The Freedom Required for Moral Responsibility

John Martin Fischer

DOI: 10.1093/oso/9780198817277.003.0013

Abstract and Keywords

Some philosophers argue that we do not need freedom to do otherwise or access to alternative possibilities for moral responsibility. These philosophers are actual-sequence theorists of moral responsibility who are motivated by Frankfurt cases, in which there is pre-emptive overdetermination. They contend that in these cases the agent is morally responsible but does not have freedom to do otherwise or access to alternative possibilities. Others have rejected the actual-sequence approach. They contend that the sort of freedom to do otherwise required for moral responsibility is indeed present in the Frankfurt cases. This essay explores the significance of the debate between these two camps. Are the two views importantly different or mere notational variants of each other? I examine these questions with attention to Terry Irwin’s discussion of Aristotle on responsibility.

Keywords: actual-sequence, alternative possibilities, Aristotle, Frankfurt cases, freedom of the will, moral responsibility
T.H. Irwin was extraordinarily influential in my development as a philosopher. He was a mentor par excellence at Cornell, where I was fortunate enough to work with him extensively in graduate school in philosophy. Here I shall select one area in which his work has had a big impact on me: moral responsibility theory. I learned a tremendous amount from Irwin’s classic essay (1980). Irwin interprets Aristotle as offering a (broadly speaking) reasons-responsiveness approach to moral responsibility. I shall begin by laying out (in a very brief form) some of the main ideas in Irwin’s ‘reconstruction’ of Aristotle’s views on moral responsibility. I then point out ways in which my own approach shares central features of this Aristotelian account, while also departing in some ways. Finally, I seek to defend my theory of moral responsibility in light of some salient and important recent challenges.

Irwin’s Aristotle on Moral Responsibility

Irwin’s ‘Reason and Responsibility’ is a brilliant and important contribution, not just to Aristotle scholarship, but to moral responsibility theory more broadly. Irwin is careful to point out that the account of moral responsibility he is attributing to Aristotle is not explicitly in any of Aristotle’s writings. Rather, he is taking key passages from Aristotle’s texts and putting them together in ways that capture what he takes to be central ideas in Aristotle, as well as ideas to which Aristotle, upon reflection, would see that he is committed. Irwin writes:

Despite the apparent simplicity and negative procedure of Aristotle’s account, I believe he needs and assumes a positive account that raises more complex questions about the general conditions for responsibility. When we consider difficulties in the account as it stands and try to solve these difficulties by appeal to other Aristotelian doctrines, the resulting theory will be less simple, but more interesting, and more capable of answering some hard questions about responsible action. I will set out this ‘Aristotelian’ theory—I will call it Aristotle’s ‘complex theory’—partly for an exegetical purpose, to show what Aristotle is committed to, whether he (p. 217) realizes it or not. But I have a philosophical purpose too, to show that the account is worth consideration in its own right.

(Irwin 1980: 117-18)

In evaluating various Aristotelian ideas about moral responsibility (and constructing an Aristotelian account of responsibility), Irwin is guided by the following extremely important insight:

We normally suppose that a responsible person can do something about the desire he acts on—that it is not compulsively strong but is responsive to his reasoning and deliberation. We assume that rational agents are capable of doing more than simply acting on the desire they find to be strongest; they can also affect the strength of their desires by further deliberation.
I fully agree with Irwin’s point here, and I have made the notion of ‘reasons-responsiveness’ central to my own ideas about moral responsibility. It has always seemed plausible to me that a compulsive desire is indeed one that is not responsive to reason; given the strength of the desire, one would still act on it, no matter what the reasons to do otherwise were. Reasons-responsiveness represents a significant way in which we are not just passive, but active; we are not just ‘passive subjects or spectators of [our] desires’, but, rather, we are agents. We don’t just act on our strongest desires, where the strength is fixed independently of deliberation and reflection. As Irwin puts it, ‘Here Aristotle is right to suggest that decision and rational desire, as he understands them, are important for responsibility’ (Irwin 1980: 131).

After considering various simpler accounts of moral responsibility suggested by Aristotle, Irwin ‘restates’ Aristotle’s view as follows:

(18) A is responsible for doing x if and only if (a) there is some deliberative argument which, if it were presented to A, would be effective about A’s doing x; (b) A does x voluntarily. (Irwin 1980: 138; the number ‘18’ represents the numbered proposition in Irwin: 1980; I shall employ his numbered propositions throughout.)

He then elaborates upon (b) to produce this ‘summary’ of Aristotle’s conception of moral responsibility:

(19) A is responsible for doing x at a time t if and only if (a) there is some deliberative argument which, if it were presented to A before t, would be effective at t about A’s doing x; (b) A believes at t that x is F; (c) x is F; (d) A’s beliefs and desires, as reasons, cause A to do x.

(Irwin 1980: 139)

This conception captures a reasons-responsiveness approach to moral responsibility. Clause (d) expresses the idea that the relevant action must be caused by reasons; and clause (a) conveys (perhaps among other things) the idea that the agent was responsive to considerations involving reasons in such a way that he could have done otherwise. This interpretation of (a) seems to be what Irwin has in mind. Consider, for instance, the remarks he offers in support of this conception:

Suppose A did x yesterday, and today we are considering his action. We might believe from what we know about A’s past that a deliberative argument applied yesterday before he did (p.218) x would have prevented him from doing x and that therefore he was responsible for doing x when he did it.
It is important that, on Irwin’s Aristotelian approach, moral responsibility requires more than causation (of the right sort) by reasons; it involves responsiveness to reasons and freedom to do otherwise.¹

Guidance Control and the Appeal to Mechanisms
I shall begin with a challenge to Irwin’s Aristotelian conception of moral responsibility, as captured in (19). I think that there could be a person who is, intuitively, morally responsible for performing an action at a time, but whose deliberative capacities would be impaired or bypassed, in any scenario in which he were presented (just prior to the time of the action) with a deliberative argument against doing it. That is, as things actually are, the individual’s deliberative capacities are intact, and he performs the action as a result of his reasons (in the right sort of way). But in all counterfactual scenarios in which he is presented a compelling deliberative argument against performing the action, he would still perform it, because of a disruption in his deliberative capacities—perhaps a nefarious neurosurgeon would be directly manipulating his brain only in the counterfactual scenarios. An adequate account of moral responsibility should imply that the individual is morally responsible for the action; but Irwin’s Aristotelian proposal does not have this result.

