



---

Review: Responsibility, Freedom, and Reason

Author(s): John Martin Fischer and Mark Ravizza

Source: *Ethics*, Vol. 102, No. 2 (Jan., 1992), pp. 368-389

Published by: [The University of Chicago Press](http://www.uchicago.edu)

Stable URL: <http://www.jstor.org/stable/2381611>

Accessed: 24/02/2011 13:22

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ucpress>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Ethics*.

# Responsibility, Freedom, and Reason\*

*John Martin Fischer and Mark Ravizza*†

*Freedom within Reason* is a book which, in its own words, is “unabashedly devoted to solving” the problems of free will and moral responsibility (p. 4). Yet, as Susan Wolf is quick to caution, the “solution” offered in this intelligent and provocative book is not intended to “put these problems, once and for all, to rest” (p. 4). Rather, the hope is more modest—to provide a new way of interpreting and understanding these age-old worries, one which, if successful, will provide “some degree of relief” from the problems of free will and responsibility which have long exercised philosophers (p. 15). In this review, we first summarize briefly the background and structure of Wolf’s argument; then we raise a number of critical questions concerning several of the central moves in that argument. In advancing these criticisms, we hope to question whether, and to what extent, the freedom Wolf finds within reason can provide the type of relief and new understanding that is needed, if not to silence, then at least to quiet traditional worries about free will and moral responsibility.

## SUMMARY

Wolf begins her argument by presenting and motivating what she takes to be the problem of responsibility—“How, if at all, is responsibility possible?”—and the related problem of free will—“What kind of beings must *we* be if we are ever to be responsible for the results of our wills?” (p. 4). According to Wolf, to judge that someone is morally responsible is to regard that person as an appropriate object of the reactive attitudes (e.g., gratitude, resentment, respect, and contempt) and as a legitimate participant in the practices which are related to these attitudes (e.g., praising, blaming, rewarding, and punishing). After reflecting on intuitions concerning agents’ responsibility, or lack thereof, in a number of paradigm cases (including hypnosis, klepto-

\* A review of Susan Wolf, *Freedom within Reason* (New York: Oxford University Press, 1990), \$24.95 (cloth). References to this book will appear parenthetically in the text.

† We thank Alexander Rosenberg for his comments. Also, we have benefited from support from the Center for Ideas and Society, University of California, Riverside.

mania, and action under duress), Wolf suggests that to be an appropriate candidate for such reactive attitudes and practices—to be a responsible agent—a person seemingly needs to satisfy three conditions. First, the agent must have control over her behavior (i.e., she must have a “potentially effective will”); second, she must, when exercising this control, be capable of being informed and governed by relevant considerations (i.e., she must have a “relevantly intelligent will”); and third, her “control must be ultimate—her will must be determined by her self, and her self must not, in turn, be determined by anything external to itself” (p. 10). This last condition Wolf terms the “condition of autonomy.”

A key claim which emerges from Wolf’s opening chapter is that the problems of free will and responsibility can be seen to derive their force and structure from a dilemma which arises once the condition of autonomy is accepted. The problem arises because although autonomy appears to be a necessary condition of responsibility, it also appears that this condition is impossible to satisfy. As Wolf argues: “Either something is behind the agent, making the agent what she is, or nothing is. The idea of an autonomous agent appears to be the idea of a prime mover unmoved whose self can endlessly account for itself and for the behavior that it intentionally exhibits or allows. But this idea seems incoherent or, at any rate, logically impossible” (p. 14). Accepting this point, one is faced with the following dilemma. On the one hand, if autonomy is a necessary condition of responsibility, then it appears (given Wolf’s reasoning) that it may be impossible for anyone to be responsible. On the other hand, if autonomy is not a necessary condition of responsibility, then it remains obscure what conditions do ground responsibility; in particular, further explanation needs to be provided which would justify attributing responsibility to nonautonomous agents while still preserving the intuitions gleaned from consideration of the paradigm cases mentioned above. As stated earlier, a central aim of Wolf’s book is to provide insight and some degree of relief from the problems of free will and moral responsibility so construed.

In the next two chapters, Wolf considers in detail each horn posed by her “dilemma of autonomy.” Starting with the latter horn, chapter 2 presents and criticizes a nonautonomous conception of free will and responsibility which Wolf terms the “Real Self View.” This view takes its lead from traditional compatibilist accounts which would seek to distinguish instances of responsible action from Wolf’s paradigm instances of nonresponsible action (e.g., action under duress, hypnosis, or kleptomania) without adverting to any requirement of autonomy. To accomplish this goal, such compatibilist accounts in their simplest formulation hold that free action requires only that “an agent is able to govern her behavior on the basis of her will which in turn can be governed by the set of the agent’s desires” (p. 28). Wolf refines this

simple conception of nonautonomous free action in light of criticisms and amendments suggested by the writings of Frankfurt and Watson, and eventually arrives at a more adequate statement of the Real Self View. Roughly put, the view holds that a nonautonomous agent is responsible for her actions just in case her actions are attributable to her real self—that is, she acts as she does in the absence of undue constraint, where this implies that “she is at liberty (or able) both to govern her behavior on the basis of her will and to govern her will on the basis of her valuational system” (p. 33).

The Real Self View has the merit of being able to explain why we do not hold kleptomaniacs and people acting under duress or hypnosis responsible for their actions, and it does so without having to appeal to the condition of autonomy with its attendant problems. Nevertheless, Wolf insists that the Real Self View suffers from difficulties of its own.

The central problem arises because consideration of examples of mental disorders, deprived childhoods, and so forth provide reason to question whether, and why, an agent should be held responsible for her real self. One response to this worry—which Wolf attributes to Hobart—is to argue that our practices of responsibility do not require us to blame people for having bad characters; rather they are blamed simply in virtue of being fully formed act-makers who act badly (p. 39). But Wolf thinks that in ascribing responsibility, we are concerned with more than simply judging whether a person is a fully formed act-maker. In addition, we want to know whether the person is responsible, in any deep sense, for being the act-maker that she is; that is, we are concerned to judge the moral quality not only of an agent’s actions but also of the individual herself. As Wolf insists, examples of “comprehensive insanity, psychological conditioning, dramatically deprived childhoods suggest that, in fact, real selves may not always be deep selves. In other words, they suggest that some individuals with fully developed real selves may not deserve praise and blame for what they do and what they are” (p. 44). In short, the conception of responsibility suggested by the Real Self View suffers because it cannot explain the difference between those fully developed, unconstrained persons who seem responsible and those who do not (e.g., the kleptomaniac, the psychologically conditioned agent, the person with the dramatically deprived childhood, and so forth).

