Symmetry, Rational Abilities, and the Ought-Implies-Can Principle

Matthew Talbert
West Virginia University

- forthcoming, Criminal Law and Philosophy -

A widely held view among philosophers working on moral responsibility is that praise and blame are symmetrical: in other words, the factors that undermine moral responsibility for good actions also undermine moral responsibility for bad actions, and vice versa. Susan Wolf is one prominent philosopher who rejects this symmetry thesis. In her 1980 paper, “Asymmetrical Freedom,” Wolf says that “[t]he freedom required for moral responsibility … is the freedom to be good.”¹ A consequence of this claim is that certain conditions—such as psychological determinism—undermine blameworthiness but not praiseworthiness. This asymmetry follows from the fact that even if an agent could not have done otherwise than to act well, she still possesses the freedom to be good (and is therefore morally responsible for her good behavior), but an agent who could not have done otherwise than to act badly lacks the freedom to be good (and is therefore not morally responsible for her bad behavior).

In her recent book, Making Sense of Free Will and Moral Responsibility, Dana Nelkin enlarges on Wolf's basic point, developing a full-scale and fully defended theory of moral responsibility that insists on an asymmetry between praise and blame.² The scope of Nelkin’s innovative and tightly argued book is remarkable, especially given its relative brevity. Nelkin manages to deal with almost every current debate in the free will and moral responsibility literature, and she often does so in fresh and provocative ways. For example, though Nelkin believes that physical determinism is compatible with moral responsibility, she makes use of

strategies generally favored by incompatibilists, such as advocating a theory of agent causation and pursuing traditionally incompatibilist lines of argument against other compatibilist views. In addition to these innovations, Nelkin offers an attractive theory of the sort of ability to do otherwise that is associated with moral responsibility as well as a compelling account of the conclusions to be drawn from our sense of ourselves as free agents.

In what follows, I will focus mainly on areas where I disagree with Nelkin, particularly with her claim that an agent is morally blameworthy only if she possessed the ability to do otherwise. While I disagree with Nelkin’s approach in several ways, I must admit that she makes a very forceful case for the asymmetry thesis. I would be surprised if her book does not win converts to her view, or at least to certain aspects of it. Indeed, as I suggest in Section 3 below, I suspect that Nelkin’s account of the ability to do otherwise is something that many compatibilists are already committed to, even if they do not realize it. I will argue, however, that a more thoroughgoing application of central compatibilist convictions leads us to reject what Nelkin says about the ability to do otherwise, particularly its asymmetrical application.

1. The Rational Abilities View

Nelkin defends a perspective that she calls the “rational abilities view.” According to this view, “one is responsible for an action if and only if one acts with the ability to recognize and act for good reasons” (3). An asymmetry between responsibility for good behavior and responsibility for bad behavior follows directly from this formulation. On the one hand, an agent who acts well, and who does so for morally good reasons, possesses an ability that Nelkin takes to be sufficient for moral responsibility. Thus, an agent who acts well is open to praise regardless of whether she could have omitted her behavior by responding differently to the reasons before
her. However, when an agent fails to respond appropriately to reasons, and acts badly as a result, he will be morally responsible only if he had the ability to respond appropriately to the reasons that confronted him. Therefore, we have the following asymmetry: “the ability to do otherwise is required for moral responsibility when one does a bad action or an action for bad reasons, but not when one does a good action for good reasons” (16).

As Nelkin notes, the asymmetry thesis is open to objection both from those who believe that the ability to do otherwise is necessary for neither praise nor blame as well as from those who believe that this ability is necessary for both praise and blame. In this section, I will sketch an objection of the former sort, focusing on Nelkin’s claim that the ability to do otherwise is a requirement on blameworthiness. (I will discuss the specifics of Nelkin’s account of the ability to do otherwise in Section 3. What is important for present purposes is that this ability involves the possibility of an agent responding differently to moral considerations.)

At the heart of the rational abilities view is a commitment to the claim “that what bestows responsibility is the possession of certain rational abilities” (7). I agree that moral responsibility depends on the possession of rational abilities, but the details of this commitment can be spelled out in different ways—for Nelkin, it is tied up with the idea that wrongdoers are excused when they lack the ability to respond to good reasons. However, as I will try to show, it is possible for an account of moral responsibility to acknowledge the importance of rational abilities while still allowing that wrongdoers are sometimes open to blame even if they could not have responded appropriately to good reasons.

To see how such an account might go, consider Nelkin’s discussion of Rosa and Sylvia. Both women perform the morally good act of jumping into the water to save a drowning child. However, while Sylvia could have responded differently to the drowning child’s plight, Rosa
could not have. Is Rosa open to praise even though she could not have omitted her good action?