I believe that this sort of problem helps to motivate a relativization to the ‘actual-sequence mechanism’ in a reasons-responsiveness account of moral responsibility. This is indeed how I seek to accommodate the famous ‘Frankfurt-style cases’, which are putative counterexamples to the Principle of Alternative Possibilities, according to which moral responsibility for an action requires freedom to do otherwise (Frankfurt 1969). Here is a version of the cases, all of which contain a signature structure involving pre-emptive overdetermination:

Black is a stalwart defender of the Democratic party. He has secretly inserted a chip in Jones’s brain, which enables Black to monitor and control Jones’s activities. Black can exercise this control through a sophisticated computer that he has programmed so that, among other things, it monitors Jones’s voting behavior. If Jones were to show any inclination to vote for anyone other than the Democrat, then the computer, through the chip in Jones’s brain, would intervene to assure that he actually decides to vote for the Democrat and does so vote. But if Jones decides on his own to vote for the Democrat (as Black, the old progressive would prefer), the computer does nothing but continue to monitor—without affecting—the goings-on in Jones’s head.

(Fischer 1982)
Now suppose that Jones decides to vote for the Democrat on his own, just as he would have if Black had not inserted the chip in his head. It seems, upon first thinking about this case, that Jones can be held morally responsible for his choice and act of voting for the Democrat, although he could not have chosen otherwise, and he could not have done otherwise. In this sort of case, the ordinary deliberative mechanism actually operates, and, intuitively, the individual is morally responsible for his action. But in the alternative scenario, there would be direct intervention in the brain producing the same action as in the actual sequence. We can of course generalize the case so that there would be such an intervention in all counterfactual scenarios. In any case, my preferred account of moral responsibility relativizes the question of reasons-responsiveness to the actual-sequence mechanism. In the Frankfurt-style case above, the actual-sequence mechanism (ordinary practical reasoning, which does not include Black and his device) is responsive to reasons, even though the alternative scenarios all involve non-responsive mechanisms. Given that the actual-sequence mechanism is suitably responsive to reasons, the agent is morally responsible (on my approach), even though I claim that the agent himself is not responsive to reasons and he cannot do otherwise.

On my view (following Aristotle), one is morally responsible for one’s behavior insofar as one meets both an epistemic and a ‘freedom-relevant’ condition. On my view, we can learn from the Frankfurt-style cases that the freedom-relevant condition involves ‘acting freely’, rather than freedom to do otherwise; and Jones acts freely in the actual sequence. That is, the freedom-relevant condition involves the exhibiting of a distinctive kind of control: ‘guidance control’, which does not require control over one’s behavior, or ‘regulative control’ (Fischer and Ravizza 1998).

What follows is a very brief account of guidance control (Fischer and Ravizza 1998). An individual exhibits guidance control to the extent that he acts from his own, suitably reasons-responsive mechanism. There are thus two important components of the analysis: ownership of the actual-sequence mechanism, and reasons-responsiveness of that mechanism. Clearly, the analysis presupposes that we can distinguish (in a reasonable way) the actual-sequence mechanism from various alternative-scenario mechanisms. As I have previously explained, this is the key to accommodating the Frankfurt cases. That is, in these cases, the actual-sequence mechanism may be the agent’s own, suitably reasons-responsive mechanism, even though the alternative-scenarios involve mechanisms that are not the agent’s own or not appropriately reasons-responsive (or both). So, for example, in the Frankfurt case sketched, Jones’s actual-sequence mechanism is ‘ordinary practical reasoning’ or ‘the human deliberative mechanism’, whereas the alternative scenario features a manipulation mechanism.
On my view, an agent’s mechanism becomes his own when he takes responsibility for it; he thereby makes the mechanism his own. Note that a mechanism could be inculcated in an agent in such a way as to be reasons-responsive; in this case it might not be the agent’s own mechanism. Thus, we need to supplement reasons-responsiveness with ownership to get an adequate account of the kind of control implicated in moral responsibility. Here are the three conditions for taking responsibility and thereby making one’s mechanism one’s own:

a) The individual must see himself as an agent; he must see that his choices and actions are efficacious in the world. This condition includes the claim that he sees that if he were to choose and act differently, different upshots would occur in the world.

b) The individual must accept that he is a fair target of the reactive attitudes as a result of how he exercises this agency in certain contexts.

c) The individual’s view of himself specified in the first two conditions must be based, in an appropriate way, on the evidence. (Fischer and Ravizza 1998: 210–14)

In order for an agent to meet the freedom-relevant condition on moral responsibility for his behavior, it must be the result of his own, suitably reasons-responsive mechanism. The crucial idea here is that we must hold fixed the actual-sequence mechanism (say, unimpaired human practical reasoning), and we then ask whether they agent would (given the operation of that same mechanism) recognize reasons to do otherwise and respond to those reasons. This is a very brief and oversimplified account, but I do not believe that the details matter for the purposes of this essay. (For a full exposition, see Fischer and Ravizza: 1998: 28–91.)

