, by itself, of no consequence.

I pursue three lines of thought against Doris' skepticism. First, I argue that skepticism is unnecessary for animating Doris' collaborativism. Second, I raise some objections to the skeptical argument itself, at least as it plays out in the dialectic with a normative reflectivist. Third, I maintain that the argument for skepticism fits uncomfortably with his pluralism. As we'll see, an important wrinkle to Doris' skeptical argument concerns the kind of certainty that matters. He writes that, "The requirement that defeaters be 'confidently'—not conclusively—eliminated can be understood as a demand for a kind of moral confidence: the conviction that, were one's initial judgment that an exercise of morally responsible agency obtained (and defeaters did not) overturned in light of new evidence, one would not be guilty of wrongdoing for having apportioned burdens and benefits according to this judgment" (66). At the core of much of what follows is the thought that we can

have a suitable level of moral confidence in our judgment that the sources of an action are not entirely populated by defeaters.

First, does Doris need the skeptical challenge to go through? I can't see that he does. Consider the picture sketched in the preceding section, i.e., reflectivism as an ideal, where common but less-than-ideal forms of reflectivism are the stuff of everyday life. In light of this possibility, it is notable that a variety of accounts of moral responsibility are prepared to acknowledge that the forms of agency and responsibility we find ourselves with in the sublunary realm are somewhat less than the most demanding forms of agency and responsibility to which we might aspire. Quite apart from the bookkeeping question of how many accounts are aptly characterized as committed to normative reflectivism, the basic normative pressure—i.e., can we make do, normatively speaking, with the suboptimal forms of agency that do exist?—is undeniably a central issue for theories of moral responsibility, *even if we grant that accurate reflective agency is relatively uncommon*.

If this is right, then Doris doesn't need the looming threat of skepticism to animate his dialogic, collaborativist proposal. Indeed, with respect to the larger ambitions of Doris' project, it is unclear why it is not enough to take seriously the various empirical issues by asking whether we can provide an account of responsibility that can accommodate them in a satisfying way. For example, suppose I am right that the troubles of psychological reflectivism will leave untroubled the idealized heart of normative reflectivism. Even so, Doris has raised an important challenge that merits attention. The challenge for reflectivists—and really, anyone wishes to show that in the ordinary course of things we have moral responsibility, at least sometimes—is to explain the possibility of agency and responsibility given the fact of potentially widespread “bypassing effects,” i.e., incongruence between the causes and reasons for behavior and cognition.

Recasting the skeptical argument as a more open-ended challenge to accommodate a naturalistic picture of agency is not just window-dressing. It changes the stakes in interesting ways, and in particular, in ways that are going to be amenable to normative reflectivists. Consider how a normative reflectivist might reply to the empirical challenge to show how we have some form of agency suitable for supporting responsibility practices and ascriptions of selves and responsible agency.

Perhaps the most promising strategy might be to emphasize the ideality of accurate reflective agency, but to concede that in ordinary life, the most we can hope for is some (relatively low) threshold of responsiveness to reasons. Such a theorist might yet hold that we are rational enough (e.g., sometimes in conscious, deliberate ways, sometimes not) to make sense of responsibility and agency if, for example, in the ordinary case agents would (given full information) see themselves as responding to *reason enough* for acting the way in which they did.¹ On such an approach, the challenge from naturalism is not that an agent is moved

¹ I take it that at least some (mainly reasons responsiveness-style) approaches to responsible agency have pursued something like this line. Arpaly's 2003 account of moral worth allows that considerations need not be thought of as moral to be to the credit of the agent.

by some causes she would not recognize as reasons. Instead, for the challenge to have teeth, it must be that there is *no* relevant cause in the etiology of the action that the agent would recognize as a reason, or as sufficient for the cognition or behavior under consideration. The most promising version of this view may need to reject the idea that there is a single cause for a given action—as opposed to a web of causes. It would also likely hold that reasons sensitivity can emerge at various levels of causal contribution, e.g., both at the level of conscious decision-making, but also, perhaps, in dispositions to respond to the world, to be moved by non-conscious motives that stand in the right relationship to morally worthwhile features of the world, and in the configuration of various perceptual affordances. However the details go, it seems clear that showing that there are no suitable causes-as-sufficient-reasons-for-action in a wide range of cases will not be a straightforward business. You certainly don't get that by showing, as Doris successfully does, that there are *some* non-rational effects on decision-making (as in the case of affect confusion, the naming effect, and so on).

All of this is compatible with thinking that many of the causes involved in a given choice are not relevant to the agent's rationality. If so, then it is frequently going to be an open and difficult-to-answer question whether we have enough instances of causes-that-count-as-reasons to do the work required by non-ideal forms of agency. Here, though, the *frequency* of defeaters seems to matter for the question of whether we should ordinarily suppose that there is enough rational guidance for our practices of blaming to remain intact.² So, on this way of recasting the stakes, questions about how much bypassing occurs, and in what situations—questions Doris mostly endeavors to sidestep—are central to whether the skeptical argument succeeds. Absent some argument about the pervasiveness of defeaters, it is hard to see how we could decide that there is insufficient rational responsiveness (or what have you) to support practices and ascriptions of responsibility and agency like ours.

