

Replies

JOHN MARTIN FISCHER

University of California, Riverside

I am very grateful to the extremely thoughtful, challenging, and insightful comments by all three commentators. I have selected some salient points, but I fear I will have only scratched the surface, at least in some instances. I hope to give further thought to many of these issues, and the challenges raised by the commentators, in future work.

Reply to Clarke

I.

Clarke notes that the idea that moral responsibility shouldn't "hang on a thread" (as I have suggested) "doesn't uniquely recommend a defense of the compatibility of responsibility and determinism. ... It might equally motivate an effort to work out a revised conception of what it is to be responsible, or, more radically, a conception of the ethical that dispenses with responsibility. One might find the resulting view less desirable, but it's unclear what, at this point, desirability has to do with it."¹

I agree with Clarke here, but I would simply point out that I think there are additional considerations that suggest that we shouldn't go in the indicated directions—revise our conception of moral responsibility or give up the connection between the ethical and moral responsibility. (I certainly do go in for a considerable revision of the conditions for moral responsibility, as Clarke points out and I discuss below, but this is different from revisionism about the concept or perhaps nature of moral responsibility.) The idea that moral responsibility shouldn't hang on a thread—that it should be resilient with respect to the discovery that causal determinism is true, or that certain kinds of indeterminism obtain—is then part of a bigger argument. I think that we assess arguments in a holistic fashion, and the plausibility of the *total package* is what's at issue.

¹ Randolph Clarke, "Determinism and Our Self-Conception," this issue, p. 243.

2.

I have suggested that the “picture” underlying our view of ourselves as morally responsible is “self-expression” of a certain sort. More carefully, I have suggested that what we care about—the value of acting so as to be morally responsible—is the value of unhindered self-expression. Clarke points out that there are several respects in which “this picture of agency as unhindered self-expression fails to do the work that is being asked of it here.” Clarke points out that this model might fit well with Gary Watson’s attributability face of responsibility, but it doesn’t appear to capture the accountability face. Clarke also ascribes to me a restriction of “underived moral responsibility” to actions or to actions and omissions. But he further notes that “it isn’t only in our actions that we may be said to express the meanings of sentences in the books of our lives. In our feelings of envy and anger, our longings and opinions, we do so as well. ... It would be an absurdly unrevealing autobiography that told us only what voluntary actions its author performed.”

But the self-expression picture, as I have sketched it, is not supposed to do *all* of the work identified by Clarke. Perhaps I was less than careful in using the term “picture”, but I was more specifically seeking to identify the value we place in acting as morally responsible (or perhaps autonomous) agents. We *do* believe there is intrinsic value in so behaving (and in being the sorts of creatures that can so behave), and I was seeking to identify this value. Of course, an adequate explanation of the accountability face of moral responsibility would not necessarily be expected to follow from the precise articulation of the value in question. (I think we can make some progress in understanding the accountability face in terms of the idea of reasons-responsiveness.)

I wish to point out that, despite some of my formulations, I certainly intend to include more than merely voluntary action in the stories of our lives. Perhaps I should have said, “In behaving (acting or omitting), where the behavior can be voluntary or involuntary, or even in having certain emotional reactions, we can be understood as writing a sentence in the book of our lives.” But this just didn’t seem to be as nice as, “In performing an action at a given time, we can be understood as writing a sentence in the book of our lives.” It just didn’t have that ring to it.

3.

Clarke points out that it is not straightforward to move from reflection on the Frankfurt examples to the sorts of conclusions I want to draw: “... in Frankfurt cases what ensures that one does a certain thing need

have nothing to do with what actually brings it about that one does that thing. What remains jarring, then, is the view that we can be fully responsible for much of what we do, even though the very processes that bring about our doing these things preclude its ever being up to us whether we do them. No reflection on Frankfurt cases can render this an entirely comfortable thought.”² Well, my goal is not to make this an “entirely comfortable thought”, although that would be great; rather, I’m hoping to make it comfortable enough.