The key departures of my approach from Irwin’s Aristotelian account are the mechanism-relativization of reasons-responsiveness and the explicit rejection of the requirement of alternative possibilities (freedom to do otherwise). In the following sections, I will defend these features of my reasons-responsiveness theory against important criticisms. But before I turn to these replies, I shall seek to clarify my contention that I have offered an ‘actual-sequence’ account of moral responsibility, that is, an account that does not require alternative possibilities (i.e., freedom to do otherwise).

An Actual-Sequence Approach to Moral Responsibility
My account of moral responsibility is an ‘actual-sequence’ model of moral responsibility. That is, it fixes on (possibly modal or dispositional) properties of the actual sequence, and features of other possible worlds are relevant only insofar as they bear on these properties of the actual sequence, not insofar as they indicate that the agent has freedom to do otherwise. Recently, Christopher Franklin has contended that everyone (and that includes me) holds that an ability to do otherwise is necessary for moral responsibility (Franklin 2015). A brief examination of this point will help to clarify my actual-sequence approach.

Franklin writes:

The issue is not whether the ability to do otherwise is necessary for freedom and responsibility: the issue is which ability to do otherwise is necessary. Everyone thinks that some ability to do otherwise is necessary for freedom and responsibility....

(Franklin 2015: 2092)

(p.221) Later in Franklin’s interesting and thoughtful piece, he writes:

Fischer’s repeated denial that alternative possibilities are relevant [to moral responsibility] is apt to mislead (and has misled). What he really intends (or should intend) by these statements is that certain species of abilities are irrelevant. ... He employs FSCs [Frankfurt-style cases] to show that the ability to do otherwise ... is not necessary for moral responsibility. Fischer’s phrasing of the point in terms of ‘alternative possibilities’ has the unfortunate suggestion that he does not think that an ability to do otherwise is required in any sense. ... That such confusions have arisen in the literature becomes clear when we consider recent compatibilists who argue that Fischer is mistaken in concluding that the ability to do otherwise is irrelevant to moral responsibility (or at least blameworthiness). This is most clearly seen in the case of the new dispositionalists [such as Michael Smith (2003), Kadri Vihvelin (2004) and Michael Fara (2008)].

(Franklin 2015: 2097)
I completely agree with Franklin that I do indeed believe that *various* kinds of alternative possibilities are required for moral responsibility (although not for the ‘grounding’ or explanation of moral responsibility), and thus that my repeated contention that alternative possibilities are not required for moral responsibility might well have caused confusion. Of course, in the Frankfurt-style cases Jones does not lose his general ability to vote for a Republican (i.e., to do otherwise in the context), just as we do not lose various general abilities when we are asleep. One might say that we keep our general abilities in these contexts, although we do not have the opportunity to exercise them (immediately). Again: I do not lose my general ability to play the piano when there is no piano in my vicinity; I simply do not have the opportunity to exercise it.

So Franklin is certainly right: we all believe that *some* kinds of alternative possibilities are present in any case in which an individual is morally responsible for his behavior (even if the presence of such alternative possibilities does not ground or explain the agent’s moral responsibility). But, as Franklin also notes, these were *not* the sorts of alternative possibilities I had in mind in contending that moral responsibility does not require alternative possibilities. I have absolutely no interest in showing that moral responsibility does not require general capacities or abilities to do otherwise, or various other kinds of abilities to do otherwise that abstract away from the particulars of the agent’s history and/or present situation. After all, there is no reason to suppose that these sorts of abilities to do otherwise are incompatible with causal determinism! I have always been interested in the sort of alternative possibility that would be (or could plausibly be thought to be) ruled out by causal determinism. And, clearly, general abilities and indeed any sort of ability to do otherwise that abstracts away from features of the agent’s past and/or current situation need not be inconsistent with causal determinism.
So perhaps I should have been clearer all along. As Franklin points out, I take the Consequence Argument very seriously; this argument purports to show that causal determinism would rule out alternative possibilities (Van Inwagen 1983; Fischer 1994). But here it is obvious that the alternative possibilities in question are conceptualized as requiring the past and current situation of the agent to be held fixed. It is this sort of alternative possibility that the incompatibilist believes is both ruled out by causal determinism and required for moral responsibility. I have sought to engage with the incompatibilist by stipulating, for the sake of the argument, that this is the sort of alternative possibility that is envisaged by Principle of Alternative Possibilities: that is, an alternative possibility that is not abstract in the way a general capacity is abstract or any similar way, i.e., by subtracting features of the past or current situation of the agent. Given this stipulation, I have argued that Frankfurt-style cases show that precisely this sort of alternative possibility is not required for moral responsibility. Given my dialectical project, it is completely unobjectionable that abstract alternative possibilities are present whenever there is moral responsibility; after all, there is no reason to suppose that such alternative possibilities are ruled out by causal determinism.

I shall return to this point below (when I discuss the ‘new dispositionalists’). But first I shall consider objections to my mechanism-based version of a reasons-responsiveness approach to moral responsibility.

A Defense of the Mechanism-Based Approach to Moral Responsibility

In a way, my reasons-responsiveness approach to moral responsibility is the ‘offspring’ of Irwin’s Aristotle and Robert Nozick, because I was also deeply influenced (in developing my account) by Robert Nozick’s accounts of knowledge and ‘tracking goodness’ or ‘aligning with value’ (Nozick 1981; Fischer 1994 and Fischer and Ravizza 1998). Nozick’s theory of knowledge is an ‘actual-sequence’ theory of knowledge, just as mine is an actual-sequence theory of moral responsibility; in both cases we focus on the (dispositional) properties of the actual-sequence mechanisms (or, ‘methods’, in Nozick’s terminology).