As a final criticism, Wolf notes that while the Real Self View places a great importance upon the fact that only persons can act in accord with their valuational system, the view does not offer any explanation for why acting in accord with value should matter so deeply, and so “at a fundamental level, [fails] to explain why responsibility makes any sense at all” (p. 47). Thus, Wolf claims that the Real Self View leaves

important gaps in our understanding of responsibility, and in order to fill these shortcomings, it seems natural to appeal once again to the condition of autonomy. Invoking this condition, however, returns Wolf to the first horn of the dilemma of autonomy—that is, exploring how the seemingly impossible requirement of autonomy might be satisfied—and it is to this task that she turns in chapter 3.

In this third chapter, Wolf sets out first to present an autonomous conception of free will and responsibility (the “Autonomy View”) and then to criticize this view. The strategy here is to show that as a more complete picture of autonomous agency is developed, two important claims become clear: (1) that autonomy is not something that we would particularly like to have and (2) that it is not something that we need in order to be responsible. The initial attraction of the Autonomy View derives from the hope that the condition of autonomy might supplement and complete the account of fully developed and unconstrained persons that is presented in the Real Self View. The Real Self View holds that values differ from desires not in their content but in their source: values arise, or at least are supported by, the agent’s Real Self; they are motivations that the agent cares about having. Given this claim, a picture emerges in which the valuing self seems to be able to judge and rise above the merely desiring self. Once this picture is accepted, one easily is led to the thought that “our values unlike our mere desires, can be controlled or chosen in accordance with Reason. Our valuing selves may then be identified with our rational selves” (p. 51). A natural move at this point (at least for one sympathetic to the Kantian tradition) would be to cash out the condition of autonomy by appeal to some conception of rational agency. But Wolf resists this move, claiming that “if Reason is similarly a property or faculty that an agent cannot help having (or lacking), and if it generates motives that an agent cannot help acting upon, then an agent who acts in accordance with Reason is likewise not in ultimate control of her actions” (p. 52).

Wolf argues that an autonomous agent must be no more bound by Reason than she is by desire; the autonomous self cannot be equated simply with the Rational self. Rather, Wolf insists that autonomous selves “must be agents who not only *do* make choices on no basis when there is no basis on which to make them, but who also *can* make choices on no basis even when some basis is available. In other words, they must be agents for whom no basis for choice is necessitating” (p. 55). Given this characterization of autonomous agency, Wolf goes on to argue that we have little reason to want to be autonomous: why, after all, would anyone want to have the power to act against all rational considerations or to make a completely ungrounded choice that ignores all interests and reasons against so deciding? Why would anyone want

the ability to “make choices on no basis even when a basis exists” (p. 55)<sup>21</sup> Since “Reason” is defined as referring to the highest faculty a person possesses, Wolf further claims that anyone who valued such autonomy—that is, who valued the power to act against Reason—must believe that even the reasons and values which are supported by one’s highest faculties are without any intrinsic normative force. Such an advocate of autonomy, then, might be committed to a type of skepticism about the objectivity of value.

After showing that an ability to act autonomously is not to be confused with an ability to act in accordance with Reason, Wolf goes on to argue that the latter ability is more relevant to ascriptions of responsibility than is the former. To support this point Wolf considers the class of agents who share the ability to act in accord with reason, and notes that within this class the only difference between autonomous and nonautonomous agents is that the autonomous agents have an ability to act against Reason which the nonautonomous agents lack. Wolf claims that at best this ability seems irrelevant to ascriptions of responsibility (for agents who have the ability to act in accord with Reason appear to be responsible even though they are not autonomous) and at worst this ability to act against Reason may be incompatible with responsibility (for her earlier arguments suggest that “in extreme cases, involving, for example, self-destructiveness or extraordinary evil, the ability to act in discord once with Reason may indicate some form of insanity” [p. 151]). Wolf concludes, therefore, that autonomy is not a necessary condition for morally responsible agency.

Having presented and criticized both an autonomous and non-autonomous conception of free will and responsibility, Wolf turns in the final three chapters of her book to developing and defending what she takes to be a more promising view—a heteronomous conception of free will and responsibility which she dubs the “Reason View.” In briefest form, this view holds that “the freedom necessary for responsibility consists in the ability (or freedom) to do the right thing for the right reasons . . . to choose and to act in accordance with the True and the Good” (p. 94). In chapter 4, Wolf develops the Reason View by contrasting it with both the Real Self View and the Autonomy View. The point which emerges from this pair of contrasts is that the Reason View may be seen as something of an intermediary position between these other two positions. On the one hand, the Real Self View is too weak; the ability to act in accord with one’s real self is not sufficient to justify ascriptions of responsibility, for questions can arise concerning

1. A similar point is also developed in Susan Wolf, “Asymmetrical Freedom,” *Journal of Philosophy* 77 (1980): 151–66; for a discussion and criticism of this view as it is presented in “Asymmetrical Freedom,” see Mark Ravizza and John Martin Fischer, “When the Will Is Free,” forthcoming in *Philosophical Perspectives*.

one's responsibility for that real self. On the other hand, remedying this insufficiency, as does the Autonomy View, by demanding that one's real self be completely undetermined by anything external to the self is too strong; such a demand for autonomy not only seems impossible to satisfy, it also bestows a power to act against Reason which no one could have any reason to want. Accepting these criticisms, the Reason View locates the freedom needed for responsibility between these two views: the freedom needed to be responsible is the freedom to pursue the True and the Good, to do the right thing for the right reason.

One of the most striking features of the Reason View is that unlike either the Real Self View or the Autonomy View which take the problems of free will and responsibility to be purely metaphysical problems, the Reason View sees these problems as being inextricably linked with normative considerations. That is, the question of responsibility is no longer simply a question of whether one's actions are completely determined by a real self that in turn is metaphysically independent from any external determining forces. For the Reason View, what is required for responsibility is not a solely metaphysical power, but a "distinctive intellectual power [which has both a metaphysical and a normative component], the power to exercise right Reason and to govern one's actions accordingly" (p. 71). Given this claim, a striking asymmetry emerges from the Reason View: a person who does the right thing for the right reason need not have been free to do otherwise in order to be held responsible for her action; however, a person who does the wrong thing and acts on bad reasons must have been free to do otherwise in order to be blamed and held responsible for the act.