Nelkin believes that she is because Rosa’s decision to save the child parallels Sylvia’s decision—which Sylvia could have omitted—in important respects:

… by hypothesis, Rosa acts for precisely 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 terrible, terrible thing. Recognizing all of this, Rosa jumps in in order to prevent the child from dying. (40)

It is important for my purposes to emphasize that on Nelkin’s view Rosa can act for good reasons—and do so in a way that reflects well on her as a moral agent—even though she could not have responded differently to the situation with which she was confronted.

Nelkin’s analysis of Rosa’s behavior seems exactly right to me. Regardless of whether she could have responded differently, Rosa is praiseworthy insofar as she acted for good reasons in a way that reflects well on her. Nelkin is not explicit about why Rosa cannot do otherwise than to save the child, but the example suggests that the source of Rosa’s inability is her deep concern for the child and the prominent role that the child’s welfare plays as a reason for her. As Susan Wolf puts it in a discussion of a similar case, perhaps Rosa “lacks the ability to do otherwise simply because her understanding of the situation is so good and her moral commitment is so strong.”

These facts do not make Rosa unfit for praise—indeed, they make her an especially good candidate for it. By contrast, if Rosa lacked the ability to do otherwise because her behavior was the result of mind control (or some other mechanism that eliminated her ability to act for reasons), then Rosa’s behavior would not reflect well on her in the way that makes praise appropriate. Praiseworthiness seems to depend, then, on whether an agent’s behavior has the

---

4 Suppose Rosa acted habitually or instinctively. Whether this undermines praiseworthiness will depend on what we mean by “habit” and “instinct.” If we think of these things as entirely disconnected from rational processes, then Rosa might not be praiseworthy, but if we think of habits and instincts as themselves honed by an agent’s judgments about reasons, then moral praise might still be apt.
sort of moral significance that is appropriate to praise, where this is a function of the kind of action the agent performed, the reasons for which she performed it, and the way that these things reflect on her as a moral actor.

In contrast to Nelkin, I believe that our approach to blame ought to be symmetrical to the foregoing account of praise: what matters for blame, on my view, is whether an agent’s behavior has the sort of moral significance appropriate to blame, and this is largely a function of the way the agent responds to reasons and what this reveals about her orientation toward other people and the value that she assigns to their needs and interests. On Nelkin’s view, this is not enough for blame: in addition to responding to reasons in a way that reveals, say, a lack of concern for the welfare of others, a blameworthy agent must also have the ability to respond appropriately to moral considerations. So, for example, in order to confidently hold soldiers morally responsible for their war crimes, we would need “to determine the extent to which soldiers have the ability to recognize and act on good reasons at the moment in question” (11). It would not be enough, on Nelkin’s view, to know that a particular soldier’s bad behavior was intentional and voluntary and that it issued from an objectionable lack of concern for the welfare of others; we must also know (or have good reason to believe) that the soldier could have responded appropriately to the moral considerations to which he did not respond appropriately.

There is undeniable appeal to the idea that a person must have been able to respond to good reasons if she is to be blamed for failing to do so. This appeal seems to stem from our confidence that moral responsibility is absent in certain cases in which an agent cannot respond to good reasons. It is important to note, therefore, that we can allow that agents who are unable

5 Similarly, Nelkin says that doubts about the moral responsibility of subjects in Stanley Milgram’s famous experiments on obedience (in which subjects mistakenly believed that they were administering painful shocks to other people) arises from the thought that they were “somehow lacking in the ability to do the right thing for the right reasons” (11).
to respond to good reasons are often not morally responsible without having to explain their lack of responsibility in terms of their inability. If, as I suggested in the last paragraph, what really matters for moral blame is whether an agent’s behavior has a certain sort of moral significance, then it might be that many agents who cannot respond to good reasons are not open to blame simply because their behavior lacks the sort of moral significance that is appropriate to blame.

For example, a person might be unable to respond to good reasons because she is deranged, or subject to a compulsion, or forced to make a decision under conditions of extreme stress. Each of these circumstances might render a person who acts badly blameless, but not primarily because the circumstances impair her ability to respond to good reasons. Rather, these circumstances provide access to an excuse for the more basic reason that they make it difficult to interpret the agent’s bad behavior as stemming from a lack of concern for the way her actions affect other people.6 If this is right, then in these sorts of cases the inability to respond to good reasons is merely associated with the thing that actually excuses the agent. Thus, while the agent in this case is not blameworthy, her inability may not be why she is not blameworthy.7 I believe that this account explains some of the appeal of the idea that people ought not to be blamed if

---

6 The case of stress is an outlier. If a person acts compulsively, we are unlikely to interpret his behavior as expressing contempt or disregard, but people acting under stress can certainly treat us this way. However, we still seem to make allowances for people in stressful situations by not holding them to the moral standards that we otherwise would. It is tempting to say that this is because it is difficult for people to act well under stress, but I think that we are also reluctant to fully attribute an agent’s behavior to her (in the way required for blame) if we think that she acted as she did only because of the stressors to which she was exposed. When we think this about a person, we see her behavior as attributable more to her circumstances than to her.