As Nozick points out, it is possible for an individual to know that \( p \), where he believes that \( p \) via a certain method. And yet it might also be the case that if \( p \) were false, the individual would use another method instead—a method that would lead him to believe that \( p \), even though it is false. Nozick writes, ‘A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset’ (Nozick 1981: 179). In this sort of case, Nozick holds that the individual does indeed know that \( p \) in virtue of the relevant properties of the actual method of belief-production: this method ‘tracks truth’, even though the method employed in the alternative scenario does not (Nozick 1981: 179).
Nozick is not the only philosopher who employs a ‘method-based’ (or ‘mechanism-based’) theory of knowledge. Because of the possibility of cases of the sort just mentioned, and for various other reasons, many epistemologists have adopted ‘mechanism-based’ theories of knowledge (and/or justified belief). Just two prominent examples are Alvin Goldman and Fred Dretske, both of whom defend ‘reliabilist’ accounts of knowledge. (p.223) (The classic early presentations of theirs views are in: Goldman 1967 and 1976; and Dretske 1970 and 1971.) On these approaches (which differ in their details), an individual’s knowledge that \( p \) requires that the mechanism that produces the individual’s belief that \( p \) be reliable. Many epistemologists have concluded that the best way to analyse knowledge is via mechanism-based accounts.

Just as Nozick argued that one could—and should—give an account of a certain kind of exemplary agency (‘tracking value or bestness’) via a method-based approach, which is structurally similar to his method-based account of knowledge, I have argued that the best way to give an account of moral responsibility is via a mechanism-based model (Nozick 1981: 317–26; Fischer 1994). I agree with Nozick that the analogy between the epistemological context and the context of agency is deeply illuminating, and I have sought further to develop the analogy.

Various philosophers have criticized my employment of a mechanism-based approach to moral responsibility (McKenna 2001 and 2013; Ginet 2006; Watson 2001; and Brink and Nelkin 2013). Perhaps the fullest development of this objection is in insightful work by Michael McKenna, and it will be helpful to lay out salient features of McKenna’s nuanced critique here.

McKenna is skeptical about the Fischer/Ravizza mechanism-based approach, because (in his view) we offer too little explicit guidance on mechanism-individuation. (And, of course, relativization of reasons-responsiveness to the actual-sequence mechanism is a crucial ingredient in my account, so it is important to be clear on what counts as the same mechanism.) He concedes that it is relatively clear that there is a different kind of mechanism operating in the actual sequence and alternative scenario in a Frankfurt-style case (McKenna 2013: 161). But he writes:

> The problem, however, is that there are some contexts crucial to the development of their theory where, absent some principled bases for mechanism individuation, it is hard to assess their broader claims. ... it is an indispensable part of their theory that we are to test how the same mechanism behaves in response to different reasons. How are we to settle whether, when different reasons are put to an agent, the same mechanism is operative? We need some purchase on what it is that we are holding fixed when we hold fixed ‘the same mechanism’ while testing its degree of responsiveness.
McKenna goes on to express the worry that because I (and my co-author, Mark Ravizza) have offered no ‘principled’ basis for mechanism-individuation, ‘we are left to settle the matter exclusively by appeal to our intuitive reactions to varying cases’ (McKenna 2013: 162). And, of course, such intuitions will vary widely, and may be influenced by one’s theoretical commitments about the relationship between causal determinism and moral responsibility (McKenna 2013: 162–3).

I have replied to these sorts of worries (especially as articulated in McKenna 2001) by emphasizing that my reliance on intuitive ideas about mechanism-individuation is no different from similar reliance on non-reductive views about mechanism-individuation (or similar ideas) in other areas of philosophy (Fischer 2004). Will one have to give up a Nozickean method-based epistemology, because one offers no fully reductive account of method-individuation? Similarly, will one have to give up a Dretske/Goldman sort of reliabilism, because one offers no fully reductive account of mechanism-individuation? Perhaps in the end this lacuna will be deemed a significant cost of such theories; but should they be ruled out from the start? I doubt it; these are very important and widely influential epistemological theories—approaches that deserve careful attention, even though they are not fully reductive (as regards mechanism-individuation). And a similar point applies to generalization strategies in ethics—consequentialist and Kantian. Will one give up rule-consequentialism or the universalizability version of Kant’s Categorical Imperative because one has no fully reductive account of rule-individuation or maxim-individuation?

To this reply to his objection, McKenna has offered this thoughtful and challenging response:

Fischer has responded to my objections by likening the problem I raise here to generality problems in other areas of philosophy. A certain level of generality, and thus lack of specification, is acceptable in other domains of inquiry, such as ethical theory. So, too, Fischer argues, is a degree of generality acceptable when theorizing in terms of mechanisms of agency. (Fischer 2004: 169) Fischer also points out that it is unreasonable to demand of a successful theory that all of its elements are fully analysed (168). Thus, he resists the burden of offering any ‘purely “principled” account of mechanism individuation—an account that did not at some level appeal to intuition’ (167–8). He thus remains committed to relying exclusively on appeal to intuitions in response to difference [sic] cases.
While I sympathize with Fischer’s thoughtful reply, it is not enough to put the objection to rest. Requesting *some* principled basis of mechanism individuation is not the same as demanding a full analysis of mechanisms—one that is ‘purely’ principled. And a degree of generality is of course acceptable in theorizing in areas such as this one. But when the degree of generality and an *exclusive* reliance on intuition will not help adjudicate differences between cases crucial to assessing the theory, then demanding *some* principled basis for settling the dispute is only dialectically fair and reasonable.

