

Responsibility and Reason: Defending an Asymmetrical View¹

Dana K. Nelkin

(To appear in Pacific Philosophical Quarterly 2008)

In this paper, I defend a view according to which one is responsible for one's actions to the extent that one has the ability to do the right thing for the right reasons. The view is asymmetrical in requiring the ability to do otherwise when one acts badly or for bad reasons, but no such ability in cases in which one acts well for good ones. Despite its intuitive appeal, the view's asymmetry makes it a target of both of the main camps in the debate over responsibility. In addressing objections, I explore the relationship between fairness and responsibility, and the nature of the ability to do otherwise.

I. Introduction

Does one need the ability to do otherwise to be responsible for one's actions? Some answer this question in the affirmative.² Others argue that the ability to do otherwise is never required for responsibility.³ This might seem to exhaust the reasonable answers to our initial question, but, interestingly, there is a set of views, worth taking seriously, that answers the question differently: "sometimes yes, and sometimes no."

I will argue for a view of this third kind in what follows. In particular, I will argue for what I call the "rational abilities view," according to which people are responsible when they act with the ability to do the right thing for the right reasons, or a good thing for good reasons.⁴ The ability to act on good reasons in turn requires both an ability to recognize reasons and an ability to act on them. And these abilities might in turn depend on having a variety of other abilities, including perceptual and emotional ones.⁵ The appeal of the view is manifested across a wide

range of our practices and judgments. For example, we believe that children gradually acquire responsibility as they gradually acquire rational abilities of various kinds, and those who are severely mentally impaired are exempt from responsibility for their actions. I believe that its appeal is manifest in the judgments of responsibility codified in a variety of legal systems, as well. To take just one example, consider the insanity defense. There has been a great deal of controversy among legal scholars about just how to understand it, and I will focus briefly on an attempt to capture the basic idea in the Model Penal Code, an ideal system of laws proposed by the American Law Institute in 1962. According to the Model Penal Code, insanity is defined as a lack of substantial capacity to control one's behavior. In turn, substantial capacity is defined as: “the mental capacity needed to understand the wrongfulness of [an] act, or to conform...behavior to the...law.” The idea is that people are not responsible for their actions when they lack either the capacity to grasp reasons for acting (or not acting, as the case may be) or the capacity to translate those reasons into action (or omission). (One can see the history of the legal debate as an attempt to capture both cognitive and volitional abilities to respond to reasons.) In general, when we find out that someone lacks a capacity for recognizing reasons or cannot control her behavior in light of them, we are tempted to excuse a person for her actions. Otherwise, in both legal and non-legal contexts, we tend to treat people as responsible for their actions, praising and blaming them.⁶

In fact, the rational abilities view seems so natural that it is somewhat surprising that it is not more widely accepted among philosophers who put forward conditions for responsibility. One reason for this phenomenon might be that when we delve under the surface, we very quickly come up against worries concerning the view's asymmetry and its implications in a deterministic world, some of which I will address in what follows.⁷

Now a number of philosophers have recognized the importance of rational capacities or powers to responsibility. For example, some defend the idea that general rational capacities are required for responsible action (but not necessarily the ability to exercise them on each particular occasion one is responsible). R. Jay Wallace (1994, 183), for one, writes that “what matters is not our ability to exercise our general powers of reflective self-control, but simply the possession of such powers...”.⁸ And according to John Martin Fischer and Mark Ravizza’s influential view, one must act on a mechanism that is responsive to reasons.⁹ But few defend what I see as the more literal reading of the claim that one must be able to do the (or a) right thing for the right (or some good) reasons. Susan Wolf is a notable exception, articulating what I believe is a view implicit in many of our practices, including legal ones. According to Wolf (1990, 69), one must “have the ability to do otherwise” in some cases; in particular, one must have the ability to exercise one’s general rational powers in cases in which one does not do so. And it is this sort of view that leads to the conclusion that the ability to do otherwise is sometimes required and sometimes not required for responsibility.

Although perhaps not obvious at first, views of this sort are asymmetrical in an important way, and it is this asymmetry that explains why the ability to do otherwise is required for responsibility in some circumstances and not in others. The asymmetry of this view reveals itself in examples like this one: suppose that you have promised to help a friend during a difficult time, and that following through is the right thing to do. If you follow through on your promise because you made the promise and value helping your friend, then you have the ability to do the right thing for the right reasons, and thereby meet the relevant condition for responsibility. If, on the other hand, you fail to keep your promise, then whether you are responsible for your actions depends on whether you have the ability to do what you fail to do. So in the case in which you

do the right thing for the right reasons, no ability to do otherwise is required; in the case in which you do not do the right thing for the right reasons, such an ability is required. As I will try to show, I think this is a very common-sense view, and quite pervasive in our own practices.¹⁰

At the same time, the asymmetrical character of the view has been the target of serious criticism, and indeed it does seem odd on the face of it that sometimes an ability to do otherwise would be required for responsibility and sometimes not.¹¹ There are at least two general types of criticism that can be (and have been) leveled against the rational abilities view. The first is a general sort that targets all views that allow for responsibility in a world in which we cannot do otherwise. Since the rational abilities view allows for responsible action in such worlds, at least in the case of right actions done for right reasons, it is vulnerable to general objections of this sort. A second sort of criticism targets the rational abilities view, in particular, in virtue of its asymmetrical character. According to one version of this kind of objection, the asymmetry of the rational abilities view has a particularly counterintuitive result if our world is one in which we cannot do otherwise. For it would turn out that if our world is indeed such a world, then although we can be responsible for our right actions done for right reasons, we cannot be responsible for any wrong actions. This naturally suggests that we could be praiseworthy, but never blameworthy, and while this may be a *welcome* result to some, it is counterintuitive at best. Of course, much hinges on what our world is like for this objection to have its full force. But even if we set aside questions about what our world is like, one might still object to the asymmetrical aspect of the view.