7 The Milgram subjects mentioned in Note 5 might be like this. The subjects who believed that they were administering shocks seem not to have done so because they were contemptuous of the welfare of the people they thought they were shocking. In fact, many subjects appear to have been very concerned about the welfare of their supposed victims; they shocked them because features of Milgram’s experimental setup exerted a surprising pressure on them to follow the experimenter’s instructions even when this conflicted with their concern for others. It may be, then, that many of the Milgram subjects are not open to blame because their behavior is not best explained by attitudes and judgments that make people appropriate candidates for blame. It may also be true that the subjects were unable to respond to good reasons, but this is not the most direct route to explaining why they are not open to blame. See Stanley Milgram, Obedience to Authority (New York: Harper and Row Publishers, 1969).
they lack the ability to respond to good reasons, and it gives us some reason to deny the principle that this ability is necessary for blameworthiness.

What would an agent who is blameworthy even though she could not have responded to good reasons look like? She would be an agent who, despite her inability, acted for reasons that indicate things about her as a moral agent that are generally good grounds for blame. A case involving such an agent would be the inverse of Nelkin’s example involving Rosa, and I think that this is the sort of case that we ought to be thinking about when we ask whether the ability to respond to good reasons is a necessary condition on blameworthiness.

Imagine, then, a case of serious wrongdoing—a contract killing, for example—in which the behavior of our wrongdoer is explained by reference to judgments and attitudes that normally open a person up to moral blame even though he could not have responded appropriately to the relevant moral considerations present in the context in which he acted. Let us suppose that whereas Rosa, because she is so good, cannot help but see that there is overwhelming reason to try to save a drowning child, our killer, because he is so bad, cannot help but conclude that he has decisive reason to carry out a particular murder. Perhaps, given the killer’s allegiance to his employers, his induction into and commitment to a certain way of life, his desire for money, and the way that he weighs the taking of human life against these things, the option of deciding to not kill is not on the table for him. Now suppose that our killer knowingly and intentionally, and perhaps even after reflection, murders an innocent person for money and that the possibility of not killing did not appear as a serious option for the killer because of his profound lack of concern for his victim’s welfare. In this kind of case, why should it be a barrier to moral blame that the killer lacks realistic access to a different decision about how to behave? After all, what explains the killer’s difficulty in responding appropriately to moral considerations are his
morally objectionable values and evaluative judgments—in other words, just the sorts of things that normally make a person blameworthy.\footnote{Nelkin notes that I have argued that even psychopaths may be open to moral blame (I make this argument in “Blame and Responsiveness to Moral Reasons: Are Psychopaths Blameworthy?” Pacific Philosophical Quarterly 89 [2008]: 516-535). She introduces several reasons for rejecting this thesis (76-79) and I admit that the case of the psychopath is complicated because of his utter inability to engage with moral reasons in the way that non-psychopaths do. However, the killer I am imagining here is not a psychopath. He is a very bad fellow, no doubt, but he may well be able to appreciate certain moral reasons, to show genuine concern for the welfare of at least some people, and so on (even if it is impossible for him to respond to good reasons on some occasions).}

I assume that Nelkin would say that the killer I just described would be open to blame only if, in addition to the facts mentioned above, he had been able to respond differently to reasons. However, I do not think that adding the possibility of responding differently to reasons helps us see how the killer would be more blameworthy than in the case I just described, and this is because the actual reasons on which the killer acts already seem to include enough for blame. Consider again Rosa and her counterpart Sylvia. Sylvia could have refrained from saving the drowning child but Nelkin does not think that this possibility adds anything to the story of Sylvia’s behavior that distinguishes her from Rosa in terms of praiseworthiness. I am making a corresponding point about the killer. If we hold fixed the reasons on which the killer acts, there does not seem to be much reason to distinguish, in terms of blameworthiness, between the killer described above and a counterpart of his who acts for the same reasons but who could have acted on different ones.\footnote{It may be that the counterpart can reject moral considerations in a way that the original killer cannot. This might mean that the counterpart is capable of a sort of moral badness of which the original killer is incapable; however, the facts about the original killer’s behavior still seem to me sufficient to make him blameworthy. I develop the strategy of comparing a wrongdoer who cannot respond to moral reasons to a counterpart who can in “Moral Competence, Moral Blame, and Protest,” The Journal of Ethics 16 (2012): 89-109.}

Now one might think that there could not be someone like the killer I describe because it is impossible for a wrongdoer to genuinely lack the ability to act for good reasons and to still guide his behavior by judgments about (other) reasons. Perhaps this is so, but a number of authors allow that there can be wrongdoers who retain general capacities for responding to
reasons (even to moral reasons), but who cannot respond to specific moral considerations, and these authors argue that such wrongdoers are not open to blame on account of their moral blind spots. Take, for example, Susan Wolf’s assessment of the blameworthiness of slave owners in the antebellum U.S. south. Even if these slave owners were generally rational and capable of making certain moral judgments, their bad values—which were imposed on them by their cultural context—may have made it impossible (or unreasonably difficult) for them to recognize the iniquity of slavery, in which case they would not be blameworthy for engaging in the practice, according to Wolf.¹⁰ I take Nelkin to be open to this conclusion as well, but I reject this view because it is possible for a slave owner to treat his slaves with the sort of contempt that justifies blame even if it is impossible for him to recognize the immorality of his behavior.