(McKenna 2013: 180)

McKenna’s critique here, and his worries about the role of mechanism-individuation in my theory, are eminently fair and reasonable. Still, I think more can be said in defense of the sort of mechanism-based theory I have presented. McKenna is willing to concede that there can be a certain residual ‘generality’ in a theory of responsibility, and he is also willing to accept a role for intuition. But he seems to think of these components as entirely distinct. That is, he seems to think that I am somehow forced to an ‘exclusive reliance’ on intuitions in applying my theory (which involves a ‘general’ component) to cases. This would certainly be problematic; here the theory would not be doing any significant conceptual work, but all the important work would be done by intuitions about cases. The theory would just be a façade lending some (illicit) credibility to the intuitive judgments.

But I am not committed (as far as I can see) to thinking of the generality and intuition components as *separate*. Rather, I think of the level of generality of the mechanism as (p.225) in part determined by intuitions, although not ‘first-order’ intuitions about certain cases (that is, intuitions about whether some individual is—or is not—morally responsible in a particular case); thus, the reliance on intuition is not separate and entirely distinct from the issue of the level of generality or abstraction of the mechanism.
On my view, we start with some clear intuitions about cases in which individuals are morally responsible, and other clear intuitions about cases in which individuals are not morally responsible. Based on these cases, we ‘try out’ a certain intuitively plausible answer to the question of mechanism-individuation. We then apply this provisional notion of mechanism-individuation (a provisional answer to the question of how specific the specification of the mechanism must be) to other cases, adjusting it if it yields intuitively implausible results in these cases. This process of adjustment results in a notion of mechanism-individuation that is not reductively defined (insofar as there is a residuum of guidance by intuition). The key is that it has to be applied consistently across the various cases, in some intuitive sense. That is, intuition also plays a role here in testing the consistency of the level of generality or abstraction of the relevant mechanism, as the theory gets applied across a range of cases.

Similarly, how does one individuate ‘maxims’ in relation to Kant’s universalizability formulation of the categorical imperative? (A parallel question: how does one individuate moral rules, in relation to rule-consequentialism?) This is a notoriously vexing problem—but it has not diminished the interest or importance of Kantian approaches to ethics! Is it permissible to tell a lie in a certain situation? Well, what is the maxim of the action? Telling a lie? Telling a lie on Monday? Telling a lie on Monday while wearing a blue shirt? Telling a lie when no one else will? Telling a lie when no one will know about it? The level of generality or abstraction of the maxim is not somehow ‘given’ or written into the fabric of the moral universe. Rather, the proponent of Kantian universalizability might try out a level of description and see whether it yields the result it is ‘supposed to’; since Kant holds that it is never permissible to tell a lie, this should be the result of the universalizability test. Now the key is to ‘hold fixed’ roughly this view of maxim-individuation when applying the test to other kinds of actions. Although the level of specificity of maxims, for the purposes of the universalizability test, is not somehow fixed in advance, it is nevertheless important to be consistent in the application of the test; otherwise, it would not be the test, but one’s antecedent ‘first-order’ intuitions and views about what actions should pass the test, that are doing all of the work.

Consider the similar point made by Nozick:

Since we have not specified a precise theory of subjunctives or specified precisely how to identify a method and tell when it is held fixed, there is some leeway in our account. It may be this leeway that enables the account to cope with these examples and other cases, by using the constituent notions loosely and intuitively. This is not an objection but a reason to think the notions can be specified more precisely to handle the cases—a condition on their specification is that they handle the cases adequately—provided the discussion of the cases did not exploit the leeway or wobble inconsistently, first leaning in one direction, then in another.
(Nozick 1981: 193)

(p.226) The last part of the quotation gets at the worry that an ‘exclusive reliance’ on intuition is what is really doing all of the work in my theory. If what drives the approach are in the first instance intuitions about cases, and then one leans toward a narrow specification of mechanisms to accommodate some intuitions, and a wide specification of mechanisms to accommodate others, this would be objectionable.

But this is not how I conceptualize my methodology. A crucial part of my approach requires that mechanisms be individuated consistently. At this point in the development of the overall theory I must rely to some extent on intuition in evaluating whether this is indeed the case, i.e., that mechanisms are being individuated consistently in the application to various cases. Perhaps Nozick is correct that if it ‘feels’ like the application of the theory across cases involves a consistent view of mechanism-individuation, then this provides reason to hope that in principle there could be a more precise (and reductive) account of mechanism-individuation. I am, however, a bit less sanguine than Nozick about the possibility of a reductive analysis of mechanism-individuation.

In sum, I do not think it is fair to saddle me with the view that I must rely exclusively on intuitions about the moral responsibility (or lack of it) of individuals in various actual and hypothetical cases, then gerrymandering the level of generality of the mechanisms so as to fit the intuitive judgments. Intuition plays a role in evaluating the consistency of the inchoate view of mechanism-individuation involved in the theory. Of course, I am still not out of the woods. It still must turn out that an overall evaluation of the application of my theory to cases must pass the test; that is, it must seem intuitively that the level of specificity of the mechanism is consistent across cases. It would be completely legitimate for a critic to inquire more specifically into this issue; but this is a different problem than the alleged sole reliance on intuitions about cases.