In fact, the asymmetry of the rational abilities view invites criticism from both of the main camps in the responsibility debate on this point. On the one hand, many incompatibilists—those who believe responsibility and determinism are incompatible—object to

the way the view treats *good* actions that are done for the right reasons. In its simplest form, the worry is that there must be at least an additional requirement for being responsible for such actions, namely, the ability to do otherwise. On the other hand, many compatibilists in turn object to the way the view treats *bad* actions or actions done for the wrong reasons, and worry that it is too much to require the ability to do otherwise to be responsible for such actions. In this way, the rational abilities view is under attack from opposite directions.

In this paper, I will explore what I take to be an especially strong version of each of these kinds of objections, and will argue that the rational abilities view has significant resources to answer them. The first, considered by Gary Watson, suggests that the rational abilities view appears plausible only insofar as we fail to distinguish between two notions of responsibility, and goes on to raise deep questions about the relationship between fairness and responsibility. The second, raised initially by John Martin Fischer and Mark Ravizza, and elaborated by Derk Pereboom among others, calls for examination of a wider set of intuitions that appear to undermine the plausibility of asymmetry, and ultimately forces us to confront questions concerning the nature of “ability” in this context. While my focus here will be on answering these two objections to the asymmetry of the rational abilities view considered in itself, doing so will have some implications for objections of the other sorts, and also for competing symmetrical views that each share *one* of these important challenges with the rational abilities view.

II. Two Faces of Responsibility and Unfairness

In “Two Faces of Responsibility”, Watson distinguishes between two notions of responsibility, “responsibility as attributability”, on the one hand, and “responsibility as

accountability,” on the other. Watson draws a number of conclusions after drawing this distinction, contributing to an extended objection to the rational abilities view.

To see how, we need an explication of this key distinction between two sorts of responsibility. Let us focus first on blameworthiness, as Watson does. There are two kinds of blame. An example of the first is the judgment that a person’s stealing of books is “shoddy.” Following Van Inwagen, Watson notes that this is a judgment that even a philosopher who is sincerely skeptical about moral responsibility can make. Or to take another example, suppose that “someone betrays her ideals by choosing a dull but secure occupation in favor of a riskier but potentially more enriching one, or endangers something of deep importance to her life for trivial ends...then she has acted badly—cowardly, self-indulgently, at least unwisely.”¹² These sorts of judgments attribute moral faults to agents, and to make these judgments just *is* to blame the agents in this sense. This is a kind of appraisal that concerns agents’ virtues and vices and so Watson calls it an appraisal from the *aretaic* perspective. The blame in question is blame in the “attributability” sense. In contrast, when we blame someone in the “accountability” sense, we invoke moral norms to which we hold people, and we treat them as deserving of sanctions for their violation. Notable among those sanctions are the reactive attitudes such as resentment. The examples Watson gives are meant to illustrate the possibility of being responsible, and even blameworthy in the attributability sense, while not in the accountability sense. In the case of the woman who betrays her ideals, we blame her, but do not necessarily think she is accountable to anyone for her actions. In the case of the book stealer, it seems at least coherent for the skeptic about responsibility in the accountability sense to blame someone who has stolen his books in the attributability sense.¹³

With this distinction in hand, Watson draws a number of related conclusions including the following: first, attributability is no small matter. While some have thought that this sort of appraisal is no different in kind from our appraisal of, say, car engines, it can be a kind of deep appraisal itself. This is best seen if we consider a class of views Watson calls “self-disclosure views”, those according to which responsible actions are those actions that “express” ourselves in some sense.¹⁴ Watson argues that if we think of self-disclosure views as embodying the conditions for attributability, then we can see that being responsible in the attributability sense allows for a kind of important, and deep, appraisal. For example, Watson (1992/2004, 271) writes, “To adopt an end, to commit oneself to a conception of value in this way, is a way of taking responsibility. To stand for something is to take a stand, to be ready to stand up for, to defend, to affirm, to answer for.” To be responsible in this way, while not entailing responsibility in the accountability sense, is nevertheless to be responsible in an important way. To praise people for their commitment to certain values is of a different order than positively evaluating a car engine, according to Watson.

Second, issues concerning avoidability, or the ability to do otherwise, arise in connection with responsibility as accountability, and not with responsibility as attributability. Since accountability entails holding people to standards associated with sanctions, it seems it would be unfair to do so unless people have a reasonable opportunity to avoid those sanctions. Thus, the question of avoidability arises from the perspective of accountability.¹⁵

Third, the asymmetrical rational abilities view gains in plausibility because of a blurring of this distinction between two notions of responsibility. Recall that on that view, one is responsible to the extent that one acts with the capacity to respond to good reasons. It turns out then, that when you do so respond, you thereby display the relevant capacity, even if you

couldn't have done otherwise. So, for example, consider a woman who jumps into the waves in order to save a drowning child, keenly focused on the need of the child and the great good of her life that is at stake. Now even if she couldn't have done otherwise, she is responsible and praiseworthy. (Following Watson, call this woman Rosa.) On the other hand, if someone does not respond to relevant reasons (for example, by not jumping in to save the child) then whether she is responsible or not will depend on whether she could have done otherwise. If she could not have done otherwise, she lacks the relevant ability, and so is not responsible. In short, the ability to do otherwise functions in an asymmetric way—you need it if you act badly, but not if you act well. According to Watson, to the extent that this is plausible, it depends on a slide from one sort of responsibility to another. When we agree that Rosa, the woman who saves the child and can't do otherwise, should be praised, we are thinking in terms of the aretaic sense. But, in contrast, when we withhold blame in the case of, say, the victim of the deprived childhood who fails to act well, we have moved to the accountability sense of responsibility. (Call this second woman, Joan.) Note that we can still find Joan to be selfish and mean-spirited, and so blameworthy in the attributability sense, while denying that she is responsible in the accountability sense.