In this context, it is perhaps worth revisiting at least one argument Doris offers for the more demanding skeptical version of the naturalistic challenge. Consider these remarks:

“Once we see that there are some arbitrary influences on cognition and behavior, we are bound to admit that there may be others; if something like *that* can make a difference, there could be *many* goofy influences in any particular instance. While the impact of each individual goofy influence may be statistically small, just as with medical interventions, the aggregate effect

My own implementation of the reasons-responsiveness idea does not require that an agent even be aware of being moved by a reason to count as being suitably responsive to moral considerations (2013: 337). On a variety of views, just because something is a product of automatic, non-conscious causes does not automatically mean it is insufficiently rational to meet some standard of responsible agency. That said, it strikes me as interesting and under-explored question the extent to which agents are rightly counted as sensitive to or rational with respect to (ecologically bounded) reasons even when there is bypassing going on at the level of particular mechanisms.
² Doris' consideration of the “triage” and “mediation” reflectivist responses are surely relevant (71-76), but those presume psychological reflectivism, and not the non-ideal forms of rational responsiveness that would appeal to the normative reflectivist.

may be quite potent; for all one knows, any decision may be infested by any number of rationally and ethically arbitrary influences. Now, there's a large, odorous, and ill-tempered animal under the awning of agency" (Doris 2015: 64).

Now consider the following parallel of that argument:

"Once we see that there are some rational influences on cognition and behavior, we are bound to admit that there may be others; if something like *that* can make a difference, there could be *many* not-so-goofy influences in any particular instance. While the impact of each individually not-so-goofy influence may be statistically small, just as with medical interventions, the aggregate effect may be quite potent; for all one knows, any decision may be infested by any number of rationally and ethically relevant influences. Now, there's a large, appealing, and well-tempered animal under the awning of skepticism"

If we accept the looser naturalistic challenge to merely accommodate the empirical facts as we find them—and if we suppose that the normative reflectivist can live with non-ideal forms of agency so long as they achieve some threshold of sensitivity to reasons—then an important idea emerges. Whether we have adequate moral certainty about the presence of a sufficient reason, or alternately, the presence of only defeaters, is a matter of *both* the empirical data *and* the nitty-gritty of a given theory's story about responsibility. If we set the bar of rational responsiveness sufficiently low—or if one is prepared to be permissive about the standards for fallibilist blaming, i.e., holding that it takes relatively little warrant to defeasibly blame someone—then the moral certainty required to hold people responsible is relatively easily met. However, as always, the tenability of the account is also hostage to the empirical facts. Lax standards for agency can be regularly defeated if defeaters are sufficiently ubiquitous.

Stepping back, I'm inclined to think that Doris' permissiveness about what counts as responsible agency (i.e., his afore-mentioned pluralism) rests uneasily with the rejection of reflectivism he announces at the outset of the book. This seems especially so when we focus on the skeptical argument. That is, given an antecedent openness to pluralism, it is unclear what turns on defeating all possible forms of reflectivism. What does matter is that one can offer a scientifically plausible picture of agency that makes sense of responsibility, agency, and selves, without requiring accurate reflection. The scientific image would remain, for all parties, something that must be accommodated, and if reflectivists can do so, then this is no threat to one's pluralism. In meeting the scientific challenge, though, Doris has a leg up. He's got a rich and interesting account of collaborationist, dialogic agency that comports pretty well with the going empirical data. In contrast, most

theories of moral responsibility help themselves to a moral psychology blissfully untainted by empirical considerations. If I'm right that the skeptical arguments won't move the strongest form of reflectivist—the normative reflectivist— and if I'm right that the skeptical argument isn't required to see the appeal of collaborativism, then Doris doesn't need the skeptical argument to succeed.

3. VALUES AND DIALOGICAL COLLABORATIVISM

In this section, I focus on some details of Doris' positive account, and in particular, the nature and role of values. Doris holds that we can anchor agency in instances of action that are structured by values. He accepts a broadly Bratmanian account of values as, "the desires properly associated with value are those desires the actor accepts in a determinative role for her practical planning," (27). On this account, values must also have a justificatory role—they must be something that the planner is amenable to employing in justification or defense of her plan (27).

Given the cognitive and behavioral sciences, self-conscious, accurate reflection on one's reasons seems elusive. Thus, one advantage of the valuational approach is that it doesn't require anything like that. Still, one might wonder, how do we get value-structured action? Doris' account leans on the dialogic and social character of what we say about ourselves. When an agent acts and is asked to provide an account of what he or she has done, the agent offers a rationalization—a performance that makes sense of judgment and behavior (141). Oftentimes, the rationalization is a local bit of biography or narrative about one's action. And, if Doris is right, it oftentimes amounts to a kind of confabulation. However, there is something special in the social significance of the rationalization, even when it is confabulatory. Agency emerges from—or is perhaps constituted by—the dynamic and socially evaluated nature of these rationalizations.