2. Negative Reactive Attitudes and Sympathy

Despite my commitment to the view outlined in the previous section, I must admit that some of what Nelkin says has shaken my confidence. She does an especially good job demonstrating the way in which expressions of our moral responsibility judgments seem to be characterized by an asymmetry. Specifically, the attitudes and responses associated with blame sometimes seem to be inhibited by factors that do not inhibit the attitudes and responses associated with praise. The case of Jeremy Gross provides a useful example.

Gross was found guilty of a brutal and premeditated murder, but the jury that heard his case refused to impose the death penalty after mitigating evidence was presented regarding Gross’s abusive upbringing. Nelkin suggests that the jury’s decision depended (at least in part)

on doubts about Gross’s full moral responsibility and that these doubts stemmed (at least in part) from the thought that Gross had difficulty responding to good reasons. Nelkin notes that,

… one wavering juror felt that though Gross had a ‘moral compass’ it had been shattered by the abuse he suffered as a child. He said, ‘I began to think not that there’s an excuse for what happened, but I had an understanding of his torment … Sympathetic is too strong a word, but I can’t think of a better way to describe how I felt. I struggled with whether he knew the difference between right and wrong.\footnote{Nelkin is quoting Alex Kotlowitz, “In the Face of Death,” \textit{New York Times Magazine} (July 6, 2003).} (14, Nelkin’s ellipsis)

Nelkin says that the “juror’s report fits well with the idea that lacking robust rational abilities—including the ability to distinguish right from wrong—diminishes responsibility” (14).

It is important to note that Gross’s case does not fit very well with my proposal in the last section that a person who is unable to respond to good reasons might be blameless for his bad behavior \textit{for some other reason than this inability}. For one thing, the juror explicitly cites Gross’s defective moral compass in explaining his decision to abstain from imposing the death penalty, and there is not much reason to think that Gross’s behavior did not express seriously deficient moral regard for his victim. So what can I say about Gross’s case? He probably does not have access to the excuse I identified in the last section (i.e., an absence of ill will), but surely many people besides the juror find their retributive impulses mollified in this case. Isn’t it natural, then, to think that it must be Gross’s defective moral compass that is doing the excusing work?

In response, it is worth noting something interesting about the juror’s statement: he says that Gross is \textit{not} excused for what he did, which would seem to entail the attribution of some degree of moral responsibility to Gross. Of course, the juror might be mistaken about what it means to excuse someone: perhaps he does not want to say that Gross is excused because he thinks this would mean that his behavior was justified. On the other hand, the juror might be
hinting at the idea I want to suggest. Perhaps the judgment that Gross is morally blameworthy for his bad behavior stays in place, and he is therefore not excused, but there is still hesitancy about imposing sanctions on him as a means of expressing moral blame. The source of this hesitancy might be sympathy.

The juror says, “Sympathetic is too strong a word …” but I suspect that sympathy for Gross’s terrible upbringing played a major role in the juror’s unwillingness to impose the death penalty.12 Similarly, when we sympathize with a wrongdoer because of her upbringing, this can make it hard to sustain toward her the negative reactive attitudes that express blame. However, this does not necessarily mean that these attitudes are unwarranted; it may just be that some aspects of the wrongdoer’s story elicit negative attitudes and other aspects elicit contrary emotional responses.13 The fact that we are inclined to feel sympathy for a wrongdoer does not negate the judgment that she is open to moral blame, even if it mutes the sentiments by which we express these judgments.

And, of course, sympathy is much more likely to be aroused in cases involving blame than in cases involving praise. The factors that contribute to bad behavior—such as a bad upbringing—are more likely to make a person appear sympathetic than are factors that contribute to good behavior, and even if we did sympathize with a person who acted well, this would not inhibit praising responses since it is typically pleasant to be on the receiving end of these responses. The responses associated with blame, on the other hand, are usually unpleasant, and

12 Indeed, it may be that Gross’s defective moral compass is significant for the juror mainly because of the way it was brought about. Would the juror have had the same response toward Gross if he believed that Gross had an exemplary childhood?
the fact that a person is subject to such unpleasant responses may contribute to our sympathy for her.

