On John Doris’s *Talking to Our Selves*

*John Martin Fischer*

I am a big admirer of John Doris’s first book, *Lack of Character* (Doris 2005). It is beautifully written, and it highlights the importance, for philosophical theorizing about agency, of a broad range of empirical results. The main thrust of the first book was to call into question “virtue” approaches in ethics and moral responsibility theory, which presuppose that we often act from stable character traits. Instead, Doris defends “situationism.” Part of his argument was that the empirical results show (or at least strongly suggest) that we are often (or perhaps typically) unaware of the factors that actually lead to our behavior. The book has been highly influential, and it helped to start a “movement” in philosophy.

The new book, *Talking to Our Selves* (Doris 2015), in part builds on these results and further develops (on their basis) a skeptical challenge for certain views about moral responsibility. More specifically, he identifies a “standard” account of moral responsibility, “reflectivism,” which, he argues, is subject to skepticism. He defines “reflectivism” as the view that “the exercise of human agency consists in judgment and behavior ordered by self-conscious reflection about what to think and do” (Doris 2015: x). He also claims that it typically comes with a corollary: “the exercise of human agency requires accurate reflection” (ibid.). He summarizes the crucial points as follows: “In an exercise of agency, as construed by reflectivism, a person correctly divines the beliefs, desires, and other psychological states relevant to her decision, makes her decision in light of these states (sometimes called reasons), and acts accordingly.”

The skeptical worry that Doris develops in the new book contends that insofar as reflectivism cannot rule out the (widespread) existence of “defeaters,” the attribution of agency and thus moral responsibility is not warranted. A defeater is a case of “incongruence,” where the agent would not count the causes of her behavior and cognition as reasons.

But the book is not entirely (or, in my view, fundamentally) “negative” in its orientation. Doris argues that agency, or at least some instances of agency (as he eventually adopts a kind of pluralism about agency and also responsibility), consists in the expression of the actor’s values. In this way, he follows Gary Watson in understanding moral responsibility as a matter of the expres-
sion of values. But Doris offers a bold and original account of the expression of values. He calls it “dialogic.” On the dialogic approach, an individual embedded in a social context negotiates the rationalizations by which he or she explains and justifies his or her judgment and behavior. As he puts it: “Human beings living in groups shape their lives, not as isolated reflectors, but as participants in an ongoing negotiation—a negotiation that simultaneously constrains and expresses who they are. In a slogan, agents are negotiations. If you like, call this notion of agency dialogic” (Doris 2015: 148). Crucially, Doris argues that agents can express their values in the required way, even if they are unaware of doing so. That is, the rationalizations need not be “historically accurate,” and thus his account of agency is compatible with significant self-ignorance (Doris 2015: 196). And one need not be aware of one’s values in order for one’s conduct to express those values in the way characteristic of agency. Thus, Doris claims that the skeptical worries that defeat reflectivism need not defeat his values approach.

This is a big, ambitious, nicely-written, and (as one would have expected) highly intelligent book. I have a learned a great deal from reading it and thinking about it in a preliminary way, and I have no doubt that I will continue to benefit from further reflection. I found the honest and sometimes self-deprecating tone appealing, and I enjoyed John’s inexpressible sense of humor.

There is at least one salient way (no doubt there are others) in which John stands in start contrast with Peter Strawson. In his classic and deeply influential essay “Freedom and Resentment,” Strawson wrote of the “optimist” (i.e., compatibilist): “Well, people often decide to do things, really intend to do what they do, know just what they’re doing in doing it: the reasons they think they have for doing what the do, often really are their reasons and not their rationalizations. These facts, too, are among the facts as we know them” (Strawson 2013: 65). Here Strawson articulates the conventional philosophical wisdom, with which Doris takes issue. But elsewhere in the article Strawson argues for the “quality of will” approach to moral responsibility, without any “metaphysical” requirement of freedom or control. On this approach, our reactive attitudes are keyed to the quality of will shown by others to us, but there is no requirement to justify the reactive attitudes by reference to a “metaphysical” proposition to the effect that the relevant individuals have freedom or control. As I read Doris, this is his view as well, although he supplements (or, perhaps, fills in) the view with the dialogical theory.

The problem, in my view, is that one’s quality of will can be manipulatively implanted in such a way that we would not deem the individual morally responsible. If your ill will or indifference to me is the result of clandestine manipulation (or similar processes) to which you have not consented, I don’t think that my resentment toward you would be justified. My reactive attitudes
are not just keyed to your attitudes toward me, but how those attitudes developed. Moral responsibility is, in my view, essentially historical. I believe that Strawson’s quality-of-will approach needs to be supplemented with a historical condition. (I—and my co-author—develop a historical approach in Fischer and Ravizza 1998, and I discuss its relationship to Strawson’s approach in Fischer 2014.) Similarly, I worry that Doris’s expression of values approach needs to be supplemented by a historical condition that captures the importance of control or freedom in the development of the attitudes in question. Now it might be that the dialogical theory, or some other element of Doris’s approach, has the resources to address the worries stemming from the possibility of responsibility-undermining methods of inculcation of attitudes. But I’m not sure, and I’d like to know a bit more about how John would address such worries.