In the remainder of the book, Wolf considers some ramifications of the Reason View and works to dismiss a number of criticisms which inevitably will be leveled against it. The close of chapter 4 seeks to defend the asymmetry found in the Reason View and dissolve "the air of paradox" which surrounds this asymmetry by applying the view to a variety of good and bad actions (pp. 80 ff.). Chapter 5 turns to the task of reconciling the Reason View's freedom requirement with causal determinism. Here Wolf reflects on an imaginative story, reminiscent of Leibniz, in which God chooses to create the best possible world containing morally responsible agents. The purpose of the story is to encourage us to see first that a psychological ability to do otherwise is all that is needed by the Reason View, and second that there is little cause to worry that this type of ability might be incompatible with determinism.<sup>2</sup> Finally, in chapter 6, Wolf examines the meta-ethical

2. We discuss this claim and Wolf's characterization of psychological ability in greater detail below.

assumptions behind the Reason View. She argues, in particular, that although the Reason View presupposes a certain objectivity of value (because it demands an ability to act in accord with the True and the Good), this normative requirement is, in fact, "quite weak and extremely plausible" (p. 21). After describing what she takes to be the spectrum of meta-ethical positions, Wolf maintains that the only such positions which are incompatible with the Reason View are quite strong anti-objectivist positions which reject completely the objectivity of values and deny that there are any nonarbitrary, agent-independent standards that can show one value or system of values to be better than others (p. 126).<sup>3</sup> In closing this final chapter, Wolf propounds a meta-ethical view she calls "Normative Pluralism." Roughly put, Normative Pluralism holds that "although Reason constrains values, it does not constrain them completely, even when the conventions and shared understandings of a person's community and the concrete details of her more specific situation are taken into account" (p. 125). When conjoined with Normative Pluralism, the Reason View's demand that a responsible agent have the ability to act in accord with the True and the Good becomes a demand that the agent be able to recognize (and act upon the recognition) that there may be more than one course of action which is right. Since Reason does not always uniquely determine one course of action as the right one, a responsible agent must appreciate whatever reasons there are for pursuing the available options, and this recognition will allow a degree of freedom for choosing between good alternatives that critics might have feared was precluded by the Reason View.<sup>4</sup>

3. Wolf considers two versions of this antiobjectivist position: "nihilism" which holds that there is no objective good, and "Conceptual Subjectivism" which claims "that while there may be a Good, it is not the sort of thing that *can* be recognized and appreciated" (and hence no moral judgment is better, more rational, more correct than any other). Wolf argues that, although the spirit behind the nihilist's view is "deeply contrary" to the Reason View, these views are not strictly incompatible. For the nihilist the ability to appreciate the True and the Good will simply reduce to an ability to recognize the True, and although this reduction will undermine judgment of praiseworthiness and blameworthiness, it will leave a "thin" notion of responsibility which can mark off actions which "can most properly be regarded as examples of a Romantic sort of Self-Expression" (p. 131). In contrast, conceptual skepticism's denial that there is any possibility of being guided by the True and Good commits such a position to arguing either that the Reason View is false or that the concept of responsibility is incoherent.

4. Wolf recognizes that requiring that responsible agents have the ability to appreciate whatever reasons there are for different courses of actions raises a difficult question concerning "how extensive, how deep, and how clear one's vision has to be" (p. 142). In addressing this question, Wolf appeals to the sense that a complete awareness of all possible reasons certainly is not required in order to be responsible; rather she claims that "our interest in freedom and responsibility is an interest in *enough* freedom and responsibility to satisfy the purposes and interests at hand" (p. 143).

In the discussion that follows, we focus on two central claims that emerge from Wolf's presentation of the Reason View. First, we discuss Wolf's defense of the metaphysical asymmetry embodied in the Reason View and argue that when a more complete set of examples is considered, the asymmetry thesis is seen to be false. Then, we examine the Leibnizian story Wolf presents in chapter 5 and consider whether the Reason View in conjunction with this story really does present some degree of relief from traditional worries about free will and responsibility.

### THE ASYMMETRY THESIS

Wolf argues that, whereas one cannot be morally responsible for a bad action which one could not have avoided performing, one can be morally responsible for a good or right action which one could not have avoided performing. She concludes that what we care about is the freedom to act in accordance with the True and the Good. That is to say, Wolf believes that the important sort of freedom is not a "neutral" freedom which consists in open options to do any of various things, good or bad; rather, it is the freedom to do the Good.

Wolf motivates her claim with the following examples. First, she claims that there are cases in which an agent is morally responsible for doing something good, even though she cannot do otherwise. In one such case, a woman spies a young boy drowning in the ocean, and without consciously deliberating, she jumps into the water and rescues him (pp. 58 ff.). The example stipulates that given the woman's character and the circumstances in which she finds herself, it is literally impossible for her to refrain from saving the child.<sup>5</sup> According to Wolf, one reason the woman might be unable to do otherwise is "simply because her understanding of the situation is so good and her moral commitment is so strong" (p. 82). Further Wolf claims that the woman can be held accountable—can be praised—for her rescue of the child even though she could not have done otherwise.

In contrast, Wolf adduces types of cases in which an agent does not seem morally responsible for what she does and in which the agent cannot do otherwise. For example, she suggests that if conditions like kleptomania, a deprived childhood, or a poor upbringing render an agent unable to do the right thing for the right reasons, then intuitively the person is not responsible for what he has done (pp. 80–81). Further,

5. Why should one think that in such a situation the woman literally is unable to do otherwise? Wolf suggests that the woman might be unable to do otherwise because she finds it "unthinkable" to allow the child to drown (p. 59); elsewhere Wolf contends that the woman might lack "the ability to do otherwise simply because her understanding of the situation is so good and her moral commitment is so strong" (p. 82). These claims are of course controversial. However, for the sake of argument we will accept Wolf's examples as presented.

it seems that in these cases the agent cannot be held morally responsible for what he has done precisely because he could not have done otherwise. Wolf's claim is that, whereas there are at least some cases in which an agent cannot do otherwise and is nevertheless morally responsible for doing something good, there are no cases in which an agent cannot do otherwise and is morally responsible for doing something bad.