3. The Ability to Do Otherwise

Concerns about physical determinism have been central to arguments about whether an agent must have the ability to do otherwise in order to be morally responsible for his behavior. For the most part, compatibilists believe that moral responsibility is compatible with determinism because neither praise nor blame requires the sort of ability to do otherwise that determinism would exclude. And, for the most part, incompatibilists regard moral responsibility as incompatible with determinism because they believe that both praise and blame depend on the ability to do otherwise. It might seem, then, that Nelkin’s position is a compromise between compatibilism and incompatibilism. That is, she might have a view according to which, if determinism is true, then people may be open to praise but not to blame. However, this is not Nelkin’s position (18, 72). She believes that the sort of ability to do otherwise that is required for blameworthiness (but not for praiseworthiness) is compatible with determinism.

Nelkin develops her account of the ability to do otherwise in the context of responding to John Martin Fischer and Mark Ravizza’s argument against Susan Wolf’s version of the asymmetry thesis. Fischer and Ravizza’s argument appeals to two cases. In the first case, called “Hero,” Martha deliberates about whether to jump into the surf to save a drowning child; she quickly decides to try to save the child and attempts to do so. However, had Martha “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’” (64).¹⁴ Martha

seems morally responsible for her good behavior, and since she could not have done otherwise, Fischer and Ravizza conclude (along with Wolf and Nelkin) that the ability to do otherwise is not necessary for an agent to be morally responsible for her good behavior.

In Fischer and Ravizza’s second case, called Villain, Joe plans to push a child off a pier so that the child will drown. Another person, Max, is aware of Joe’s plan and “has secretly installed a device in Joe’s brain …. [that] can be employed by Max to ensure that Joe decides to drown the child” (65, my ellipsis). As things transpire, Joe sticks to his plan and pushes the child off the pier for his own reasons. Max’s device is irrelevant to Joe’s behavior, but given its presence, Joe could not have done otherwise than to try to kill the child. Since Joe acted for his own reasons, Fischer and Ravizza say that he is morally responsible for his behavior and that the ability to do otherwise is therefore unnecessary for an agent to be morally responsible for bad behavior. Thus, on their view, praise and blame are symmetrical.

In response, Nelkin says,

… a proponent of the asymmetry thesis can 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.

Neither Martha nor Joe can do otherwise insofar as it is inevitable that they will perform their respective actions. What both agents lack, then, is the ability to do otherwise in the “inevitability-

---

15 Again, Nelkin is quoting Fischer and Ravizza, “Responsibility, Freedom, and Reason.”
16 Just to be clear, when Nelkin says that we don’t have to reject “the asymmetry in question” she is not, I think, saying that we don’t have reason to reject the asymmetry that Fischer and Ravizza are encouraging us to reject. What we don’t have reason to reject, on Nelkin’s view, is a different asymmetry, one that Fischer and Ravizza don’t recognize. As we will see, Nelkin distinguishes between two senses of an ability to do otherwise. I take it that she agrees with Fischer and Ravizza that possession of this ability in one of these senses—the sense in which this ability conflicts with deterministic—is unnecessary for either praise or blame. So Nelkin accepts one sort of symmetry between praise and blame, but she insists that possession of the ability to do otherwise in the other sense is a necessary condition on blameworthiness but not on praiseworthiness, which gives rise to the asymmetry that she endorses.
undermining” sense (66). For an agent to have this ability with respect to an action, it would have to be false that it is inevitable that she will perform the action.

However, an agent might lack the ability to do otherwise in the “inevitability-undermining” sense but possess the ability to do otherwise in the “interference-free” sense. On this second conception of the ability to do otherwise,

… 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.17 (66)

The interference-free sense of ability is clearly different from the inevitability-undermining sense, but it is also distinct from the general sense of ability upon which compatibilist accounts often rely. A person may retain the general ability to type—the knowledge and skills required for typing—even if the nearest keyboard is hundreds of miles away. However, such a person would not possess the interference-free ability to type since the lack of a keyboard interferes with the exercise of the relevant skills (67).

According to Nelkin, what it would take for an agent’s interference-free ability to be undermined “is either the removal of [her general] capacities, talents, skills, and so on … or the actual interference with or prevention of the exercise of those capacities” (66). Later she says, “[t]he idea of being interference-free that I prefer, then, is that nothing is actually preventing you from acting otherwise (though it would under different circumstances)” (67). It appears, then, that if you possess a general ability to perform an action, the only way you can fail to have the interference-free ability to perform that action is for there to be an actual (as opposed to counterfactual) interference with your exercise of the general ability.18