Call the factors that actually bring a certain thing about the “A-Factors,” and the factor or factors that render the thing in question inevitable “B-Factors”. As Clarke points out, in the Frankfurt cases the B-Factor is not an A-Factor; indeed, this reflects the signature structure of preemptive over-determination that seems to be distinctively potent in calling into question the Principle of Alternative Possibilities (PAP). Thus, the Frankfurt cases have “disjoint factors.” In contrast, causal determination is (or involves) factors that are (arguably) both A-Factors and B-Factors. Thus, causal determination is a context of “overlapping factors”. Of course, one cannot move straightforwardly from a context of disjoint factors to one of overlapping factors.

I grant this point, but I have never supposed one could make the sort of transition here envisaged. Rather, the argument takes place in steps. First, the Frankfurt-cases are supposed to show that the *mere* lack of alternative possibilities (of the relevant sort) does not imply that the agent in question is not morally responsible for the behavior. Second, my claim is that *if* the mere lack of such alternative possibilities does not imply that the agent is not morally responsible, then it is puzzling as to why the lack of alternative possibilities *as a result of causal determination* would have that implication. That is, if the end-state of not having alternative possibilities (of the relevant sort) does not *in itself* rule out moral responsibility, why does *this particular way of getting to that end-state* rule out moral responsibility? Now I’m not saying that there is no way to answer this question; rather, my view is that it is at least a legitimate question, and it puts the ball in the court of the defender of PAP.

The question under consideration at this point is essentially this: if the end-state of lacking alternative possibilities does not in itself rule out moral responsibility, why does a particular path *to that end-state* rule out moral responsibility? That is, why would causal determinism’s precluding alternative possibilities (or, for that matter, making a difference) rule out moral responsibility? Now it might be that causal determination *directly* rules out moral responsibility (i.e., not in virtue of

² Clarke, “Determinism and Our Self-Conception,” this issue, pp. 247-8.

precluding alternative possibilities); but then we could consider this claim—a claim I have attended to carefully. Having systematically canvassed various features of causally deterministic sequences, I have concluded that there is no strong argument that causal determination *directly* rules out moral responsibility.

So the mere lack of alternative possibilities arguably at least does not rule out moral responsibility, and causal determination *in itself and apart from ruling out alternative possibilities* (arguably, at least) does not rule out moral responsibility. So how could causal determination's precluding alternative possibilities rule out moral responsibility? I suppose it is possible that, even though each factor (the lack of alternative possibilities and causal determination) does not *in itself and in the absence of the other* rule out moral responsibility, they work together (perhaps synergistically) to expunge moral responsibility. This is a possibility, but I don't find it plausible. I have never suggested that invocation of the Frankfurt cases generates a decisive argument for the rejection of PAP or for compatibilism, but I do think that the examples pose a good challenge for the defender of PAP: if the facts identified above do not individually rule out moral responsibility, how is it that their combination would? This would seem to be magic, and, although I love magic, I do not love it in philosophy.

I have no doubt that we ordinarily think we can “make a difference” in various contexts; no doubt, the capacity to make a difference is indeed part of our ordinary conception of our agency. My point here is not to deny this, but to suggest that it does not survive careful philosophical scrutiny. More specifically, the idea that what we value in acting freely (and thus behaving so as to be morally responsible) is *making a certain sort of difference* does not survive critical scrutiny, especially in light of the Frankfurt cases.

Suppose that, having seen the “make a difference” sign in front of the church, Randy decides to work for Obama every Sunday afternoon (after church, of course) until the election. But imagine that we know that Randy would have been caused by a counterfactual intervener to make the very same decision, were Randy about to decide to spend his Sunday afternoons (say) watching his favorite NFL team. When Randy freely decides to volunteer for Obama on Sunday afternoons and does indeed do so, it is plausible that Randy's free decision and behavior has value. But is the value in question the value of “making a difference? Well, Randy would have made the same kind of decision (individuated in the relevant way) and would have behaved in the same kind of way (individuated in the relevant kind of way) in *either* scenario; and there was no other scenario accessible to Randy.