We believe, however, that there are cases in which an agent who cannot do otherwise is morally responsible for doing something bad. Thus, because we agree with Wolf that there are cases in which an agent who cannot do otherwise is morally responsible for doing something good, we (contrary to Wolf) believe that good and bad acts are symmetrical with respect to the requirement of alternative possibilities for moral responsibility: in neither the case of good actions nor the case of bad actions is moral responsibility dependent upon alternative possibilities.

To support both our contentions that there are cases in which an agent is morally responsible for a bad action to which there was no alternative possibility and that good and bad acts are symmetrical with respect to the requirement of alternative possibilities, we shall develop two parallel "Frankfurt-type" cases.<sup>6</sup> In such a case, there is a "fail-safe mechanism" which does not actually come into play but which suffices to guarantee the actual result.

Let us first present a Frankfurt-type case in which the agent performs a good act. Call this case "Hero."<sup>7</sup> "Hero" is intended to be similar to Wolf's example of the woman who rescues a drowning child. Martha is walking along a beach, looking at the water. She sees a child struggling in the water, and she quickly deliberates about the matter, jumps into the water, and rescues the child. We can imagine that Martha does not give any thought to not trying to rescue the child, but that if she had considered not trying to save 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. We simply stipulate that in the alternative sequence the urge to save the child would be genuinely irresistible.

Apparently, Martha is morally responsible—indeed, praiseworthy—for her action, although she could not have done otherwise. Martha acts freely in saving the child; she acts exactly as she would have acted, if she had lacked the propensity toward strong feelings of guilt.

Consider, now, a parallel case, "Villain."<sup>8</sup> An evil man, Joe, knows that a young child stands at the end of a long pier each day in order

6. Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66 (1969): 828–39.

7. This example is presented in John Martin Fischer and Mark Ravizza, "Responsibility and Inevitability," *Ethics* 101 (1991): 258–78.

8. For a similar case, see *ibid.*

to watch the sunset. Joe has decided (for his own rather perverse reasons) to push the child off the pier causing her to drown in the violent surf below. Joe is bad, but Max is no better. Max is pleased with Joe's plan to drown the young girl, 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; the electronic device simply monitors Joe's brain activity, but it does not have any causal influence on what actually happens. Joe acts exactly as he would have acted, had no device been implanted in his brain.

Evidently, Joe is morally responsible for what he has done. Indeed, Joe is blameworthy for pushing the young girl off the pier. But whereas Joe is morally responsible for his action, he could not have done otherwise. Joe could not have done otherwise because of the existence of a "counterfactual intervener" (Max) who would have caused him (in a certain manner) to behave as he actually did, had Joe been inclined to do otherwise. We might say that Joe acts freely and is morally responsible for what he does because no "responsibility-undermining factor" operates in the actual sequence leading to his action. Rather, such a factor—Max's use of the electronic device to stimulate Joe's brain—operates in the alternative sequence. In cases in which a responsibility-undermining factor operates in the alternative sequence but not in the actual sequence, an agent can be held morally responsible for an action, although he could not have done otherwise. And this seems to hold even in cases in which the agent does something bad.

"Hero" and "Villain" are parallel Frankfurt-type cases. In "Hero" an agent is morally responsible for doing something good, although she could not have done otherwise. In "Villain" an agent is morally responsible for doing something bad, although he could not have done otherwise. Thus, Wolf's asymmetry thesis is false; rather, good and bad actions are symmetric with regard to the requirement of alternative possibilities for moral responsibility.<sup>9</sup>

9. This account of Wolf's asymmetry thesis and our criticism of it are based on Wolf's repeated claim that "according to the Reason View, the freedom necessary for

A more careful scrutiny of "Villain" will help to illuminate the motivation for the asymmetry thesis and its ultimate inadequacy. Let us say that a factor which intuitively rules out moral responsibility is a "responsibility-undermining factor." Now we can divide the class of cases in which an agent is not morally responsible for what he does and cannot do otherwise into two proper subclasses: those in which the responsibility-undermining factor operates in the actual sequence

---

responsibility consists in the ability (or freedom) to do the right thing for the right reasons" (p. 94). Of course, any interpretation of this claim crucially depends upon how one understands the phrase "the ability to do the right thing for the right reason." In interpreting this phrase, we have relied upon the characterization of ability which Wolf offers in chap. 5. Here Wolf states that the claim that an agent is able to do *X* may be understood as a claim that (1) the agent possesses the capacities, skills, talents, knowledge, and so on which are necessary for *X*-ing and (2) "that nothing interferes with or prevents the exercise of the relevant capacities, skills, talents and so on" (p. 101). Given this characterization of ability, it seems clear that in "Villain" Joe is not able to do the right thing for the right reason, for were he to try to do so, Max's device would prevent his exercising the relevant capacities, skills, etc., required to refrain from pushing the child into the water. Thus, "Villain" offers a case in which an agent is not able to do the right thing for the right reason, and nevertheless, he does intuitively seem responsible and blameworthy for his act. We, therefore believe that "Villain" offers a strong counterexample to Wolf's view. It might be objected that in more cautious moments, Wolf seems to qualify her claims in a way that would protect her position from our criticisms. In particular one might cite a formulation of the asymmetry thesis which Wolf offers in chap. 4. Here she writes, "The Reason View is thus committed to the curious claim that being psychologically determined to perform good actions is compatible with deserving praise for them, but that being psychologically determined to perform bad actions is not compatible with deserving blame" (p. 79). Wolf clarifies her notion of psychological determinism later by saying that agents are not psychologically determined to make the particular choices they do, if "when those agents chose what to do, it would have been equally compatible with their psychological histories in conjunction with all the psychological laws that apply to them that they had chosen something else" (p. 103). Given these definitions, our example of "Villain" is arguably not a counterexample to Wolf's asymmetry thesis. This is because Joe is not psychologically determined, in Wolf's sense, to push the child off the pier; hence it is no objection to this formulation of the asymmetry thesis to note that he should be blamed for his misdeed. To address this type of objection our previous example of "Villain" may be reformulated in such a way that it undermines even this more qualified version of Wolf's asymmetry thesis. In the reformulated version of the example, we stipulate that Max uses his device to implant in Joe an irresistible desire to push the child off the pier. Since Joe already is inclined to push the child, the implanted desire plays no role in Joe's actual deliberation, choice, and action; nevertheless, were Joe even to begin to consider not pushing the child, the implanted desire would come into play, literally making it impossible for Joe to refrain from shoving the child into the treacherous surf. Since the implanted desire plays no role in producing Joe's action, it intuitively seems that he is still blameworthy for his crime, and this is true even though he is psychologically determined to perform the bad action. We, therefore, conclude that the version of Wolf's asymmetry thesis which is qualified in the way just suggested (employing the notion of psychological determination) is false: considerations parallel to those which undermined the original version of the thesis also controvert this version.