17 Nelkin draws this description from Wolf’s, Freedom Within Reason.
18 Here is an obvious worry that Nelkin considers: does determinism entail that whenever an agent acts, his ability to do otherwise was actually interfered with? (72-76). Nelkin’s response is very interesting. First, she argues that
In Villain, since Joe did not try to behave differently than Max wanted him to, Max’s device did not actually interfere with an attempt on Joe’s part to behave differently. Thus, Joe retained his interference-free ability to do otherwise in Villain. So Villain is not a counterexample to the claim that the ability to do otherwise, understood as Nelkin proposes, is necessary for an agent to be open to moral blame.\(^9\)

I suspect that many compatibilists will see good reason to adopt at least some of what Nelkin says here. It seems to me that what compatibilists have mainly been eager to show (with their counterfactual intervener cases and emphasis on general abilities) is that it is the inevitability-undermining sense of the ability to do otherwise that is unnecessary for moral responsibility. But this commitment doesn’t give compatibilists reason to deny the necessity of interference-free abilities for moral responsibility. In fact, it’s easy to construct cases in which an interference-free ability to do otherwise seems clearly necessary for moral responsibility. Consider again the case of Joe and Max. As I noted above, the kind of ability that Nelkin thinks Joe needs for moral responsibility is not undermined by the fact that Max would intervene with Joe’s behavior; it is undermined only if Max actually intervenes. Once we are clear on this detail—and Nelkin suggests that Fischer and Ravizza misread Wolf with respect to this issue (66)—I should think it becomes difficult to deny that Joe needs the interference-free ability to do otherwise in order to be morally responsible for pushing the child off the pier. If my reading of Nelkin is correct, then Joe’s having this ability simply means that his behavior was not caused by an actual intervention on Max’s part. To deny that Joe needs this ability would be to say that he

---

\(^9\) Indeed, it seems that no counterfactual intervener case could provide a counterexample to this principle since in such cases the intervener does not intervene, so there is no actual interference with an agent’s abilities.
might be blameworthy even if he decided not to push the child off the pier and Max intervened and made him do it anyway.Compatibilists should not be interested in holding Joe morally responsible for the child’s death under these circumstances. The fact that this seems so clear (to me, at least) does not mean that Nelkin’s point isn’t important and innovative. It is both of these things, but it is also the sort of point that can seem obviously right once someone has gone to the trouble, as Nelkin has, of articulating it in a clear and forceful way.

Having just emphasized how sensible Nelkin’s view is, I now must say that I do not accept the claim that people are never blameworthy if something actually interferes with their general ability to respond to good reasons. Of course, some kinds of interference undermine blameworthiness. As I suggested above, if Joe had decided not to push the child off the pier but Max made him do it anyway, then Joe would not be blameworthy for pushing the child off the pier. This is because the action that Max caused would not stem from Joe’s judgments about reasons and so would not be attributable to Joe in the right way to ground blameworthiness.

However, it is implicit in Nelkin’s account that not every way in which an ability can be interfered with will function like the device that Max installed in Joe’s brain. Note, for example, that on Nelkin’s view, Martha (in Hero) does not need the interference-free ability to do otherwise in order to be praiseworthy. But since an agent is praiseworthy only if she acts for good reasons, this means that it must be possible for Martha’s ability to do otherwise to be actually interfered with and for her to still act for good reasons. So what sort of actual interference with the ability to do otherwise is compatible with Martha saving the child for good reasons? Perhaps the original Hero case is like this. In that case, if Martha had considered not saving the child, she would have been overwhelmed by guilt and would have tried to save the

---

20 Strictly speaking, Nelkin would not endorse this claim either because there can be cases in which a wrongdoer is responsible for the actual interference with his ability and is thus blameworthy for his subsequent behavior. I set this complication aside below.
child anyway. Whether Martha would be praiseworthy under these actual-interference conditions would depend, on Nelkin’s account, on whether being overwhelmed by guilt is a way of responding to good reasons.

Similar to what I suggested in Section 1, when we evaluate the claim that an interference-free ability is necessary for blameworthiness, we should consider cases of wrongdoing that are symmetrical with Martha’s. Suppose, then, that Joe, in Villain, lacked the interference-free ability to refrain from pushing the child off the pier because at the time when he considered refraining, all the reasons that he counted in favor of killing the child reoccurred to him with overwhelming force. In this case, I should think that Joe is not excused from blame. The lesson is that it matters what kind of actual interference with abilities we are considering. Perhaps Nelkin will not count the case I just described as one of actual interference with Joe’s ability to refrain, but as I have said, she must allow for a case in which Martha’s behavior is actually interfered with but she still responds to good reasons. My claim is that when we have a case that is symmetrical with Martha’s, but involves a wrongdoer who acts for bad reasons, we will have a case in which it is plausible to say that the wrongdoer is blameworthy even though he lacked the interference free ability to do otherwise.