We can say that Randy made a difference in the sense that he made it the case that he decided to volunteer for Obama (and did volunteer work for Obama) as a result of *his own free choice and not the intervention of an external agent or force*. But I doubt that *this* is the kind of difference the minister had in mind when he put up that sign, or the kind of difference that could plausibly be invoked in seeking to articulate the value of Randy's free decision and action. Is it really attractive to suppose that what we value in Randy's exercising free will is that he has knocked on doors as a result of the normal human deliberative process, as opposed to knocking on doors as a result of direct stimulation of the brain? After all, Randy will decide to knock on the same doors for Obama, knock on those same doors, have the same conversations, and perhaps even influence the same potential voters, in *both* scenarios accessible to him. Why would the capacity to make the sort of difference under consideration be of value, especially given that the possibility of making a difference of this sort is not even part of Randy's practical reasoning that issues in his decision?

When we talk about the value of making a difference, we are typically thinking in terms of making a difference to *end-states* individuated in *broad* ways. So, for example, perhaps the minister is thinking that his flock (or even innocent passersby) can help to alleviate poverty, or help to build houses for the poor, or provide food to the hungry, and so forth. Perhaps Randy is thinking that in working for Obama, he can at least in a small way help to change American foreign policy or make the American distribution of income and wealth less unfair, and so forth. That is, insofar as the value of making a difference is relatively clear and uncontentious, we are (typically, at least) thinking in terms of making a difference to end-states, broadly construed. But it is quite a different matter when the end-states (so construed) are fixed, and we can only make a difference to the *route* to the given end-state.

I doubt, for instance, that the minister would be excited to know that a given house was built for a poor family as a result of a process that involved Jones's unimpaired decision rather than as a result of a process that was triggered by direct stimulation of Jones's brain; the same house would have been built for the same poor family in both scenarios. Now of course this is not to say that we don't care about—and indeed value—Jones's free decision and activity; it is simply to say that it is awkward to find that value in the value of making a difference. One *does* make a difference, but is it a difference of the right kind? (This point is parallel to my contention that it is not enough to identify *any old* alternative possibility in a Frankfurt-case; it must be an alternative possibility *of the right kind*.)

The story of Callie reminds me of Alfred Mele's story of Diana, which he presents as part of his Zygote Argument. I do think that a compatibilist must here bite the bullet; that is, a compatibilist must admit that the world could have been set up by an agent with the intention that it unfold just as it actually does. (After all, there are religious views, such as those of Luis de Molina, that go pretty much just like this!) Elsewhere I have sought to argue that the issue of whether the world—or a particular part of it—was set up distally by an agent with particular intentions is simply irrelevant to an agent's moral responsibility.³ I believe that moral responsibility is an essentially *historical* notion, but, on my view, we need not (and should not) go so far back in history in our responsibility judgments. Moral responsibility is a matter of how the actual sequence unfolds; on my view, the history of a bit of behavior must contain guidance control. But this does not entail that it couldn't have been the case that an agent was there at the beginning, and created the world with precisely the intention that that bit of behavior be produced in just that way.

Reply to Yaffe

Yaffe's comments are particularly kind and gracious; they are the ones I forwarded to my mother and father! But I should also say that I remain somewhat puzzled by Yaffe's views, ingenious as they are. In the big picture, we are philosophical friends: we both think that causal determinism is compatible with moral responsibility and with ethics, as it were. I get there by denying Ought-Implies-Can, whereas Yaffe gets there by a different route, which involves a certain sort of distinction between different kinds of not-doings. I defend my route because I think it is the most attractive way to get to the conclusions we share, but, in a more expansive mood, I suppose I could simply say that there is more than one way to skin the cat.