This “diagnosis” of our intuitions provides an initial challenge for the rational abilities view. There is a variety of ways a defender of the rational abilities view might respond at this point, including an appeal to the intuitive plausibility of the view itself. Further, as Watson notices, his account actually *makes room* for a certain asymmetry when it comes to praise and blame, and it might seem tailor-made to support an asymmetrical view when it comes to avoidability. First we can note that our vocabulary for talking about blame is much richer than that for talking about praise (e.g., there is no counterpart to ‘holding responsible’ for good actions). Second, this

corresponds to a second asymmetry: much more seems to be at stake when it comes to blame than what is at stake when it comes to praise. “Hence,” Watson (1992/2004, 284) writes,

considerations of fairness might still support an asymmetry solely within the perspective of accountability. For if the requirement of avoidability derives from the idea that we should not be made to suffer from sanctions which we had no reasonable opportunity to avoid, then the requirement will have no relevance to the conditions of appropriate praise. The special objection to responding adversely to those who could not do otherwise simply does not apply to the case of favorable treatment.

No worries about fairness of sanctions, then, arise in the case of praise, and so, perhaps being praiseworthy in the accountability sense does not require avoidability.

All looks well for the rational abilities view. But, alas, Watson identifies a different worry that remains unanswered.¹⁶ Again, the problem concerns fairness, but this time it is a matter of *interpersonal* fairness. Suppose that Rosa is praised for her unavoidable action in saving the child. Joan, who failed to act well but couldn’t help it, can complain that a system that benefits Rosa and not Joan is unfair. After all, Rosa is no more deserving of benefit than she is.

In what follows, I will focus on two key assumptions in the reasoning set out so far: (i) Fairness in terms of assigning rewards and sanctions is associated with accountability and not with attributability, and (ii) Rosa deserves her rewards no more than Joan, and the reason is that neither one can do otherwise.

Let us briefly consider the first assumption first: fairness in terms of assigning rewards and sanctions is associated with accountability. I want to press the idea that fairness and

accountability can come apart, at least to some extent, in both directions. It seems possible to have a fair system of sanctions—understood simply as deprivations of some sort—even if no one is responsible in the accountability sense for anything. It could at least be not *unfair* to lock people up if they act badly in the attributability sense, that is, if they violate moral norms, like killing or torturing. What would be unfair is anything that purported to be *retributive* sanctions. But just plain sanctions would seem perfectly acceptable under some circumstances. Fairness of sanctions, then, does not entail accountability, as long as we do not think of sanctions as having a built-in retributivist component. (This raises a challenge to the claim that the appropriateness of sanctions must require avoidability, and here we skip over accountability altogether.)

Moving in the other direction, *must* accountability blame and praise be understood in terms of the fairness of providing sanctions and rewards? There certainly seem to be important connections here, but the possibility of a Gandhi-like figure who does not believe we ought to respond with the reactive attitudes or with punishment, raises an important set of questions, raised by Watson himself in other work.¹⁷ Must such people abandon the notion of accountability altogether, in the sense of thinking that they themselves and others are obligated to meet certain standards? If the link between accountability and the fairness of sanctions *could* be broken in this way—so that accountability did not *by itself* entail fairness of sanctions—then the claim that avoidability is required for fairness of sanctions could simply fail to apply to its target.¹⁸ This is a complex issue, and I will set it aside for now in order to focus on another.¹⁹

Let us turn then to the second assumption in Watson's argument. There is reason to doubt that the complaint of unfairness that arises in the reasoning about Rosa and Joan is one that targets asymmetrical views in particular (or even the more general category of views that accept some circumstances in which responsibility does not require the ability to do otherwise). To see

why, suppose that a third woman, call her Sylvia, acts well, and yet could have acted badly, too. She has all the capacities one could hope for, and more. Still, Joan is stuck with only the capacity to act badly. Could Joan not still make a charge of unfairness, just as she did in connection with Rosa? It could go like this: “Here is Sylvia, getting all sorts of kudos for saving a child, and I, Joan, didn’t even have that opportunity.” If so, then the charge of unfairness made against Rosa does not hinge on the idea that she gets the benefit of praise despite lacking the ability to do otherwise; rather, the charge is more general than that: Joan happens to lack *whatever* capacities are required for reward (whether they include the ability to do otherwise or not). But the problem isn’t with a particular account of responsibility; rather it appears to be with any account.

One response to Joan’s complaint would be to say that Joan’s situation exemplifies the problem of moral luck. But then Joan’s complaint, albeit quite serious, does not make essential appeal to the requirement of avoidability. Another response is to say that Joan’s complaint of unfairness does not impugn our judgment of responsibility on either Rosa’s or Sylvia’s part. It is unfair that they get to be accountable agents (with all the benefits and sanctions that go along) and she doesn’t, but there is no incoherence in saying this; there is no contradiction in attributing accountability while acknowledging unfairness in its distribution. Further, it shows that this interpersonal complaint of unfairness need not be especially problematic for views that do not require avoidability for responsibility, since the same complaint applies equally to those views that do require avoidability.