4. The Ought-Implies-Can Principle

Nelkin believes that the Ought-Implies-Can principle (OIC) supports her asymmetry thesis. Here is Nelkin’s formulation of OIC:

(i) If S ought to have performed action \textit{a}, then S could have performed action \textit{a}, and (ii) if S ought not to have performed action \textit{a}, then S could have refrained from performing action \textit{a}. (100)
If OIC is true, it follows that “[a] person is morally blameworthy for what he has done only if he could have done otherwise” (101). After all, if a person is blameworthy for performing an action, then he ought not to have performed that action, and if OIC is true, then the action that he ought not to have performed must have been one from which he could have refrained. However, OIC does not entail a similar claim about actions that one ought to have performed: a person who did what she ought to have done was able to do what she ought to have (thus satisfying OIC) regardless of whether she could have done something else.

Nelkin is inclined to regard OIC as a fundamental principle that sheds light on our moral practices, but that (probably) cannot be defended by appeal to more basic principles or rationales. She considers some rationales for OIC but believes that they most likely fail and that these failures are “suggestive of the idea that the principle is indeed axiomatic, in the sense that it is fundamental” (108).

Though I do not endorse the use to which Nelkin puts OIC, I suspect that she is too hasty in rejecting at least one of the arguments meant to support it. The argument I have in mind is the one proposed by John Martin Fischer (though he does not ultimately accept OIC). Fischer suggests that if OIC (or something like it) were false, “‘then there could be cases in which an agent ought to do X but cannot do X,’” but if this agent “‘would be blameworthy for not doing X, there could be cases in which an agent is blameworthy for not Xing and yet he cannot X. And this is objectionable—even unfair’” (109).²¹

Nelkin believes that Fischer’s reasoning relies on the assumption that moral blame is, or involves, a sort of sanction, and that blame is therefore fair only when such sanctions are appropriate. Earlier in the book, Nelkin rejected this reasoning because she believes that it is appropriate. Earlier in the book, Nelkin rejected this reasoning because she believes that it is

possible to hold people morally responsible for their bad behavior without being inclined to sanction them. For example, “a Mahatma Gandhi-like figure” might eschew sanctioning attitudes like resentment without giving up on moral obligation and moral accountability (42).

I am willing to grant that the Gandhi example shows that there can be a conception of moral responsibility that does not make use of sanctions. My concern is that most people do not employ a Gandhian conception of moral responsibility. I suspect, rather, that when most people blame a wrongdoer, they usually regard it as at least permissible to target the wrongdoer with attitudes like resentment. So perhaps a Gandhian would do nothing unfair when she holds responsible a wrongdoer who could not have done otherwise, but the rest of us may be acting unfairly insofar as we are committed to the view that it is permissible to resent those we blame.22

At any rate, since Nelkin rejects arguments for OIC because “the principle is more basic than the claims being made in support of it,” her endorsement of the principle rests on “recognizing what is supported and explained by it” (111). “[T]he principle can,” she says, “offer mutual support to other principles and to conclusions about particular cases” (111). As an example, Nelkin considers cases in which it is impossible for a person to do something that she would otherwise be obliged to do and therefore a “second best option” arises as a new obligation for the agent.23 For example,

A boulder falls on your friend. You try to lift the boulder and cannot. You immediately acquire the obligation to run for help. You are not obligated to lift the boulder, precisely because you cannot. The explanation for your new obligations … is that you cannot be obligated to do what you cannot do. The existence of new, or second-best obligations, points precisely to this fact. (112)

I have no quarrel with the idea that a second-best option can become an obligation, and I agree, of course, that the person in the example is not obligated to lift the boulder and that she is

22 As I indicate in Note 27 below, I don’t actually accept this conclusion about the potential unfairness of blame.
23 Nelkin is drawing here on Frances Howard-Snyder, “‘Cannot’ Implies ‘Ought Not,’” Philosophical Studies 130 (2006): 233-46.
not blameworthy for failing to do so. I wonder, though, how much light OIC really sheds on this example. I suspect that the person’s inability to lift the boulder is not the whole story about why she is not obligated to do so and that the example does not generalize to all cases in which a wrongdoer cannot do otherwise.

If there were an obligation to lift the boulder in Nelkin’s example, it would presumably arise out of a more general obligation to alleviate suffering, which might itself derive from an even more general requirement to show other people due regard. I tentatively suggest, then, that the person who cannot lift the boulder does not have an obligation to do so because her failure to do so does not violate the obligation to show others due regard. This is why the failure to lift the boulder does not reflect poorly on the person as a moral agent and it is why she is not blameworthy for failing to lift the boulder.