and those in which the responsibility-undermining factor operates in the alternative sequence. What is striking about Wolf's examples of situations in which an agent is not responsible for doing something bad and cannot do otherwise is that they are all examples in the first proper subclass: they are all cases in which the responsibility-undermining factor operates in the actual sequence. But one cannot generalize to all cases from features of cases in a proper subclass. That is, one cannot conclude that there are no cases of agents being responsible for bad actions which they could not avoid doing simply because there are no such cases in the proper subclass of cases in which the responsibility-undermining factor operates in the actual sequence. And Frankfurt-type cases such as "Villain" are precisely cases in which the responsibility-undermining factor operates in the alternative sequence. Thus, it can be seen that Wolf is guilty of inappropriately generalizing from a proper subclass of the relevant cases.

One way of attempting to save Wolf's asymmetry thesis would be to restrict its application to those cases in which the responsibility-undermining factor operates in the actual sequence. But whereas this does rule out Frankfurt-type cases, it clearly does not preserve the asymmetry thesis insofar as there would be no cases of moral responsibility for good actions.

Let us investigate a more promising way of revising Wolf's asymmetry thesis. Let us say that a factor which renders some result inevitable is an "inevitability-generating factor." Now notice that all of Wolf's examples (of good and bad actions) are cases in which some inevitability-generating factor operates in the actual sequence. If we restrict the application of the asymmetry thesis to such cases, we can rule out as irrelevant "Frankfurt-type cases."

Of course, there are some cases in which an inevitability-generating factor operates in the actual sequence, the agent performs a good action, and the agent is not morally responsible for what he does. Consider, for example, a case similar to "Hero." In this version of the story, another woman, Maxine, has secretly installed a device in Martha's brain which enables her to control Martha's decisions and actions by electronically stimulating her brain. Let us suppose that Maxine's device is activated in the actual sequence, causing Martha to rescue the child. Presumably in such a case Martha would not be morally responsible for saving the drowning child. This need not be any sort of problem for the revised asymmetry thesis, however. This thesis need not be construed as holding that in all cases in which an agent does something good and some inevitability-generating factor actually operates the agent is morally responsible for what he does. Rather, the revised asymmetry thesis, properly construed, is that there are some cases in which an agent does something good, an inevitability-generating factor actually operates, and the agent is morally responsible for what he

does; but there are no cases in which an agent does something bad, an inevitability-generating factor actually operates, and the agent is morally responsible for what he does. Of course, this is a highly attenuated asymmetry thesis. More important, even if this weaker version of the asymmetry thesis is true, it is doubtful that it offers any relief from traditional worries about free will and responsibility. We will argue for this point at the end of the next section of this review, but first we need to explain more clearly what the traditional worries are and how Wolf hopes to address them.

## THE ABILITY TO DO OTHERWISE

### *The Story and the Consequence Argument*

One of the most distinctive and original features of the Reason View is its claim that the freedom necessary for responsibility is the freedom to do the right thing for the right reason. Thus far we have suggested one way in which Wolf's formulation of this view may need to be amended, but a key claim of her project still remains to be discussed—namely, that the Reason View meets the stated objectives of a book that is unabashedly devoted to solving the problems of free will and responsibility. Does identifying the freedom to do the right thing with the freedom required for responsibility provide a fruitful way of interpreting and understanding our accessibility to the reactive attitudes? Does the Reason View provide the promised relief from the worries about free will and responsibility that have long plagued philosophers?

In one obvious sense, the Reason View does meet the book's objectives. To the extent that the problems of free will and responsibility are taken to be structured and motivated solely by Wolf's "dilemma of autonomy," the Reason View does provide a solution: it resolves the dilemma by holding simply that autonomy is not a necessary condition of responsibility; rather, the freedom required for responsibility is the freedom to act in accord with the True and the Good. Unfortunately traditional worries about free will and responsibility extend beyond the set of concerns captured by Wolf's formulation of the dilemma of autonomy. Thus, in a deeper sense, the question remains whether the Reason View does provide the resources to relieve some of these other more traditional, metaphysical worries about free will and responsibility. To take but one example, consider the traditional debate over the compatibility of free will and causal determinism.<sup>10</sup> A natural way to formulate this worry is to begin by claiming that it

10. Another traditional set of worries which Wolf does not address, and which are raised by her own Leibnizian story are concerns about the compatibility of God's foreknowledge and freedom to do otherwise; see John Martin Fischer, ed., *God, Foreknowledge, and Freedom* (Stanford, Calif.: Stanford University Press, 1989).

is surely possible that the thesis of causal determinism is true. And if this thesis were true and responsibility required the ability to do otherwise, then no one would be responsible insofar as causal determinism is not compatible with having the relevant ability to do otherwise. Roughly put, the argument for the incompatibility of determinism and freedom to do otherwise runs as follows: "If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us."<sup>11</sup> Certainly, this "Consequence Argument" (as van Inwagen dubbed it) needs to be presented with greater detail and care if it ultimately is to prove defensible, but such refinements need not detain us here.<sup>12</sup> What is important is that this simple argument raises the worry that, if causal determinism is true, then no one has the relevant ability to do otherwise, and this conclusion directly threatens the attributions of responsibility and the reactive attitudes Wolf is anxious to defend.

In raising this worry, the first thing to note is that Wolf initially distances herself from trying to answer such metaphysical concerns directly:

This book will not even attempt to refute the view that negative answers to the metaphysical questions of responsibility and free will are unavoidable. To the extent that this book offers answers to these questions, they are answers which will not guarantee that we are, in the full metaphysical sense of these concepts, free and responsible beings. In fact, if the views in this book are correct, there could be no such guarantee. Nonetheless, the book addresses the metaphysical questions in a spirit that takes pessimism about them to be unwarranted, and in the hope that patience, in philosophy as elsewhere, will have its own proverbial rewards.  
[P. 22]

At least part of Wolf's point here is to anticipate that the Reason View does not require that one be free in a full metaphysical sense (i.e., autonomous) in order to be responsible. Indeed, as she later stresses, an important way in which the Reason View places the problem of responsibility in a new light is by replacing the old requirement for complete metaphysical autonomy with a new, normative requirement that responsible agents simply be able to act in accord with the True

11. Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), p. 16.