If a special sense of unfairness lingers when one thinks about Rosa benefiting when she couldn’t have done otherwise, then it seems to come from an *antecedent* reluctance to see Rosa as accountable. One already judges that Rosa is not accountable, and then reasons to the

unfairness of rewarding her and not Joan. But then the initial intuition that Rosa is not accountable is what is really doing the work. In that case, a defender of the asymmetrical view could say, “I have located a genuine asymmetry when it comes to *non*-interpersonal fairness. And now we can discuss our conflicting intuitions about the particular case.”

However, we cannot yet rest easy that all considerations about unfairness have been accounted for. For there is one more set of cases that raises a question of interpersonal fairness.²⁰ So far, we have considered a question of interpersonal fairness between Rosa and Joan, and I argued that the same charge could be made by Joan against Sylvia, as well, so that any unfairness that arises for Joan does not arise in virtue of Rosa’s *inability* to do otherwise. Now let us explore the relationship between Rosa and Sylvia directly. Suppose that Rosa, who cannot do otherwise, and Sylvia, who can, each save one of two simultaneously drowning children. Can Sylvia justifiably complain about having to share kudos or a prize, with Rosa? If so, it seems that there is a kind of interpersonal unfairness associated with Rosa’s rewards, given her inability to do otherwise.

This is an interesting and challenging case. But I believe that filling out the details of Rosa’s inability can undermine any initial sense of unfairness one might feel. For, by hypothesis, Rosa acts precisely for the same clear reasons that Sylvia does: she sees a child drowning and realizes that the child’s death is an easily preventable (though inconvenient) and a terrible, terrible thing. Recognizing all of this, Rosa jumps in in order to prevent the child from dying. Do we still feel that we would begrudge her a share of the kudos or prize in this case? One reason why it could, for some, seem that there would be unfairness in this case is that we understand Rosa’s not being able to help it as a kind of “automatic” response, or a psychological compulsion. But when we think about Rosa’s reaction as arising from her recognition of the

powerful reasons there are for her acting in the way that she does, I believe such an initial response loses at least some of its strength.

To the extent that a sense of unfairness remains, it might be because of another unstated background assumption, namely, that what Sylvia did was in some relevant sense *harder* than what Rosa did. One might think that the fact that Rosa could only do the one thing, while Sylvia had a genuinely causally open choice, suggests that what Sylvia did was harder. Sylvia, unlike Rosa, had to overcome the temptation to pursue, let us suppose, a less inconvenient opportunity. Here, I believe, we face two questions: (a) Does difficulty matter to one's praiseworthiness? And (b) Is Sylvia's task more difficult than Rosa's? While I do not believe that the answer to (a) is at all obvious, I want to focus here on (b). The reasoning for an affirmative answer given above, which I believe is very natural, is not in the end convincing. Why does the mere ability to do otherwise reveal difficulty, and, conversely, why does the lack of ability to do otherwise reveal ease? Answering these questions depends on getting clearer about what is meant by "degree of difficulty". One natural way of understanding it is this: amount of effort and/or sacrifice required. But on this understanding, it may very well be that Rosa's act calls for an equal degree of difficulty. Perhaps she had to work extremely hard (perhaps even harder than Sylvia if she were a less fluid swimmer) to reach the child she saved. Now there might be other interpretations of difficulty that build in "resistance to a genuinely causally open alternative". But since so resisting an alternative might be quite easy in the first (effort) sense, we need some non-question begging reason for requiring this condition. Until we have such a reason, it seems that an initial sense of interpersonal unfairness between Rosa and Sylvia is left without support, and indeed, is undermined by a fuller description of the case.

Now the “unfairness” objection we have been considering is ultimately about the verdict of the rational abilities view on the “good” cases, cases in which one acts well and for good reasons. The bottom line of all variations of the unfairness objection is that you might need the ability to do otherwise for both good and bad actions to be responsible in the accountability sense. Thus, far, I have raised questions about each variation. Now let’s turn to the opposite complaint—the view gets it wrong in the “bad” cases and you don’t need the ability to do otherwise in either “good” or “bad” cases.

III. Heroes, Villains, and Abilities

Fischer and Ravizza agree with advocates of the rational abilities view that one can be responsible for acting well without the ability to do otherwise, but disagree with their claim that one cannot be responsible for acting badly without such an ability. That is, they reject the asymmetry that is built into the rational abilities view. Their objection depends largely on appealing to “Frankfurt-style” cases, arguing that such cases work for cases of good and bad actions, showing that we can be responsible for *either* kind, despite lacking the ability to do otherwise.²¹ Their first case is called “Hero,” and in it a woman, Martha, is walking along a beach when she sees a child struggling in the water. As they write, “...she quickly deliberates about the matter, jumps into the water, and rescues the child.” Had she considered *not* saving the child, “she would have been overwhelmed by literally irresistible guilt feelings which would have caused her to jump into the water and save the child anyway.” (1992, 376). Intuitively, Martha is morally responsible, even though she could not have done otherwise. Her disposition to guilt feelings played no role in what she did, despite making it impossible for her to do otherwise.

Now consider a second case, “Villain”. Joe is an evil man who knows that a child watches the sunset at the end of a long pier every day. Joe has decided to push the child off of the pier, causing her to drown. Max is just as evil.

Max is pleased with Joe’s plan...but Max is a rather anxious person. Because Max worries that Joe might waver, Max has secretly installed a device in Joe’s brain which allows him to monitor all of Joe’s brain activity and to intervene in it, if he so desires. This device can be employed by Max to ensure that Joe decides to drown the child and that he acts on this decision; the device works by electronic stimulation of the brain. Let us imagine further that Max is absolutely committed to activating the device to ensure that Joe pushes the child should Joe show any sign of not carrying out his original plan. Also we can imagine that there is nothing Joe could do to prevent the device from being fully effective if it is employed by Max in order to cause Joe to push the child into the treacherous surf.