Rather than to say that the person who cannot lift the boulder has no obligation to lift it because she cannot lift it, it would be better to say that she has no obligation because she cannot lift the boulder because it is too heavy. It does not follow that just any reason for which a person cannot lift a boulder would entail that she is not obligated to do so. As I have argued above, the precise form that an inability takes matters a great deal. Suppose, for example, that what makes it so hard for the person to lift the boulder is that she is filled with contempt and hatred for the person under it. In this case, I would not think it out of place to say that the person still has an obligation to lift the boulder if she has the physical strength to do so, and that her failure to do so

\[24\text{ One might say that what it is hard for this person to do is to try to lift the boulder, which is different from the case of a person who can try to lift the boulder but who must fail to actually lift it. I don’t think this presents a real problem for my example, though. My point is just that it is one thing to explain that a boulder was not lifted by referencing its weight, and it is something very different—morally different—to explain this fact by referencing a person’s moral attitudes. At any rate, Nelkin’s original example could be changed a bit so that the relevant inability is an inability to try to lift the boulder: we could, for example, compare the contemptuous person I describe to someone who is so physically exhausted that she can’t bring herself to try to lift the boulder. My claim would be that while an inability predicated on physical exhaustion might well count as an excuse, an inability predicated on objectionable moral attitudes would be unlikely to do so.} \]
reflects poorly on her as a moral agent. I also think the person is blameworthy for failing to lift the boulder because her failure indicates a lack of regard for the person under it. The sort of moral incapacity involved in this case does not interfere with the attribution of blame-grounding, condemnable attitudes toward others because the incapacity here seems to be largely constituted by such attitudes.

However, if there really is something preventing the contemptuous person from lifting the boulder, then it may seem odd to tell her that she ought to do it. What makes this odd is that “ought” claims are often meant to guide behavior, and if we really believe that someone cannot do otherwise, then there is little point in trying to guide her behavior in that direction. However, talk about what someone “ought to do” does not serve only to guide behavior; it also serves to describe behavior: it can note that someone has fallen below a certain standard. This is just the sense in which the contemptuous person ought to lift the boulder even if, in some sense, she cannot. She ought to lift the boulder in the sense that she will be a worse person if she does not do so (if she has the physical strength for it) because a person who has appropriate regard for others would lift the boulder (if she has the physical strength for it).

When we blame people, we are implicitly committed to the claim that they ought to have behaved differently, but the point of this claim is not, I think, exhausted by the possibility of guiding the actions of others. When we blame a wrongdoer and tell him he ought to have behaved differently, we may also be interested in pointing out facts about what sort of treatment other people deserve. This sort of “ought” does not imply “can,” so it can be used even in the case of someone who could not have done what she ought to have. This descriptive sense of

25 Nelkin acknowledges the possibility of drawing a distinction like this (108-109), but she does not put it to the use I am suggesting.

ought will be particularly apt in cases in which an inability to behave as one ought to have stems from the agent’s own moral shortcomings, rather than from physical incapacity.  

Conclusion

There is a great deal in Nelkin’s admirable book with which I am sympathetic but that I have not discussed in any detail. This includes Nelkin’s effort to put a compatibilist spin on agent causation as well as her subtle reflections on our apparent commitment to viewing ourselves as free agents. Instead, I have focused on a central point where I disagree with Nelkin. As I have argued in different ways throughout this essay, it seems to me that a wrongdoer who could not have responded appropriately to moral reasons may still be open to blame. This is most clear in cases in which the wrongdoer’s moral inabilities and impairments do not undermine the thought that her behavior stems from the kind of contempt, or lack of concern, for others that tends to make blame appropriate. In many cases, factors that make it difficult for an agent to respond to moral reasons will also make it unreasonable to construe her behavior as having this sort of blame-grounding moral significance, but other cases of inability won’t be like this. In my view, these other cases show us that blameworthiness (like praiseworthiness) does not depend on whether an agent could have responded differently to moral reasons. I said in the introduction that my conclusion follows from a more thoroughgoing application of compatibilist convictions than seems attractive to Nelkin. The central conviction that I have in mind is the one that says it is the nature of an agent’s actual behavior, and the actual reasons to which she responds, that

\[27\] What I have said so far does not address the issue (raised above) of whether it is fair to blame someone who could not have done otherwise and therefore could not have avoided the sanctions associated with blame, particularly the negative reactive attitudes. Very briefly, my view is that though it may be unpleasant to be targeted with negative reactive attitudes, it is not the point of these attitudes to sanction wrongdoers. These attitudes are simply natural responses to the recognition that another has wronged you. Therefore, facts about the wrongdoer will not tend to show that these attitudes are unfair except insofar as they tend to show that no wrong was committed. For a detailed argument along these lines, see Pamela Hieronymi, “The Force and Fairness of Blame,” *Philosophical Perspectives* 18 (2004): 115-48.
matters for moral responsibility and not whether she might have acted differently. It seems to me that this compatibilist conviction applies not just to the problem of physical determinism but also to cases in which psychological limitations or impairments make it difficult for agents to act for good reasons.²⁸

²⁸ I would like to thank Dana Nelkin for kindly giving me comments on this paper and for fielding questions about her view.