It might be helpful (and, I hope, not excessively self-indulgent) to offer (in very brief and sketchy form) my approach to injecting a control requirement into the process that leads to the expression of the relevant kinds of attitudes—the kinds of attitudes (and behavior) to which our reactive attitudes are responses. I argue that guidance control is required for moral responsibility. On my view, an agent exhibits guidance control in her behavior insofar as it issues from her own, appropriately reasons-responsive mechanism. Regulative control—freedom to do otherwise—is not necessary for guidance control (and thus, for moral responsibility), and guidance control is compatible with causal determination and also with indeterminism (of the relevant sorts).

The sort of control or freedom I posit avoids many of the puzzles and problems that plague compatibilist approaches that embrace regulative control (genuine metaphysical access to alternative possibilities) and certain libertarian accounts, and it does not imply that our moral responsibility and personhood “hang on a thread” (in the sense that they depend on certain deliverances of theoretical physics). Although I agree with John that certain debates about the relationship between causal determination and freedom to do otherwise may well issue in Dialectical Stalemates (Fischer 1994), I have suggested that we can sidestep these stalemates by positing only guidance control, and not also regulative control, as necessary for moral responsibility. And, although I concede that it is contentious whether guidance control is the freedom-relevant condition for moral responsibility, I believe that a plausibility argument can be made for this claim. I think this style of compatibilism is on firmer footing than classical compatibilism (according to which we require regulative control for moral responsibility, and such control is consistent with causal determination). So I am not as pessimistic as John about our options, given the Dialectical Stalemates in the literature about the relationship between causal determinism and freedom to do otherwise. There is, in my view, a robust, important, and
non-mysterious kind of freedom (that goes back in philosophical spacetime to Chrysippus, Spinoza, Locke, and, more recently, Frankfurt) that doesn't require freedom to do otherwise, and positing this sort of freedom as a requirement for moral responsibility seems to be necessary (at least in my view) in order to address manipulation (and related) worries.

But my account of the relevant kind of control invokes reasons-responsiveness. How does the huge skeptical literature about our abilities to identify and act on reasons relate to my account? This has intrigued and vexed me for a long time. As John takes a certain manifest delight in highlighting, the results from the social sciences suggest that we are not good at identifying the actual motivational states that result in our behavior; there is often, as John puts it, a problematic "incongruence" between what we believe is motivating us and what actually is. If we are particularly bad at identifying and understanding the real causes of our behavior—if we think we are acting for certain reasons, but in fact we are simply being driven by our instincts, emotions, or other contextual features that we would not take to justify our actions—then how could guidance control provide the account of the freedom required for moral responsibility?

As I wrote just above, I have found this question disturbing and vexational. With the emerging difficulties in replication of results in the social sciences, and, in particular, psychology—and these difficulties pertain precisely to some of the results relevant to reasons-responsiveness, such as the "priming" results—I had hoped I could finally just put this question aside and give it the benign neglect I desperately hope it deserved. But John's thoughtful discussion of the situation in the wake of the replication difficulties gives me pause. He argues that, despite the undeniable fact that many of the results have not proved to be resilient and replicable, there is still enough consensus to have considerable doubts about our abilities to identify the real causes of our behavior. This is a deep challenge for a view like mine.

I do want to clarify my view a bit, however. I want to emphasize that my sort of view is completely compatible with our often being wrong about what is actually leading to our actions; that is, it is compatible with a good deal of incongruence (where this incongruence would not crowd out moral responsibility). It is just obvious that we human beings are not perfect at identifying the causes of our behavior—we are often deceived, unaware, or just wrong about this. I want an account of moral responsibility that leaves room for a good deal of incongruence; we are still morally responsible, in a broad range of contexts, in which we are mistaken about our "real motivations," in my view. But I do think that it is at least required that we be capable of responding appropriately to the reasons that exist in the context of our behavior. More carefully, on my view, an agent is morally responsible for her behavior only
if she acts from (her own) mechanism, a mechanism that has the capacity to respond appropriately to the reason that actually exists in the situation. This is what explains why the reactive attitudes are justified, when targeted to a morally responsible agent: she is acting from a mechanism that she owns and that has the capacity to respond appropriately to the reason she actually has in the context of evaluation. Note, as above, that a manipulated agent might not deserve to be the target of (say) resentment or indignation, even though he manifests ill will; quality of will is not enough to justify the relevant reactive attitudes, insofar as mental states with the relevant quality can be induced in intuitively responsibility-undermining ways.