In fact, Joe does push the child off the pier on his own, as a result of his original intention. He does not waver in any way. Max thus plays absolutely no role in Joe’s decision and action...” (1992, 377).

Joe seems to be morally responsible for his actions, just as Martha is for hers. In neither case does what prevents them from doing otherwise actually play any role in causing an action. Thus, Fischer and Ravizza conclude, “Wolf’s asymmetry thesis is false; rather, good and bad actions are symmetric with regard to the requirement of alternative possibilities for moral responsibility” (1992, 377).

Interestingly, a proponent of the rational abilities view could accept Fischer and Ravizza's intuitions about the cases without rejecting the asymmetry in question. The reason is that we can say that while Joe in Villain and Martha in Hero both lack the ability to do otherwise in some sense, they both *have* such an ability *in the relevant sense*. As Wolf (1990, 110) herself understands the possession of an ability, an agent has an ability to X if (i) the agent possesses the capacities, skills, talents, knowledge and so on which are necessary for X-ing, and (ii) nothing interferes with or prevents the exercise of the relevant capacities, skills, talents and so on. Since in Villain, Max's presence actually plays no role in Joe's action, and, we can presume, Joe has the talents and skills to refrain from pushing the child off the pier, Joe's ability to so refrain is intact, despite Max's presence. Similarly for Martha in Hero. Now of course, there is *a* sense of "ability to do otherwise" in which both Martha and Joe lack such an ability. Joe will push the child off the pier, even if he wavers, and Martha will save the child even if she considers not saving the child. In this sense, having an ability to do X is precluded when it is inevitable that the agent will not do X (call this the "inevitability-undermining" sense). But in another sense, both have their abilities to do something different from what they do intact.

Fischer and Ravizza recognize that the success of their argument hinges on what notion of "ability" is at stake. Yet, they argue, even on Wolf's understanding of "ability", Joe lacks the ability to do otherwise "for were he to try to [do the right thing for the right reason], Max's device would prevent his exercising the relevant capacities, skills, etc., required to refrain from pushing the child into the water" (1992, 378, note 9). I believe that this is a misreading of Wolf's characterization. What is needed to remove one's ability to do something in the relevant sense (call it the "interference-free capacity") is either the removal of the capacities, talents, skills, and so on (the presence of which is not in dispute) or the interference or prevention of the

exercise of those capacities. The fact that Max's device *would* interfere or prevent such an exercise in counterfactual circumstances does not entail actual interference or prevention.

For this reason, Fischer and Ravizza's Frankfurt-style case does not constitute a decisive objection to the asymmetry of the rational abilities view. It does raise the important question of how we should understand "ability" in this context. But particularly given our judgments of responsibility in Hero and Villain, we have more reason to think that responsibility is given by the conditions that allow for Hero and Villain to be responsible rather than ones for which they are not. To some extent (although not entirely), advocates of the view get to explicate the notion at issue,²² and a sense of ability according to which both Hero and Villain are responsible would be the preferable choice. At the same time, it must be acknowledged that even if the Hero and Villain case does not undermine the rational abilities view built on the interference-free capacity conception of an ability, other Frankfurt-style cases might.

A case proposed by Derk Pereboom might seem to be just such a case. The first part of the case proceeds as follows: Joe is considering whether to claim a certain tax deduction, all the while knowing that it would be illegal, but that he would probably not be caught and convicted. Joe has a

very powerful but not always overriding desire to advance his self-interest no matter what the cost to others, and no matter whether advancing his self-interest involves illegal activity. Furthermore, he is a libertarian free-agent. Crucially, his psychology is such that the only way that in this situation he could fail to choose to evade taxes is for moral reasons...In fact, it is causally necessary for his failing to choose to evade taxes in this situation that a moral reason occur to him with a certain force. A moral reason can occur

to him with that force either involuntarily or as a result of his voluntary activity...However, a moral reason occurring to him with such force is not causally sufficient for his failing to choose to evade taxes. If a moral reason were to occur to him with that force, Joe could, with his libertarian free will, either choose to act on it or refrain from doing so (without the intervener's device in place)...But to ensure that he chooses to evade taxes, a neuroscientist now implants a device which, were it to sense a moral reason occurring with the specified force, would electronically stimulate his brain so that he would choose to evade taxes. In actual fact, no moral reason occurs to him with such force, and he chooses to evade taxes while the device remains idle (2001, 19).

It seems that Joe is responsible, despite lacking the ability to do otherwise. Does Joe lack that ability to do otherwise *in the interference-free capacity sense*? Not obviously. It is plausible to say that Joe has the relevant capacities to do the right thing, despite failing to do so. Is there interference? It seems not; the device in Joe's brain functions much like Max in the Villain case in that it never does anything. It is only "waiting" to act, so to speak, in the event that it is triggered.

At this point, however, Pereboom suggests a modification to the case that he claims shows that Joe does in fact lack the interference-free capacity to do the right thing:

Suppose that a patient has a tumor that puts pressure on his brain so that he can no longer do cutting-edge mathematics. If the tumor were not putting pressure on the brain, he could do the mathematics. But imagine that it is causally impossible to remove the tumor, or for its existence to cease in any other way, without the patient dying. Then, it

would seem, he does not have the capacity, free of interference, to do cutting-edge mathematics. Analogously, suppose that in Tax Evasion the intervener has implanted his device in Joe's brain, which is triggered by the requisite level of attentiveness to moral reasons, but she has also made it causally impossible to remove or disable the device without killing him. As a result, he permanently cannot choose to refrain from evading taxes. Under these circumstances Joe would appear not to have the capacity, free of interference, to choose to refrain from evading taxes. But still, it seems he could be morally responsible.²³

If this is right, then Joe is a counterexample to the rational abilities view; one can act badly and responsibly without being able to act well, even in the interference-free capacity sense. However, this example depends heavily on the use of an analogy to a tumor, and it is here that we can begin to resist the reasoning. In the tumor case, it is claimed, we think of the mathematician as lacking the ability to do high-level mathematics even in the interference-free capacity sense. If we think of the implanted device as relevantly similar to the tumor, we should draw the parallel conclusion in Joe's case.