12. Interested readers can find careful developments of different versions of this argument in John Martin Fischer, "Incompatibilism," *Philosophical Studies* 43 (1983): 127–37; van Inwagen; and Carl Ginet, *On Action* (Cambridge: Cambridge University Press, 1990).

and the Good. Wolf writes, "Once we have identified the ability that this radical metaphysical independence seemed necessary to secure, however, we can see first, that metaphysical independence is not necessary after all, and second, that we can state the condition of freedom and responsibility more directly by referring outright to the ability to act in accordance with (and on the basis of) the True and the Good" (p. 73). But the point to be stressed here is that even though the Reason View's introduction of a normative requirement does frame the conditions for responsibility in a new, more normative way, it is not a way which completely erases all metaphysical demands. Some degree of metaphysical independence still is required, for, according to the Reason View, if an agent is to be blamed for her bad acts, she must have been able to do otherwise—she must have been able to do the right thing for the right reason. Granting this point, the Consequence Argument's conclusion that determinism is incompatible with having the ability to do otherwise appears nearly as worrisome as ever. Should determinism prove to be true and the Consequence Argument sound, then no one would be appropriately blamed for performing bad actions because no one would be free to do otherwise.<sup>13</sup>

In chapters 4 and 5 Wolf does take up the worry that determinism might pose a threat to the Reason View's conception of responsibility, and she provides some suggestions for how her view could address this type of objection. Wolf notes that the ability required by the Reason View can be analyzed into two narrower abilities: "The first is an ability of thought, the ability to *know* what is in accordance with the True and the Good; the second is an ability of execution, the ability to convert one's knowledge into action. With respect to a bad-acting agent, then, we need to know whether she could have known better and whether, knowing better, she could have acted better" (pp. 87–88). Later, in a footnote, Wolf acknowledges that one might worry that neurophysiological determinism might provide reason to think that agents are unable to know or do otherwise, and hence that agents who act wrongly are not blameworthy because they are unable to act in accord with the True and the Good (p. 88, n. 6). To answer this

13. One might object that in raising this criticism we are failing to take seriously Wolf's qualification in the above quotation. After all, in this passage Wolf states that she deliberately will not even attempt to refute certain negative, metaphysical arguments, and hence given her own construal of her project, it may be that she is not concerned to address traditional worries such as those captured by the consequent argument. We certainly do not mean to suggest that Wolf's project requires her to address all traditional formulations of the problems of free will and responsibility. However, even if Wolf's own project does not entertain certain formulations of traditional worries, we still believe that the question of how the Reason View might provide the resources to meet these worries is of interest in its own right. (Also see Wolf's n. 2, p. 151, and our discussion below.)

concern, Wolf offers a Leibnizian story which is intended to illustrate two morals: first, that the ability to do otherwise which is relevant to responsibility is a psychological ability, not a physical one; second, that such a psychological ability to do otherwise is compatible with causal determinism.<sup>14</sup>

The story, briefly summarized, runs as follows. Imagine there is a Leibnizian-type God who chooses to create the best of all possible worlds. In this world, God wants there to be persons who have the ability to exercise practical and theoretical reason, and who are able to deliberate and act on the basis of their deliberations. It is also extremely important to God that these persons not be psychologically determined to choose and act as they do. After considering all possible worlds, God actualizes the best one, and, for reasons not specified, this world happens to be one in which the psychologically undetermined persons are “composed of stuff that is subject to completely deterministic physical laws” (p. 107). In this best of all possible worlds, Wolf tells us, there is a professor named Rose who chooses one evening to watch a movie on TV rather than grade her students’ exams. And the question Wolf asks is whether Rose could have done otherwise—whether she could have graded papers instead of watching the movie? Wolf readily concedes that given the details of her story, it is both physically and, so to speak, divinely impossible for Rose to grade the papers. Nevertheless, Wolf insists that in a sense relevant to an assessment of Rose’s freedom and responsibility, she could have done otherwise.<sup>15</sup>

In support of this claim Wolf emphasizes that Rose was created such that it is compatible with “her psychological history up to the

14. According to Wolf the attribution of a psychological ability to an agent consists of two claims: a positive claim which holds that the agent has the relevant capacities, skills, and talents, and a negative claim which holds that nothing prevents the agent from exercising the relevant capacities. Wolf argues that whereas there is little reason to fear that determinism would undermine the positive claim that the agent possesses the pertinent capacities and talents, there may be reason to worry that determinism would prevent the agent from exercising these capacities, and a primary aim of her Leibnizian story is to dismiss this fear.

15. It is worth noting that Wolf’s story relies upon what we might call “conditionals of freedom.” That is, it assumes that God can know what agents would freely do in different circumstances, and based upon this knowledge God chooses to create the best possible world. As Wolf writes, “He [God] simply knows (or notices) that a certain possible world contains a woman named Rose who decides to watch TV and He chooses to create that possible world. Omnipotence is not fully comprehensible” (p. 106). The appeal to such conditionals of freedom is, of course, controversial. For discussion of the issues surrounding such conditionals, see Robert Merrihew Adams, “Middle Knowledge and the Problem of Evil,” *American Philosophical Quarterly* 14 (1977): 109–17; Alfred J. Freddoso, “Introduction,” in Luis de. Molina, *On Divine Foreknowledge (Part IV of the “Concordia”)*, trans. Alfred J. Freddoso (Ithaca, N.Y.: Cornell University Press, 1988); Alvin Plantinga, *The Nature of Necessity* (Oxford: Oxford University Press, 1974); and William Hasker, *God, Time and Knowledge* (Ithaca, N.Y.: Cornell University Press, 1989).

moment in question, in conjunction with all the psychological and psychophysical laws that apply to her, that she choose TV, and it is also compatible with these that she choose to grade papers" (p. 108). Since at the level of psychological explanation Rose is able to do otherwise, Wolf argues that she is responsible, and further that it is not relevant that Rose is physically unable to do otherwise. After all, Wolf asks, "Do the physical facts make Rose choose (or explain why she chooses) to watch TV, or does Rose's choice to watch TV make (or explain why) the physical facts come out as they do" (p. 110)? And to this question Wolf answers, "Given God's basis for choosing which world to create, the latter seems more reasonable than the former. For God's choice of the physical facts does not condition but rather is conditioned by His choice of the psychological or personal facts" (pp. 110–11).