But we philosophers traffic in disagreement, so perhaps it will be helpful at least to explore the specific locus of disagreement between Yaffe and me. I find the fixity of the past idea, interpreted in a certain way, deeply plausible (although certainly open to doubt by reasonable and thoughtful philosophers). Simply put, I adhere to the idea stated elegantly by Carl Ginet: our freedom is the power to add to the given past, holding the laws of nature fixed. Put slightly differently, our freedom is the power to "get there from here"—to extend the actual past,

³ John Martin Fischer, "Review of *Free Will and Luck*, by Alfred Mele," *Mind* 117 (2008), pp. 195-201.

holding fixed the laws of nature. On this picture, which I contend corresponds to the commonsense view of the future as a garden of forking paths—paths that are branches off a given, single path—it would not make any sense to distinguish between doings and not-doings, or (among the not-doings) refrainings and mere omissions. If I am now free to behave in a certain way, then I must have it in my power to get there from here; there must be a path that goes through the past and present to a scenario in which I behave in the indicated way. Admittedly, some philosophers do not accept this picture. But my point is that if one accepts it, one should not say that it only applies to doings, rather than not-doings; and one should not say that among the not-doings, it only applies to refrainings.

Yaffe disagrees. He thinks that the standards with respect to the fixity of the past are stricter when we are considering refrainings than when we are considering mere omissions. I believe that Yaffe gets to this conclusion by taking a detour through deliberation. That is, he says,

Note that premise (3.2) enshrines the thought that every accessible possible world is one which is similar to the actual world in its past and its laws. Why this constraint on accessibility? The reason offered by the Consequence Argument is that the past and the laws are not things that agents can affect. So, the thought is, when deciding whether or not a possible world is accessible, we must check to see if all those things that the agent is powerless to affect are as in the actual world; if not, then the world in question is not accessible.⁴

Yaffe then asks why we should focus only on worlds in which those things that the agent is powerless to affect are as in the actual world. He answers as follows:

There is probably more than one way to capture this intuition, but here's one: When an agent is deliberating correctly about what to do, he considers only those courses of conduct that he believes he has the ability to do. In deciding what he has the ability to do, he is really asking himself what he might do *given that the world is in some ways unchangeable by him*. That is, the deliberating agent holds fixed those things that he is, or believes himself to be, powerless to affect.⁵

But now Yaffe wonders “how ... we get from this observation about deliberation to the view that worlds are inaccessible, in the relevant sense, if they differ from the actual world in ways the agent is powerless

⁴ Gideon Yaffe, “Comments on John Fischer’s *My Way*,” this issue, p. 255.

⁵ Yaffe, “Comments on John Fischer’s *My Way*,” this issue, p. 256.

to change?” That is, Yaffe wonders how we can go from the relevant sorts of claims about deliberation to claims about our freedom. He contends that the only way to “fill this gap” is by assuming something that both of us would reject, and he concludes that we should have low standards for accessibility when it comes to what an agent can omit, but higher standards for accessibility when it comes to what an agent can refrain from doing.

Before I consider a final argument Yaffe offers, I simply wish to state that I would not be inclined to seek to explain the fixity of the past, as I construe it, *via* considerations about deliberation. This seems to me to be a needlessly roundabout way of proceeding. In my view, it is not helpful to seek to explain the fixity of the past (construed as I have construed it) by reference to considerations about deliberation; rather, the fixity of the past is a basic notion that *constrains* deliberation, but is not *explained by* features of deliberation. In my view, Yaffe gets to a dead end because he starts off on the wrong path in seeking to explain the fixity of the past. Although I admire his ambition and ingenuity here, it might just be that there is *no* deeper explanation of the fixity of the past—the notion that our freedom is the power to extend the given past. (I am not sure we *need* a deeper explanation of the fixity of the past; if so, would we also need a deeper explanation of the deeper explanation, and so forth?)