We need more detail about the tumor case in order to see how the parallel is supposed to work. In general, I believe, we think of tumors as destroying brain tissue, and so destroying the capacities and skills needed to engage in certain mental activities. In fact, many brain tumors operate in this way, destroying brain cells directly or indirectly by creating inflammation, compression or swelling of brain tissue.²⁴ A less realistic scenario would be one in which the tumor is more akin to the counterfactual intervener in the Villain case—as “waiting in the wings” so to speak. In this scenario, the mechanisms for doing mathematics would remain completely untouched; but somehow if they were engaged, the tumor would ensure that they could not result

in a correct mathematical judgment. Now it matters very much which of these two different scenarios we have in mind when we draw the parallel to Joe's case. On the first sort of scenario, I believe we would say that the mathematician with a brain tumor simply no longer has even the capacity to perform mathematical calculations that he did before. The parallel to our case of Joe, then, is one in which Joe lacks the capacities and skills necessary for doing the right thing. But in that case, it isn't obvious or intuitive that Joe is responsible for not doing the right thing. In clear contrast, on the second scenario, the mathematician arguably continues to possess the relevant capacities, despite the fact that the device would prevent him from engaging them were he to try or otherwise think thoughts associated with his trying to so engage them. The parallel to our case of Joe here is one in which the device still functions as a counterfactual intervener, albeit one that is there to stay. And in that case, it can be argued as before that Joe still has the ability to do the right thing for the right reasons in the interference-free capacity sense. Thus, we need not accept this scenario as a counterexample either.

Now one reason that the example might initially appeal is by an equivocation of these two scenarios. In imagining the first, we accept Joe's incapacity in the relevant sense, and in imagining the second, we accept his responsibility. But we do not generate both of these conclusions at the same time if we hold fixed the scenario.

Let us try one more way of making the counterexample work. Brain tumors might work a different way. If they are encapsulated and simply put pressure on the brain from outside it, and if they are diagnosed early and removed, patients can regain lost function. For this to constitute the basis of a counterexample, however, such a scenario must lead us to say that the mathematician lacks the capacity to do higher mathematics. Yet even if it does, we have a distinct lack of parallel to the Tax Evasion case. For by hypothesis, the device in Joe's brain is

not “triggered” unless there is a sufficient level of attentiveness to moral reasons. Yet it is unclear in the brain tumor case what is doing the “triggering”. For the counterexample to work, it seems that the mathematician’s thinking certain thoughts would need to be a trigger for the tumor to do some sort of work. But in the brain tumor case as described, it is already doing its work—namely, putting pressure on the brain. So it is simply unclear how this third scenario can function to underwrite the claim that the device in Joe’s brain deprives Joe of the relevant capacity while continuing to need triggering in order to operate. Yet the “triggering” mechanism is crucial to the example, because it is this aspect of the case that leads us to hold Joe responsible. In the end, then, the brain tumor parallel does not help generate a counterexample to the rational abilities view.

It is worth thinking more about this characterization of an ability as a capacity free of interference. Recently, participants in the debate about responsibility and determinism have explicitly distinguished between two very different ways that determinism appears to threaten responsibility.²⁵ On the one hand, determinism is thought to take away the ability to do otherwise, and on the other, determinism is thought to preclude an agent’s being the true source of her actions. Frankfurt-style examples bring out this point. For even if it turns out that having the ability to do otherwise (in the sense that is lacking in the examples) is not required for responsibility, determinism might still prevent one from acting responsibly insofar as one’s actions are mere consequences of past events outside of one’s control. This suggests that there are really two separate threats to free and responsible action that have their origin in determinism. Interestingly, if having an ability is instead understood as having a capacity free of interference, it becomes much more difficult, if not impossible, to find examples in which an ability to do otherwise is lacking while determinism is false, and the agent is acting freely and

responsibly. As a result, it becomes very difficult, if not impossible, to distinguish conditions under which responsibility is threatened by a lack of ability as a result of determinism or by the failure to be a source of one's actions as a result of determinism.²⁶ This is not to say that the two kinds of threats are conceptually indistinguishable, but it does offer one explanation for why the two threats have been confused in the past, and also suggests the possibility that the two kinds of threats are conceptually linked in a closer way than is sometimes supposed.

Before concluding, it is important to address a potential challenge to the view set out. One might worry that in the interference-free sense of ability, one can have the ability to do otherwise while lacking any genuine alternatives; thus, the intuitively appealing idea that alternatives are necessary for blameworthiness would fail to be captured if the interference-free sense of ability is used. One way to reply to this challenge is to suggest that there is a corresponding sense of "alternative" such that if one has the capacity to do otherwise and is interference-free, one does have an alternative. But even if that is not the sense often invoked in discussions of responsibility (and, in fact, the one that is meant to be precluded in the Frankfurt-style cases), one might argue that there is a natural confusion between the two senses. It might be that typically when one has the interference-free ability to do otherwise, one does have alternatives in the sense that is precluded by Frankfurt-style cases. Frankfurt-style cases are, after all, arguably not the norm. So we might mistakenly assume that if we have the capacity to do otherwise and are interference-free, then we have alternatives in the sense invoked by Frankfurt-style cases.²⁷

IV. Conclusion

In this paper, I have addressed two kinds of objections that target the asymmetry of the rational abilities view, each from a different direction. In the process, we have seen ways in which the view is incomplete, including how exactly we are to understand the ability to do otherwise. And there are other worries that I have set to the side, most notably, ones concerning the implications of the view in a deterministic world. At the same time, I believe that the rational abilities view has what are perhaps surprising and interesting resources for dealing with the two challenges, and ones that are sufficiently strong to make the view worthy of further exploration.