As I have previously written, I am less optimistic than Nozick about the possibility of a fully reductive account of mechanism-individuation. In the end, I would not be surprised if we will have to settle for a theory of moral responsibility that leaves an important role (of the sort identified above) for intuition; it will have at least a part (or parts) that are not specified in an algorithm that employs only unproblematic (or uncontested) ingredients. McKenna prefers an agent-based to a mechanism-based approach to moral responsibility. But I shall now contend that his own preferred account does not succeed in avoiding the invocation of a component that is ‘unspecified reductively’ in just the sort of way my notion of ‘same mechanism’ is. So the complaint cannot be that mechanism-individuation is not susceptible to reductive, algorithmic explanation.
To take stock, McKenna argues that in a Frankfurt-style case, the agent (Jones) is indeed responsive to reasons, although he cannot do otherwise. Jones cannot do otherwise because of the presence and dispositions of the counterfactual intervener, Black. Nevertheless, Jones is responsive to reasons because in ascertaining an individual’s reasons-responsiveness, one attends only to intrinsic features of the agent. That is, one abstracts away from extrinsic features, such as Black. And if we subtract Black from the equation, then Jones is evidently reasons-responsive. Although McKenna agrees with (p.227) me that the agent in Frankfurt-style cases cannot do otherwise (in the relevant sense of ‘cannot’), he nevertheless maintains that the agent is reasons-responsive. We are both semicompatibilists; that is, we both hold that causal determinism is compatible with moral responsibility, quite apart from whether causal determinism rules out freedom to do otherwise. But whereas I explain moral responsibility by reference to the reasons-responsiveness of the actual-sequence mechanism, McKenna does so by reference to the reasons-responsiveness of the agent. David Brink and Dana Nelkin agree with McKenna in preferring an agent-based, rather than a mechanism-based, reasons‐responsiveness approach to moral responsibility (Brink and Nelkin 2013).

McKenna works out his specific suggestion for an analysis of Jones’s reasons-responsiveness in light of similar approaches to the analysis of notions, such as water-solubility or fragility, offered by the so-called ‘new dispositionalists’ (McKenna 2013, Vihvelin 2004 and 2013, Fara 2008, and Smith 2003; the name, ‘new dispositionalists’, comes from a penetrating article by Randolph Clarke 2009). The new dispositionalists hold that we can analyse a passive power (such as solubility or fragility) in terms of a complex set of conditionals. According to these theorists, we can thus avoid the problems posed by ‘masks’ and ‘finks’. (A mask would block the manifestation of the power in the ordinarily triggering circumstances, and a fink would eliminate the power in such circumstances.)

McKenna starts with this sort of (rough and approximate) conditional analysis of the ‘disposition’ or ‘passive power’ fragility:

If this vase were toppled, and if it retained during the relevant duration of time its intrinsic properties P₁-Pₙ, and if it were not interfered with in a way that would impede the causal efficacy of those intrinsic properties, then it would break.

(McKenna 2013: 167)
Modeling his account of reasons-responsiveness on this sort of approach, he offers the following conditional pertinent to Jones’s ‘active’ power of reasons-responsiveness in the Frankfurt-style cases (here McKenna employs the original Frankfurt-style case in which Jones shoots Smith, but nothing hangs on the difference between this version and the one I presented above in which Jones votes for a Democrat):

\[ RR_I: \text{If Jones were to become aware of } R_I, \text{ and if Jones retained during the relevant duration of time intrinsic agential properties } P_1-P_n, \text{ and if Jones were not interfered with in a way that would impede the causal efficacy of those properties, then Jones would not shoot Smith.} \]

(McKenna 2013: 168; ‘RR_I’ indicates that the conditional in question is one of a collection of counterfactuals relevant to an agent’s reasons-responsiveness)

Employing RR_I as part of a general analysis of reasons-responsiveness, McKenna contends that Jones is reasons-responsive in the Frankfurt-style case. The crucial move is to abstract away from Black, who is ‘external’ to Jones; Black’s presence is not an ‘intrinsic’ agential property of Jones. But now the problem for McKenna is that he relies on an ingredient, the distinction between intrinsic and extrinsic (agential) properties, which he does not explicitly define or explain. This is a notoriously difficult (p.228) distinction to characterize in an algorithmic, reductive fashion. There is a huge literature discussing this sort of distinction, with no clear upshot about how reductively to explain it. Two important papers in this literature are Langton and Lewis 1998 and Lewis 2000; it is noteworthy that the distinction between intrinsic and extrinsic spatial properties is parallel in important ways to the equally vexed distinction between hard (temporally nonrelational or intrinsic) and soft (temporally relational or extrinsic) facts about times. (See Fischer 1989 and Fischer and Todd 2015.)
Of course, I do not object to an account of reasons-responsiveness or moral responsibility that employs an ingredient that is not given an explicit reduction to unproblematic notions. I do not object to invoking a component that is left to some degree up to guidance by intuitions (of the appropriate kind). Here I wish simply to indicate that McKenna’s agent-based approach to reasons-responsiveness is, arguably, in the same boat with my mechanism-based approach insofar as they both employ ingredients that are not explicitly reduced to unproblematic notions. It might be that we have a better purchase on the intrinsic–extrinsic property distinction than we do on mechanism-individuation, but this is not clear, and we would need an argument for this conclusion. When one seeks to be ‘constructive’ in moral responsibility theory—or in pretty much any other area of philosophy—it is hard to avoid employing some elements that are not reducible in algorithmic fashion to unproblematic or uncontentious elements. It is perhaps a cost of doing business. In any case, the presence in one’s theory of such an element cannot in itself be a fatal flaw, even if it represents a kind of incompleteness or failure to achieve a philosophical ideal. We might again seek wisdom from Aristotle, who pointed out that we should not seek more precision in our theorizing than the data allow.