Yaffe’s final (and, again, quite imaginative) argument is based on my objection to mere flickers of freedom as insufficiently robust to ground attributions of moral responsibility (given the regulative control model of moral responsibility). Yaffe says:

As I understand it, this is a symmetry argument, Fischer’s idea is that whatever the opponent takes to be required for responsibility in the actual sequence over and above alternative possibilities must be true in the alternate sequence too. Otherwise, the possible alternative isn’t sufficiently ‘robust’. But the symmetry principle that Fischer’s argument here employs commits him to the view that I am pressing with regard to “accessibility” [according to which the standards of accessibility are higher in the case of refrainings than omissions].⁶

But, as far as I can tell, my argument for the requirement of robustness for alternative possibilities is *not* a symmetry argument. If it *were* a symmetry argument, then I think Yaffe would be right that this would call into question a natural construal of the fixity of the past; but, in my view, that is even more reason *not* to employ some sort of general symmetry argument. Rather, my argument is based on the idea that

⁶ Yaffe, “Comments on John Fischer’s *My Way*,” this issue, p. 258.

you can't get something from nothing. That is, it would seem to be *alchemy* to suppose that simply by adding a flukish alternative possibility—a mere flicker of freedom—to a context of no moral responsibility, you get—presto!—moral responsibility. To suppose that adding such an exiguous alternative possibility could get you moral responsibility would be like supposing that appealing to quantum indeterminacies or random events in the brain would help to secure *control* and responsibility; the problems are parallel. These considerations certainly do not seem to be appeals to *symmetry*.

Reply to Normore

I agree with Professor Normore—and philosophers such as John Searle—that we tend to experience reasons as not *compelling* or *forcing* us to act. (I think that it is a gratuitous leap to the conclusion that we experience a “causal gap”, as Searle would say. Obviously, it is one thing to say that we do not experience ourselves as causally determined; it is quite another to say that we experience ourselves as not casually determined.) Further, I suppose I would agree with Normore that we tend to experience (at least most of) our reasons as “resistible”. I would be inclined to understand this experience as involving a belief about the intensity of the urges or desires or pro-attitudes associated with the reason; when I experience a reason as irresistible, or giving rise to an irresistible desire or urge, I experience the *intensity* of the desire or urge as such as to render it impossible that I resist. On my account, then, experienced irresistibility involves the view that the agent cannot resist *in virtue of the intensity of the associated desires or urges*.

Much depends here on one's account of practical reasoning (of which there are various Humean and Kantian versions), and I cannot go into details here. On the surface, it is not clear what exactly is meant by “resistibility” and “irresistibility”, as they apply to *reasons* (rather than the associated desires or urges). Further, I would simply note that some philosophers (including Frankfurt and Watson) think that there can be what they call “volitional necessities”; as I interpret the idea, volitional necessity involves “irresistible reasons”. Indeed, Frankfurt thinks that volitional necessities are indispensable ingredients in the formation of the self, practically speaking.

Of course, if causal determinism is true, and we grant that the Consequence Argument is sound, then it will turn out that all of my desires are such that I cannot resist them. And if I know that causal determinism is true, and I accept the Consequence Argument, then I can know that I can never resist any of my desires. But it does not follow that all of my desires are irresistible, in my sense. Surely, not even in a causally

deterministic world would all of my desires present themselves as forcing or compelling me to act—as rendering it impossible that I resist *in virtue of their intensity*.

Return to Normore's claim, "If it is feature of our experience of reasons that that we experience them as resistible and it is a feature of our conception of causal connections that they necessitate, then we cannot simultaneously take something to be a reason and a cause of our action." I reply as follows. We (typically, at least) experience reasons as resistible, in the sense that the intensity of the associated pro-attitudes does not explain why we cannot resist them. It is not a feature of causal connections—even deterministic causal connections—that they "compel" or "force"; if necessitation simply means that alternative possibilities are ruled out, then we can simultaneously regard something as a reason and as a (deterministic) cause of our actions. In a causally deterministic world, only some reasons will give rise to or be associated with "irresistible" urges—those whose intensity is so great that we cannot resist them. One can have necessitation without compulsion. Given knowledge of causal determinism and the soundness of the Consequence Argument, I would have to see all my actions as the result of choices that, it turns out, I couldn't have failed to make. But I wouldn't thereby have to regard them as irresistible—as compelling me or forcing me. After all, causal determination entails necessitation, but the relevant kind of necessitation is *different* from "irresistibility".