Department of Philosophy

University of California, San Diego

REFERENCES

- Brink, David (2004) "Immaturity, Normative Competence, and Juvenile Transfer: How (Not) to Punish Minors for Major Crimes," *Texas Law Review* 82, 1555-85.
- Fischer, John Martin (1994) *The Metaphysics of Free Will: An Essay on Control*. (Cambridge: Blackwell).
- Fischer, John Martin (2006) *My Way: Essays on Moral Responsibility*. (Oxford: Oxford University Press.)
- Fischer, John Martin (1982) "Responsibility and Control," *Journal of Philosophy* 79, pp. 24-40.

- Fischer and Ravizza (1998) *Responsibility and Control, A Theory of Moral Responsibility*.
Cambridge: Cambridge University Press.
- Fischer and Ravizza (1992) "Responsibility, Freedom, and Reason." Critical Review of *Freedom Within Reason* by Susan Wolf. *Ethics* 102, pp. 368-389.
- Frankfurt, Harry (1969), "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, pp. 829-39.
- Frankfurt, Harry (1971), "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, pp. 5-20.
- McKenna, Michael (2003) "Robustness, Control, and the Demand for Morally Significant Alternatives: Frankfurt Examples with Oodles and Oodles of Alternatives", in Widerker and McKenna, eds. (2003), pp. 201-218.
- Mele and Robb (1998) "Rescuing Frankfurt-Style Cases," *The Philosophical Review* 107, pp. 97-112.
- Mele and Robb (2003) "Bbs, Magnets, and Seesaws: The Metaphysics of Frankfurt-style Cases," in Widerker and McKenna (2003), pp. 127-138.
- Otsuka, Michael (1998) "Incompatibilism and the Avoidability of Blame," *Ethics* 108, pp. 685-701.
- Naylor, Margery Bedford (1984) "Frankfurt on the Principle of Alternate Possibilities," *Philosophical Studies* 46, 249-58.
- Nelkin, Dana Kay (in preparation) *Rationality, Responsibility, and the Sense of Freedom*.
- Pereboom (2001) *Living Without Free Will* (Cambridge: Cambridge University Press).
- Van Inwagen (1983) *An Essay on Free Will*. (Oxford: Oxford University Press).
- Wallace, R. Jay (1994) *Responsibility and the Moral Sentiments* (Cambridge: Harvard University

Press).

Watson, Gary (1975/2004) "Free Agency," *Journal of Philosophy* 72, 205-20, reprinted in Watson (2004).

Watson, Gary (1987/2004) "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," in Ferdinand Schoeman (ed.) *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (Cambridge: Cambridge University Press), 256-286, reprinted in Watson (2004).

Watson, Gary (1996/2004) "Two Faces of Responsibility," *Philosophical Topics* 1996, reprinted in Watson (2004).

Watson, Gary (2001/2004) "Reasons and Responsibility," *Ethics* 111, 374-94, reprinted in Watson (2004).

Watson, Gary (2004) *Agency and Answerability: Selected Essays*. (Oxford: Oxford University Press).

Widerker, David and McKenna, Michael, eds. (2003) *Moral Responsibility and Alternative Possibilities* (Burlington: Ashgate).

Wolf, Susan (1990) *Freedom Within Reason* (Oxford: Oxford University Press).

Wolf, Susan (1980) "Asymmetrical Freedom" *Journal of Philosophy* 77 (1980), pp. 151-66.

¹ This paper grew out a set of comments I presented while on a panel at the Conference on Actions and Values, celebrating the publication of Gary Watson's *Agency and Answerability* at the University of California, Riverside in 2005, and a part of a paper I presented at a conference on Psychopathology and Evil at the University of San Francisco in 2006. I am very grateful to the audiences for their feedback, and especially to John Martin Fischer, Gary Watson, and Manuel Vargas both for giving me these opportunities to work through the ideas that follow and

for their invaluable input. Finally, many thanks to Derk Pereboom, Sam Rickless, and an anonymous reviewer for this journal for their excellent comments on previous drafts.

² See, for example, Van Inwagen (1983) for a classic defense of this position.

³ See, for example, Fischer and Ravizza (1998, 37).

⁴ It may be that there is not a single right thing to do in a given situation, in which case one is responsible so long as one can do one of the good things available for good reasons.

⁵ I leave as open questions here just what abilities are required, and whether there are special abilities that humans need, but that are not necessary just in virtue of being responsible agents generally.

⁶ The view also explains why in both legal and non-legal contexts we tend to hold children less responsible than adults. See Brink (2004) for a discussion of reasons to resist a recent trend to decrease the differences in legal treatment between adults and children.

⁷ In this paper, I set aside certain issues concerning determinism, and focus primarily on when, if at all, responsibility requires the ability to do otherwise. The issues concerning determinism are of great importance; here I approach some of them indirectly via their connection to the ability to do otherwise.

⁸ It is important to note that Wallace also provides a second condition for moral blameworthiness, namely, that the action in question must be wrong. Further, and equally importantly, Wallace does not take these conditions to define responsibility on their own. An agent is responsible in the first instance when her actions are the appropriate attitude of the reactive attitudes. In turn, the reactive attitudes are appropriate (and fair) when these two conditions are met.