Moral Responsibility and Freedom to do Otherwise
Whereas McKenna seeks to employ resources of the ‘new dispositionalists’ (such as Fara, Vihvelin, and Smith) to give an account of reasons-responsiveness, according to which the agent in the Frankfurt-style cases is reasons-responsive, although he could not have done otherwise, the new dispositionalists employ their analyses to give an account of freedom, such that the agent in the Frankfurt cases is indeed free to do otherwise. Dana Nelkin joins the new dispositionalists in giving an account of freedom that has the result that Jones is free to do otherwise in the Frankfurt-style cases (Nelkin 2011:115). For simplicity’s sake, I shall include Nelkin in the group I refer to as the ‘new dispositionalists’; perhaps I should employ a different name (such as ‘New Alternative Possibilities Compatibilists’) for this inclusive group, but it will be simpler to use ‘new dispositionalists’ with the caveat the Nelkin departs from Smith, Fara, and Vihvelin in important ways. Although these philosophers offer different particular accounts of freedom, it will suffice here to note that they all abstract away from Black’s presence; they analyse Jones’s freedom while subtracting away Black. In Nelkin’s terminology, it is an ‘interference-free’ notion of freedom.
So the new dispositionalists offer an account of freedom that is ‘abstract’—it does not require that all features of the past and current situation of the agent be held fixed. As I have previously written, this kind of account of freedom is perfectly consistent with causal determinism, and it has thus not been my focus of attention. I agree that agents have various kinds of abilities to do otherwise in the Frankfurt-style cases, including general abilities. I further agree that they possess the kind of ability to do otherwise identified by the new dispositionalists; this might be thought to be an ability or freedom that is ‘in-between’ a general capacity and the incompatibilists’ preferred notion of freedom, which holds everything about the past and current situation fixed. Just as there are no worries about the compatibility of general abilities with causal determinism, there are no worries about the compatibility of the new dispositionalists’ notion of freedom with causal determinism. If one is primarily concerned to defend the compatibility of causal determinism and moral responsibility, one can embrace the contention that, in order to be morally responsible, an agent must possess the ability to do otherwise, as conceptualized by the new dispositionalists. Not a problem!

But now we have two routes to compatibilism about causal determinism and moral responsibility. On one approach, moral responsibility does not require freedom to do otherwise—this is the actual-sequence model. On the other approach, moral responsibility does require freedom to do otherwise. Of course, the relevant freedom is interpreted differently by proponents of each route to compatibilism. On the first route, freedom is interpreted as requiring that all facts about the past and current situation be held fixed; on the second route, the fixity requirements are not so stringent, and one does not hold fixed the presence and inclinations of Black. Is there some reason to prefer one approach to another? Why should we adopt an actual-sequence model, rather than the new dispositionalists’ strategy? Indeed, the two approaches can seem to be mere ‘notational variants’ of each other.

To explain. On my actual-sequence model, I contend that Jones is not free to do otherwise (McKenna agrees). In evaluating Jones’s freedom to do otherwise (in the relevant sense), I hold fixed the presence and inclinations of Black. However, I also contend that Jones’s actual-sequence mechanism is appropriately reasons-responsive, and thus he can be morally responsible. In evaluating the responsiveness characteristics of the actual-sequence mechanism, I abstract away from Black—he is not part of the actual-sequence mechanism. (McKenna abstracts away from Black in evaluating Jones’s [agential] reasons-responsiveness, and thus gets to the conclusion that Jones is morally responsible in the Frankfurt-case, despite lacking freedom to do otherwise.)
The new dispositionalists do not hold fixed Black’s presence and dispositions in evaluating Jones’s freedom to do otherwise. So they get to the very same conclusion (that Jones is morally responsible in the Frankfurt cases), and they do so by abstracting away from Black. The only difference between the two approaches is that the actual-sequence theorist abstracts away from Black when ascertaining reasons-responsiveness (either of the agent or mechanism), and the new dispositionalists abstract away from Black when evaluating whether Jones has freedom to do otherwise. But this might seem to be a superficial difference. At a deep level, both approaches deem Jones morally responsible in the Frankfurt cases because Jones would act differently in certain scenarios in which Black is out of the equation. Why prefer (say) an actual-sequence approach to moral responsibility? Aren’t the two approaches only superficially different?

This is a difficult question, but I think at least some progress can be made toward answering it. The first step here is to note that it is desirable to have a theory of moral responsibility that fixes on features that ground or explain moral responsibility, not merely features that are necessary conditions for moral responsibility. It might be that freedom, as interpreted by the new dispositionalists, is a necessary condition of moral responsibility. But in my view the presence of such freedom is not part of the best explanation of an agent’s moral responsibility. Rather, the best explanation of moral responsibility—what grounds moral responsibility—will fix on features of the actual sequence. What explains an agent’s moral responsibility is that she makes a certain kind of statement (I explain this point in Fischer 1990), not that she makes a difference (of a certain kind). After all, interpreted naturally and straightforwardly, Jones does not make a difference in the Frankfurt-style case, but he does make a statement (in an important sense) (Fischer 1990).