Wolf concedes that there is a strong intuition that Rose lacks the relevant ability to do otherwise; this intuition stems from the fact that it is physically impossible for Rose not to turn on the TV coupled with the sense that "where the organism goes, so must Rose" (p. 109). But according to Wolf, this intuition fails, for without further argument there is no reason to give priority to this thought over the thought that "where goes Rose, so must the organism." Why, that is, emphasize the intuition that "Rose's body controls Rose" rather than the intuition that "Rose controls her body"? In short, Wolf contends that there is no reason to take the physical level of explanation as "more basic" than the psychological level of explanation. Unfortunately, Wolf holds that she can offer no precise explanation of what it means for one level of explanation to be "more basic" than another, but she suggests that recalling her story will help us to see that the psychological level is as basic if not more basic than any other level of explanation, for it was of first importance to God to create persons who were psychologically undetermined, and only of a lesser importance that the world be physically determined (pp. 111–12). Presumably, part of the point of the story is to help us to shift perspectives and see that there is no reason to take a physical level of explanation as more basic than a psychological one. Given God's priorities in the story—the fact that his choice of the physical facts is conditioned by his choice of the psychological facts—we are supposed to see that the psychological level of explanation and psychological abilities are what matter in deciding whether an agent is able to do otherwise in the sense relevant to ascriptions of responsibility.

### *Criticism*

The problem with this strategy is that it ultimately fails to provide sufficient grounds to abandon the original intuition which prompted Wolf's story. That is, it fails to provide any cogent reason to abandon

the intuition that, if an agent is physically determined to perform a given morally bad action and is unable to do otherwise (in virtue of this physical determination), then he is not morally responsible for performing the action (pp. 107 ff.). Indeed, one would think that just this sort of intuition would seem all the more compelling and hence be all the less easily abandoned by anyone who is persuaded by that portion of Wolf's asymmetry thesis which holds that one must be able to do otherwise in order to be responsible for bad actions.<sup>16</sup>

Remember that the Reason View holds that in order to be responsible an agent must be able to do the right thing for the right reason. This means that in order to blame an agent for performing a bad action, it must be the case that she was able to do otherwise. This ability would seem to be vitiated if it were the case that in either a psychological or physical sense the agent was not able to do other than what she did. Consider a particularly heinous crime, such as murder. The proponent of the Reason View must hold that a murderer should be blamed only if he could do the right thing. Surely it would be implausible for such a theorist to claim that the murderer should be blamed even though it is physically impossible for him to do anything but pull the trigger of his gun. A proponent of the Reason View—who emphasizes that one who behaves badly must be able to behave well if he is to be morally responsible—will surely be under some pressure to interpret the relevant sort of ability as physical (as well as psychological). For such a theorist needs to answer the question of how it could be fair to hold (even a psychologically undetermined) murderer responsible for pulling the trigger (and thus committing murder) if his behavior was physically determined and that determination resulted in his inability to do anything but pull the trigger.

The story does not provide any cogent reason to abandon this claim. First, Rose's choice to watch television is a relatively innocuous one; perhaps we do not balk at holding her responsible because nothing much is at stake here. But the situation is evidently different in regard to a morally heinous action such as murder. It does not much matter whether we bother to hold Rose responsible, and no serious unfairness would result from doing so. The example of Rose appears to gain

16. We, of course, think that the Frankfurt-type examples, like those presented in the section, "The Asymmetry Thesis," above, do provide good reason to doubt the intuition that, if an agent is unable to do otherwise, then he is not morally responsible for performing the action. However, these examples do not in themselves provide reason to doubt the intuition that if an agent is physically determined to perform a given morally bad action, then he is not morally responsible for doing so. Moreover, a defender of the Reason View would hardly want to avoid worries about determinism by appealing to such examples, for as we argued above, the Frankfurt-type examples provide independent reason to reject (or at least generously reformulate) the Reason View's asymmetry thesis.

plausibility through focusing on a relatively insignificant action, which does not engage our intuitions of fairness precisely because of its triviality.

Second, it is unclear what the relationship is supposed to be between the world envisaged by Wolf and our world. That is to say, Wolf imagines a possible world in which God creates for a certain reason (and in a certain way). She generates conclusions about relative basicness of levels of explanations (and ultimately about freedom). But what exactly is the relationship between the envisaged world and the actual world? Surely it would be highly controversial to suppose that our world was created in the way envisaged in Wolf's story. It would surely be unattractive to base one's claims about our freedom in this world upon the supposition that the world was created in the way imagined in the story. But what then is the point of invoking this (presumably imaginary) world? If the point is that it is possible that the world have been created in this way, it becomes unclear how this could support the pertinent claims about explanatory basicness and freedom in our world. (Evidently, the most that would follow from the mere possibility that the world was created in the way envisaged in the story is the possibility that Wolf's claims about the explanatory hierarchy and freedom are true. But this is not something that was in dispute.)

Finally, we do not see why a ranking of levels of explanation as regards the degree to which they are basic is relevant to an assessment of which notion of ability is germane to ascriptions of responsibility. Clearly, the purposes of explanation are multifarious, and what counts as a good explanation is notoriously context-dependent. Similarly, basicness of explanation will be relative to purposes and will be context-dependent. It is mysterious how an abstract consideration of basicness apart from any particular context of explanation could be relevant to identifying the pertinent sort of ability. Further, it is completely unclear what the relevant context of explanation would be, relative to which basicness should be assessed, where this assessment would allegedly help to fix the notion of ability relevant to ascriptions of moral responsibility.

Wolf suggests that given God's priorities in the story, we can see that (in the story) the psychological level of explanation is more basic. But even if there is some perfectly reasonable context of explanation relative to which we can say that it follows from God's priorities that the psychological level of explanation is more basic, why should this fact have any bearing on which notion of ability is relevant to ascriptions of responsibility? That is, even if God created the physical facts to fit with what he intended as the psychological facts and thus there were a sense in which the psychological level of explanation is more basic than the physical, why should this have any bearing on the issue of which notion of ability is relevant to ascriptions of responsibility?