Rationalizations can't be offered willy-nilly, at least not if they are to be successful in navigating social life. They are constrained by the willingness of others to accept the rationalization, the expectation that one's behavior will be in accord with one's proffered rationalizations, and the presupposition that such rationalizations are a mutually recognized basis for how we go forward together in shared, cooperative contexts (147-8). These constraints on rationalization structure our interactions with others in forward- and backward-looking ways. Over time, these patterns of rationalization cause people to shape their lives in ways that reflect or express their values, Doris claims (153).

This is an original and intriguing approach to responsible agency and selfhood. And, I think it rightly points our attention to the rich resources involved in the social dimensions of our moral and agential practices. Doris is surely right that under a variety of conditions, we do better reasoning together, and that dialogic processes may do a better job of revealing and expressing our values. I wonder, though, about whether rationalizations can do the work Doris sets for them.

In particular, why think rationalizations express values? Why aren't rationalizations just (oftentimes purely local, one-off) culturally defensible action-explanations with no connection to an agent's values?³ Proffering a (sometimes scripted, other times improvised) story that deflects attention or concern does not seem like the right kind of candidate for value-expression, at least not as we ordinarily understand it. If there is something to this worry, we might also worry about how often action reflects values. If the answer is "sometimes," this doesn't feel like enough to overcome skepticism about agency. It certainly doesn't make the collaborativist picture look more promising than a suitably sophisticated reflectivist story about agency (see below).

As I read his account, there are suggestions that what Doris has in mind by values and their role in the collaborativist picture of agency is actually a fairly radical recasting of what it is to act from or in keeping with values. That is, there are hints that he thinks of an agent's values as constructed, constituted, or determined *external* to the agent (this is one way of reading the discussion of bias on p. 146, I think). If so, that's a pretty radical departure from standard understandings of what values are, and how they are an agent's. If that's the view, one might worry that it abandons one of the most appealing features of a valuational account. That is, on one way of understanding valuational accounts of agency, the appeal to values is supposed to provide a principled way to identify when actions are produced by agents, as opposed to the world or some non-agentive psychological factor. If Doris means to argue that valuing is constructed external to the agent in some relatively meaty sense, it becomes harder to see how valuing understood in this way gets us a sufficiently robust basis for agency, as opposed to yet another story about why what seems like agency is really the product of ideology, circumstances, or forces regarded as alien or external to the agent.

So, I'd love to hear more about why we should think rationalizations express values, and if not, whether the picture is a broadly constructivist and/or externalist story about values being importantly outside the agent's head. If constructivism/externalism is the picture—and I'm unsure whether it is—then I want to know why values of that sort count as the agent's values, as opposed to a story about why responsible agency is a kind of sham.

In keeping with disciplinary expectations, my focus in this and the preceding sections has been on points of disagreement. I conclude by noting that these disagreements mask what is otherwise a robust agreement with the motivations and aspirations of *Talking to Our Selves*. I agree with Doris that reflectivism in some or another form structures philosophical theorizing. I concur that experience does badly at securing the

³ I'm not sure if the following is a concern about values so defined, or about rationalizations. Suppose we take seriously the core lessons of situationist social psychology (e.g., Doris 2002). If so, then why shouldn't we think that many of rationalizations are not expressions of values, but instead, expressions of local situational pressures (Cf. Vargas 2018)? I would have thought values need to have some cross-situational stability of the sort that is threatened by the relatively local nature of rationalizations, which will often be situationally specific responses to contextual pressures. If so, then why isn't reading off values from rationalizations akin to reading off character traits from behavior? I take it that Nelkin (this volume) has a similar concern.

goods suggested by a reflectivist picture of agency. I'm open to at least some forms of pluralism about responsible agency and moral responsibility. And, I'm inclined to accept a picture of responsibility that, if not specifically collaborativist, shares much of collaborativism's emphasis on the socially scaffolded nature of moral responsibility. Thus, my suspicion is that our disagreements, such as they are, are less about the tenability of the project and more about the particular shape required by an adequate rehabilitation of our understanding of responsible agency. Here, as elsewhere, collaborative and dialogical efforts are the most promising way forward.⁴

⁴ Thanks to the audience and participants of the 2016 Pacific APA Book Symposium where these comments were first presented. Thanks, too, to the members of my 2016 seminar at U.C. Berkeley for discussion about many of the ideas in this paper.

References

- Arpaly, Nomy. 2003. *Unprincipled Virtue*. New York: Oxford.
- Doris, John. 2002. *Lack of Character*. New York: Cambridge University Press.
- Doris, John. 2015. *Talking to Our Selves: Reflection, Ignorance, and Agency*. New York: Oxford University Press.
- Vargas, Manuel. 2013. "Situationism and Moral Responsibility: Free Will in Fragments." In *Decomposing the Will*, edited by Till Vierkant, Julian Kiverstein, and Andy Clark, 325–49. New York: Oxford University Press.
- Vargas, Manuel. 2018. "The Social Constitution of Agency and Responsibility: Oppression, Politics, and Moral Ecology." In *The Social Dimensions of Responsibility*, edited by Marina Oshana, Katrina Hutchinson, and Catriona Mackenzie, 110–135. New York: Oxford University Press.