⁹ See, for example, Fischer and Ravizza (1998).

¹⁰ I think that things are a bit more complicated than the simple statement of the view in the text, and my own view also departs from Wolf's view set out in the text in some important respects. For example, it might be that acting with the ability to do the right thing for the right reasons is sufficient for responsibility, but not necessary; in particular, one might add a minimal "tracing" component to the view, allowing that one might be responsible for actions performed without the relevant ability if they are the result of prior actions or decisions performed with the relevant ability. I leave discussion of issues of this kind for another time, as the objections that I consider in the remainder of the paper do not hinge on settling them.

¹¹ Nevertheless, it is worth emphasizing that the rational abilities view is not alone in having an asymmetrical character. Any view that posits a single specific ability will have this kind of character, too. For example, according to one version of a "real self view", one is responsible for one's actions just in case one has the ability to conform one's actions to one's true values (where those values represent one's "real self"). Watson (1975/2004, 26) offers a view of this type, when he writes "The free agent has the capacity to translate his values into action." Here the evaluational system might be said to represent the "real self", and having the capacity to act on that system is what is necessary for free action. (At the same time, it is important to note that Watson also suggests a different sort of view in the second part of the same sentence: "his actions flow from his evaluational system." Although this second claim seems to be a gloss on the first, I think they are actually quite different in meaning. I call the model described in the first claim a "capacity" model and the second model, according to which one must actually act on one's evaluations (and not simply have the capacity to do so), the "flow" model.) On this view, if one actually conforms one's actions to one's true values, there is no need to be able to do otherwise to meet the condition sufficient for responsibility; but if one fails to conform one's

actions to one's true values, then one must have the ability to do otherwise in order to be responsible. Thus, this view, too, is asymmetrical. This illustrates that asymmetry comes naturally with any condition that requires a substantive ability, and the rational abilities view is thus not unique in incorporating asymmetry.

¹² Watson (1996/2004), 266.

¹³ One might worry that the distinction is not yet sufficiently clear. For example, in describing the woman who betrays her ideals as someone who is not accountable to anyone for so doing that, Watson does not explicitly consider accountability to herself. Thus, it would be good to have an example in which it is easier to accept the existence of blame in the attributability sense and not in the accountability sense. One case might be a toddler who behaves in a way we could accurately describe as "mean", while not at the same time thinking that he has violated any obligation or standard that applies to him. Another might be someone who has a serious mental disorder, but whom we would not unreasonably call "cruel". At the same time, if the distinction could not be made out in the end, then the objection to follow loses much of its force. I will here assume that the distinction is a real one (as illustrated by these cases), and go on to argue that even if it is, the objection based on it can be met.

¹⁴ See Watson 1992/2005, 261. The view seems to be based on a "flow" model, as described in note 11. Watson explicitly notes that, according to this sort of view, "weak-willed" actions do not count as free (262).

¹⁵ It is important to note that Watson is not committed to a particular account of the sense of avoidability that appears to be required, or even to the requirement of avoidability itself. He simply claims that the question arises here.

¹⁶ It is important to note that Watson does not claim that the problem is intractable, only that it must be addressed. As Watson (1992/2004, 285) puts it, “once we view praise and blame in terms of the fairness of assigning rewards and sanctions, as it seems we must from the perspective of accountability, we cannot dismiss this complaint out of hand.”

¹⁷ See Watson (1987/2004, 257-8), and (2001/2004, 316).

¹⁸ Also, and even more briefly, it is worth thinking about non-moral sorts of responsibility in this connection. If there are indeed other sorts (e.g., responsibility for one’s aesthetic choices or athletic performances), then we can ask here, too, whether sanctions and rewards play the same sort of role. (There doesn’t seem to be a complement to the reactive attitudes, or at least such a complement would not seem to be required.)

¹⁹ Among many possible views is one according to which responsible bad behavior is not required to be met with sanctions, but if sanctions must be suffered, those responsible, rather than those who are not, ought to be the ones to do so. (In Nelkin (in preparation) I discuss in detail the Gandhi-like examples and this general line of reply.)

²⁰ I thank a reviewer for this journal for posing this challenge.

²¹ For Frankfurt’s original case, see Frankfurt (1969). Frankfurt also credits Robert Nozick with independently constructing a similar case.

²² I say “not entirely” because part of the appeal of the view, I believe, comes from its surface plausibility, and since the notion of “ability” is on the surface of the view, it cannot have an explication that would strike us as implausible.

²³ In correspondence. This suggestion is based on a similar one that appears in Pereboom (2002), 27-28, in response to a proposal of McKenna’s, namely, one that requires a power to be the author of one’s actions or not.

²⁴ See the entry on brain tumors in adults in the National Institute of Health’s Medline Plus site.

<http://www.nlm.nih.gov/medlineplus/ency/article/007222.htm>.

²⁵ For example, Pereboom (2001) distinguishes between what he calls “leeway incompatibilism” and “source incompatibilism,” each based on a different reason for thinking determinism and freedom (and so responsibility) are incompatible. Also, see Fischer (1982).

²⁶ See Mele and Robb (1998) and (2003), and McKenna (2003) for additional Frankfurt-style examples, each of which is meant to address objections to the original Frankfurt cases. In these cases, the protagonists are intuitively responsible, determinism is false, and in an important sense, the agent lacks the ability to do otherwise. But in the interference-free sense, the agents would appear to have the ability to do otherwise.

²⁷ Actually, the situation is more complicated than this, since at least some who offer Frankfurt-style cases distinguish between robust and non-robust alternatives. I discuss these in Nelkin (in preparation).