The new dispositionalists accept a kind of alternative-possibilities requirement on moral responsibility, and thus, I believe, it is natural to suppose that they believe that moral responsibility rests on the capacity to make a difference to the world. But if one wants one’s theory to track closely what explains or grounds moral responsibility, the new dispositionalist will have to say that the freedom to do otherwise (interpreted as in new dispositionalism) does in fact explain or ground moral responsibility. But this just seems implausible to me. That is, it is implausible to explain Jones’s moral responsibility for voting for the Democrat by contending that he selected this path rather than a path in which he voted for a Republican, where both of these paths were genuinely available to him. On the new dispositionalists’ approach, this is precisely what would explain Jones’s moral responsibility: Jones is selecting from among alternative ways the world could go, where the ‘could’ is interpreted via abstraction away from Black. On this view, Jones is indeed making a difference—he is selecting from among paths that are available to him. It is in virtue of this selection that he makes a difference and is thus morally responsible.
But this picture just seems incorrect to me, even though it is not straightforward to say why. It does not seem to me that Jones is best described as selecting from among paths the world could genuinely take—either the path in which he votes for the Democrat or the path in which he votes for the Republican. This kind of selection cannot explain why Jones is morally responsible—he doesn’t really select from among such alternatives. He does not know about Black, so he believes that he can vote for either the Republican or the Democrat. So, he makes a selection from among ‘epistemic possibilities’—possibilities which are not ruled out, for all Jones knows. But it is infelicitous to conceive of Jones as selecting from metaphysically open options, thus making a difference, and thereby being morally responsible—or so it seems to me.

(p.231) Perhaps the point could be put as follows. The proponent of the alternative-possibilities model (presented by the new dispositionalists) thinks that Jones makes a difference to how the world goes in the sense that he selects a path in which he votes for the Democrat, rather than a path in which he votes for the Republican, where both paths are open to him. But, unbeknownst to Jones, if he were about to choose to vote for the Republican, he would be blocked from so choosing (and acting on such a choice). This fact—the fact that Jones would be blocked from choosing to vote for a Republican or acting on that choice—would seem to etiolate the claim that Jones really has a power to make a difference to the world, to select a path along which he votes for the Democrat, rather than a path along which he votes for the Republican. And this is so, even though the new dispositionalist holds that Jones has the freedom in a Frankfurt-style case to vote for the Republican. This suggests that the ‘freedom’ identified by the new dispositionalist does not correspond to the notion of ‘power’ in the idea that Jones has a power to make a difference to the world in the sense that matters for the grounding of moral responsibility.

Think of it this way. Suppose Jones reflected on his behavior and proclaimed, ‘In this context, I was a difference-maker. Through my choice and action, I made a difference to the way the world unfolded—I made the difference between the world’s taking the Jones-votes-for-the-Democrat path and the world’s taking the Jones-votes-for-the-Republican path.’ It would be natural to be skeptical, and to point out (gently, of course) that Jones is being rather grandiose here. After all, he really didn’t make the indicated difference to the way the world unfolds, since if he were about to choose to vote for the Republican, he would be occluded from doing so (and from acting on such a choice). Perhaps Jones makes a difference—but not this difference, and it is this difference that is putatively relevant to the explanation of his moral responsibility.
It is better to think of Jones as making a statement, rather than making a difference. (Although I do not have space to elaborate on this and the following claims here, please see Fischer 1990.) Insofar as he acts from his own, suitably reasons-responsive mechanism, he makes a statement: he writes a sentence in the narrative of his life. In acting freely, agents transform the chronicles of their lives into genuine narratives or stories (strictly speaking). That is, agents who act freely have life-stories that can evoke prototypical emotional reactions in appropriate audiences, and that are evaluated in part via a consideration of narrative relationships (and not just a function that adds up the welfare of the moments of the life). As the author of this narrative, the agent makes a statement, even if he does not make a difference. I have sought to develop these rudimentary ideas a bit more fully elsewhere (Fischer 2009). Here, unfortunately, given space constraints, this will have to suffice.

I certainly do not think that I have a knockdown argument for preferring an actual-sequence approach to the new dispositionalists’ approach. Rather, my argument here has been tentative and merely suggestive. I hold that it is desirable to have a theory of moral responsibility that gives not just necessary (and sufficient) conditions, but closely tracks what grounds (or explains) moral responsibility. And I have further suggested (p.232) that selection from among metaphysically available options—making a difference—is not a felicitous way of explaining moral responsibility in general (as illustrated by the Frankfurt-style cases). In contrast to the alternative-possibilities model (and the new dispositionalism), the actual-sequence approach to moral responsibility has it that what explains an agent’s moral responsibility is not selection from among available paths, but freely adding to the narrative of one’s life. In acting freely, one writes a sentence in the narrative of one’s life—one makes a statement, even if one doesn’t make a difference.

Conclusion
I have sought to defend a reasons-responsiveness model of moral responsibility. I am in this respect deeply indebted to Irwin’s interpretation of Aristotle. I depart from Irwin’s Aristotle in at least two important ways. First, I adopt a mechanism-based account of moral responsibility, that is, an account that relativizes the issue of reasons-responsiveness to mechanisms (rather than agents). This is why I previously explained that my theory is the offspring of Irwin’s Aristotle and Nozick. In adopting a mechanism-based model, one can embrace the deep insights of certain epistemologists—and the parallel morals about agency.
Second, I adopt an ‘actual-sequence’ model of moral responsibility, according to which moral responsibility does not require freedom to do otherwise. I thus should also mention another important part of my ancestry: Chrysippus, Locke, and Frankfurt. Embracing an actual-sequence model of moral responsibility gives one an elegant answer to the question of what grounds moral responsibility: it is not that an individual makes a difference, but that she makes a statement.²

References

Bibliography references:


The Freedom Required for Moral Responsibility


Notes:

(1) Irwin’s pathbreaking interpretation of Aristotle’s account of moral responsibility continues to be influential (and controversial) today. It is the subject of sustained critical discussion in (Echenique 2012), and, more recently, in (Cooper 2013).

(2) I am very grateful to helpful comments on a previous version of this essay by David Beglin, Benjamin Mitchell-Yellin, Dana Nelkin, and Susan Sauvé Meyer.