Normore agrees with my claim that what matters to moral responsibility is that we act from certain kinds of rational faculties; he says, borrowing the phrase from Wilfrid Sellars, that this puts us in the "space of reasons". But he says:

If we have to see reasons as resistible, then we have to suppose that the faculty is not in itself something which, given a particular set of reasons inevitably settles matters one way rather than another—that it is not a 'mechanism', as that is usually understood. ... My own suggestion is that we follow Ockham and Buridan here and characterize the faculty as one which, given that it is producing a given output with a given set of inputs, and holding constant all those inputs and everything else in the world outside that faculty, could come to produce a different output, i.e., that we characterize the faculty as not only responsive to reasons, but as indeterministic in this attenuated sense.⁷

It is an important part of the Fischer/Ravizza account of guidance control that the actual mechanism that issues in the choice could have—had the capacity to—respond differently to the *actual* reason. So

⁷ Calvin Normore, "Comments on J.M. Fischer, *My Way*," this issue, p. 265.

Normore and I are very close to agreement here. What he says, and I deny, is that this capacity of the actual-sequence mechanism (or rational faculty) must be interpreted indeterministically. I am open to the indeterministic interpretation, and nothing in the core of my view rules it out, but I simply resist the notion that we *must* interpret it indeterministically. In my view, one of the advantages of my approach here is that it renders moral responsibility resilient to the possible empirical discovery of the truth of causal determinism.

Although I do not have the space to discuss this point in any detail here, I wish simply to assert that nothing in my account of guidance control is incompatible with standard assumptions in decision theory. I do not suppose that, apart from special assumptions (such as causal determinism or the existence of an essentially omniscient God), an agent lacks the ability to choose or do otherwise (regulative control); even with the assumption of causal determinism, an agent is not in a position to know what choice is pre-determined, and must then deliberate *as if* his various options are available to him. I have presented my own proposed analysis of Newcomb's Problem elsewhere, and I hope further to discuss and defend it in future work, and to apply it to related puzzles.⁸

In the end, I wish to agree with both Normore and Clarke that if we were to learn that causal determinism is true, this would significantly alter our most basic views of ourselves. We ordinarily think of the future as a garden of forking paths, and we think of our responsibility as based on "making a difference"—selecting one path from among various genuinely open paths in the garden of the future. Given knowledge of causal determinism, and assuming the soundness of the Consequence Argument, we would have to re-configure our basic views of our agency and responsibility, and this would be no small thing. Semicompatibilism, then, together with my total "package" of related views, is a radical view—or perhaps represents a radical reconstitution of our ordinary views. I have never supposed that accepting this total package would be easy or convenient. After all, I have somewhat immodestly declared it, "The Free Will Revolution."

Chairman Mao, when asked what he thought of the French Revolution, said, "It is too early to tell." And it is certainly too early to tell whether I (and the other pioneering band of Semicompatibilists) can achieve even a small change in the traditional ways of thinking about this interrelated web of issues. Chairman Mao also apparently did not

⁸ John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control* (Oxford: Blackwell Publishers, 1994); "Newcomb's Problem: A reply to Carlson," *Analysis* 61 (2001), pp. 229-36; and "Critical Notice of J. Howard Sobel's *Puzzles of the Will*," *Canadian Journal of Philosophy* 31 (2001), 427-44.

much care about his physical appearance or clothes, when meeting foreign dignitaries; when one of his advisors once told him that he had a hole in his pants, he reportedly said, “But who will look at my bottom?” Here is where I depart company from my fellow revolutionary; I have tried my best to cover my ass.