But now perhaps an important question emerges for my approach. What if the social science literature really shows that we are terrible at identifying the real causes of our behavior? That is, what if it shows that we are almost always wrong, when we think we are acting for a particular reason—and that what is really going on is that we are driven by some weird contextual factor that we would not take, upon reflection, to be a reason (i.e., to justify our behavior)? It might still be true that we act from mechanisms that are capable of responding appropriately to our reasons, but, given our nature as depicted in the social sciences literature (and, in particular, dual-process theory), it would be implausible to deem us morally responsible. The bare capacity or ability in question does not seem enough.

I think I agree. I suppose I think that what is required is that we be “good enough,” even if we are not perfect (and, indeed, even if we have considerable flaws), in our reasons-responsiveness. If we are good enough, then it would not be unreasonable to expect individuals who act from mechanisms that are capable of responding appropriately to reasons actually so to respond. Of course, it is difficult to specify what this notion of “good enough” amounts to, and I concede that I have never done a careful and systematic study of the social sciences literature (as John has done so impressively). Perhaps this is intellectually disreputable (and certainly is, at best, somewhat lazy), but I’m just inclined to think that the skeptical results are not as bad as they are sometimes represented as being—that the phenomena are not so prevalent as some suggest, and that they do not affect us in as many contexts, especially “important” decisions.

I’m just frankly skeptical about many of these results, or perhaps the interpretations of the results as suggesting that we are always or almost always wrong about what really motivates us. I do not perhaps have the confidence of Peter Strawson—that our capacities for reasons-responsiveness are “part of the facts as we know them”—but then again I’m a bit closer to him than to the skepticism that is a hallmark of John’s work. There are obviously biases toward publishing sexy and surprising results, and somehow—although this
is just a hunch of mine—I think that these results are overblown and exaggerated. I must confess that this is not based on science, but a gut-level “sense” of mine.

But of course it would be intellectually irresponsible in the extreme just to leave it at that, and not be open to the possibility that I am wrong in this hunch. What if I were to be convinced, over time, that we are not indeed “good enough” at reasons-responsiveness? That is, what if the problems of replication recede and more and more evidence piles up that we typically or almost always don’t know the real causes of our actions; we benightedly think we are voting for the candidate who will ameliorate poverty the most, but it is really because his name starts with the same letter as ours (or there is an unpleasant odor in the voting booth, or whatever). Here I depart from John in what might to some be a surprising way. I really do think that moral responsibility requires reasons-responsiveness. So I would conclude, under the envisaged circumstance in which I were convinced of the skeptical results, that we are not morally responsible. In contrast, John thinks we are manifestly morally responsible, and thus that we should simply reject the reasons-responsiveness (or, his terminology, the “reflectivist”) approach to analyzing moral responsibility (in favor of a value-expression account). So here John is the conservative, and I am the radical—or at least willing to take very seriously the possibility that we are not morally responsible.

Elsewhere, I have objected to Galen Strawson’s skeptical argument about moral responsibility, contending that if an account of “initiation” is proposed that is manifestly impossible to satisfy, it is much more plausible to conclude that this account is not endorsed by common sense than to conclude that we are not morally responsible. That is, I would jettison Strawson’s proposal about initiation, rather than moral responsibility. But I would have to jettison moral responsibility, if I were convinced that we are not good enough at reasons-responsiveness. This is because, unlike a particular account of initiation, I really think that reasons-responsiveness is part of the core of moral responsibility. It is hard for me to imagine that it would be fair or appropriate to target individuals with the reactive attitudes, and associated practices, such as punishment, if they were not sufficiently adept in the space of reasons. In my view, moral responsibility does not depend on whether or not causal determinism is true; so it does not “hang on a thread” in this sense. But it does depend on our being capable of a threshold level of reasons-responsiveness. So it does depend on the truth of a certain empirical claim about human agents.

One final worry. John wishes to allow for moral responsibility (in virtue of value-expression), even when the agent in question is not good at reasons-responsiveness. I’m inclined to think that moral responsibility is importantly connected to the capacity for interpersonal relationships, such as love and
friendship. But it is deeply puzzling to me how one could have a friendship, or stand in a relationship of love, to an individual who systematically is wrong about the real causes of his or her behavior. How could one have a close friendship with someone who thinks she is acting from reasons, but, in fact, always or almost always acts from causes that are not only different from what she believes is motivating her behavior, but would not count as reasons at all? This is further impetus for thinking that there is a close connection between moral responsibility and reasons-responsiveness.

University of California, Riverside
fischer@ucr.edu

Acknowledgment
I am very grateful Benjamin Mitchell-Yellin for helpful comments on a previous version of this paper.

References