Suppose, for instance, that God had created the psychological facts to match what he intended the physical facts to be. That is, suppose God had wanted a certain physical configuration to obtain, and he thus created Rose with a psychological ability to do otherwise (in Wolf's sense) exactly like that in the story. Under these circumstances, presumably, the physical level of explanation would be counted as basic (or at least more basic than the psychological). But in this situation the intrinsic facts pertaining to Rose's action—including her psychological ability to do otherwise—would be precisely the same as in the story. Further, under such circumstances it would seem that the appropriate responsibility ascriptions would be the same as those in the story. Thus, the issue of relative basicness of levels of explanation (as fixed by God's priorities) would not seem to be relevant to the assessment of which notion of ability is pertinent to ascriptions of responsibility.

We have pointed out that there is a strong intuition—one which should be attractive to proponents of the Reason View—that in order to be responsible for a morally bad action, the agent must not have been physically determined to behave as he did (and thus physically unable to do otherwise). We have pointed out that the ingredients of Wolf's story—the example of Rose, the notion of relative basicness of levels of explanation, and God's priorities—do not provide any reason to doubt the intuition. Even after Wolf's story, then, we are, at best, at an impasse.

Wolf apparently disagrees with this conclusion. In a footnote to chapter 5 she writes, "this line of argument [i.e., that the story shows the physical level of explanation is not more basic than the psychological level] might be used as part of a response to the Ginet-Van Inwagen argument for incompatibilism" (p. 151). To support this claim, Wolf notes that the Consequence Argument depends upon the premise that it is not in anyone's power to affect the laws of nature or the past, and

curiously, there is a sense in which the persons in the story might be said to have the power (by way of God's foreknowledge and interests) to affect the laws of nature or states of the world prior to their birth. Thus, one might argue that the Ginet/Van Inwagen argument does not apply to the persons in the story. If this were admitted, however, along with the conclusion that these persons were free though determined, it would throw doubt on the strength of the argument even in our own case. If the psychological abilities possessed by the people in the story are sufficient for their freedom and responsibility, why wouldn't those same abilities be sufficient for us? [Pp. 151–52]

We fail to see how Rose can affect the laws of nature or the past in any way that is relevant to a judgment that Rose is free to do otherwise in the sense required for moral responsibility. Admittedly,

Rose's psychologically undetermined choices are one of the factors which lead God to actualize this best of all possible worlds, and in this "curious" sense it might be said that Rose's choice to watch TV affects (via God's foreknowledge) the states of the world before her birth. But the important point to see is that Rose cannot affect the past or the laws in any way that would give her an ability physically to do anything other than watch TV, and this is what the incompatibilist worries about. That is, there is an important difference between affecting the past and changing the past (in the sense of so acting that the past would have been different from what it actually was). In the story, Rose (arguably) affects the past, but the story provides no reason to think that she can so act that the past would have been different from what it actually was. And this latter ability is required in order to block the incompatibilistic argument.

Indeed, in some other possible world Rose (or her counterpart) does something other than watch TV, and in that possible world the past and/or the laws are different from the actual world. But to note that there is such a possible world (which God foresees and chooses not to actualize) in no way establishes that in the actual world Rose can choose to do otherwise and thereby change the past or the laws. Of course, Wolf cannot simply insist that psychological abilities of the sort possessed by Rose are sufficient for freedom to do otherwise (and responsibility) without begging the question in an egregious fashion against incompatibilists like Ginet and van Inwagen. The bottom line is that the story provides no reason to reject the incompatibilist's argument.

#### *The Asymmetry Thesis and the Consequence Argument*

We argued above that the asymmetry thesis (allegedly entailed by the Reason View in conjunction with other facts) as presented by Wolf is false. We suggested another version of the asymmetry thesis: there are some cases in which an agent does something good, an inevitability-generating factor actually operates, and the agent is morally responsible for what he does; but there are no cases in which an agent does something bad, an inevitability-generating factor actually operates, and the agent is morally responsible for what he does. This thesis—as opposed to other related theses—has a chance of being true.

Yet even if this weaker version of the asymmetry thesis were true, it would remain doubtful that it could offer any relief from the traditional worries given expression by the Consequence Argument. The weaker asymmetry thesis still claims that there are no cases in which an agent is morally responsible for a bad action which is produced by an inevitability-generating factor. And remember that the consequence argument implies that causal determination is just such a factor. Thus, whether the Reason View can countenance moral responsibility for

bad actions depends on such issues as the truth of causal determinism and the soundness of the Consequence Argument—the bad old traditional problems. Given that a reasonable theory will countenance responsibility for bad (as well as good) actions, the Reason View is evidently no less vulnerable to worries issuing from the traditional metaphysical problems than is the Autonomy View.<sup>17</sup>

## CONCLUSION

Wolf's book is highly intelligent, original, and provocative. Her criticisms of the Autonomy View and the Real Self View are both vigorous and incisive. Her alternative approach—the Reason View—is developed with considerable subtlety and refinement. It is a distinctive approach to free will and moral responsibility, which deserves to be taken seriously.

But does the Reason View meet the goal of a book which is “unabashedly devoted to solving” the problems of free will and moral responsibility? Has it lived up to its promise to provide a new way of structuring the old problems and to offer some “degree of relief” from these problems? We have argued that it does not. In particular we have suggested that the asymmetry thesis embodied by the Reason View is false, and further that even if this thesis is reformulated in a way that has a chance of being true, the View still cannot assuage traditional worries pertaining to the relationship between free will and causal determinism. Thus, whatever its other virtues, the Reason View has not provided relief from the vexing problems of free will and moral responsibility.

17. It is worth noting that the consequence argument not only raises worries about our reformulated version of the asymmetry thesis, but it raises the same worries about Wolf's original version of the asymmetry thesis which we discussed in “The Asymmetry Thesis,” above. One formulation of the asymmetry thesis which would avoid such worries is Wolf's more restricted claim (embodying her technical notion of psychological determination) that “being psychologically determined to perform good actions is compatible with deserving praise for them, but that being psychologically determined to perform bad actions is not compatible with deserving blame” (p. 79). But though this version may avoid the particular worries posed by the Consequence Argument, there are good reasons for holding that this thesis is false. (See the discussion of the revised “Villain” case in “The Asymmetry Thesis,” above.)