COMPATIBILIST VIEWS OF FREEDOM AND RESPONSIBILITY

ISHTIYAQUE HAJI

Determinism, the doctrine that the nonrelational facts of the past and the laws of nature entail one unique future, has been thought by many to be incompatible with freedom and moral responsibility for reasons that include the following. Traditionally, the most influential view about the sort of freedom required for responsibility postulates the availability, at various points in our lives, of genuinely accessible alternative possibilities. But there are powerful arguments for the conclusion that determinism expunges such possibilities and thus undermines the right sort of freedom for responsibility. A second concern clusters around the idea that there is no room for "authentic agency" in the world if determinism is true. There are different ways to crystallize this somewhat amorphous idea. According to one prominent view, an agent is morally responsible for her behavior only if the antecedent actional elements, like her values, desires, or beliefs that cause that behavior, are "truly her own"; they are not, for example, the product of direct, surreptitious implantation. But it has been claimed that if determinism is true, this "authenticity" condition can never be met because our springs of action are ultimately the product of events long before our births and hence products of external sources over which we have no control. (See, for example, G. Strawson 1986, Kane 1995, 1996a and b; Pereboom 1995, 2001.)

Compatibilist theories all share the common presumption that determinism does not undermine freedom and responsibility. The success of a compatibilist account depends partly but pivotally on how well it responds to the two aforementioned putative threats that determinism poses to responsibility. In this chapter, we will survey two broadly different kinds of compatibilist theories. Advocates of the first kind propose that responsibility is constituted by the "reactive attitudes" such as indignation, forgiveness, resentment, guilt, gratitude, and love that we display toward one another in response to behavior and traits of character. As there are no independent grounds external to the range of these attitudes that are relevant to responsibility ascriptions, the thought is that such an account of responsibility is immune to threats of determinism. Proponents of the second kind of compatibilist theory develop the view that a person is responsible for his behavior if there is an appropriate "fit" between that behavior and various psychological elements of his or various features of the world. As this fit can obtain even if the world is determined, determinism, it has been claimed, is no threat to responsibility.

1. P. F. Strawson's Account

One of the most influential advocates of the first kind of compatibilist theory is P. F. Strawson, who develops his account with the objective of reconciling traditional disputants whom he calls "optimists" and "pessimists" in the free will debate (1962: 59-60). Optimists, who are compatibilists, defend a consequentialist conception of responsibility, holding that responsibility ascriptions like blaming and praising judgments are to be understood and justified by appeal to the useful consequences that follow from them. On their view, responsibility ascriptions provide a means of regulating social behavior, and such regulation, they claim, can be efficacious even if determinism is true. Pessimists, who are incompatibilists, insist that the sort of freedom required for responsibility is "contra-causal." Against the optimists, they argue that being responsible requires that the agent be deserving of, for example, blame, but this requirement of desert cannot be captured by the social regulation view. The view, hence, leaves a glaring gap in our conception of responsibility, a gap, pessimists say, that is to be bridged by the thesis of contracausal freedom.

Strawson rejects both optimist and pessimist accounts, but he hopes to draw concessions from either in an effort to better conceptualize responsibility. From the optimist, Strawson wants the admission that the social regulation view does overlook something vital to responsibility. From the pessimist, he desires acknowledgment that the vital element overlooked is not contracausal freedom, but the
proper role that the reactive attitudes and feelings play in our interpersonal lives, which furnish the arena in which responsibility ascriptions are made (ibid.: 78–80).

In Strawson's view, the question about the conditions under which an agent is morally responsible is identified with the question of the conditions under which it is appropriate to hold an agent morally responsible. These conditions, in turn, are explained in terms of susceptibility to reactive attitudes. Strawson proposes that holding an agent morally responsible is an expression of certain basic needs and aversions: "It matters to us whether the actions of other people...reflect attitudes toward us of good will, affection, or esteem on the one hand or contempt, indifference, or malice on the other." The reactive attitudes are "natural human reactions to the good or ill will or indifference of others towards us as displayed in their attitudes and actions" (ibid.: 67); and they express "the demand for the manifestation of a reasonable degree of good will or regard, on the part of others, not simply towards oneself, but towards all those on whose behalf moral indignation may be felt" (ibid.: 71). Responsibility, then, is nothing more than—it is constituted by—our adopting these attitudes toward one another. On Strawson's view, holding responsible is as Gary Watson comments, "as natural and primitive in human life as friendship andanimosity, sympathy and antipathy. It rests on needs and concerns that are not so much to be justified as acknowledged" (Watson 1987b: 259).

Strawson's account has generated a great deal of insightful, critical discussion. A worry developed by John Fischer and Mark Ravizza (1993: 18) builds on the strong connection that Strawson makes between being responsible and holding responsible. The worry is that one can hold a person responsible for something when intuitively he be is not (and conversely, one can intuitively be responsible for something without being held responsible). As an illustration, Fischer and Ravizza ask us to imagine a society in which some class of people are systematically treated only as objects to be used in the interest of social utility, and others have no reactive attitudes toward members of the group. This fact alone, they claim, would not suffice to warrant that these persons are not morally responsible.

A cluster of criticisms has been directed against Strawson's response to the following twin claims of the pessimist. First, if determinism is true, we have reason to abandon the reactive attitudes and practices associated with moral responsibility, for determinism implies that excusing considerations apply to all human action and thus hold universally. Second, if we do have reason to abandon these attitudes and practices, we are psychologically and practically capable of doing so. In response to the first, Strawson deploys the "rationalistic strategy," whose crux is that it would not follow from the truth of determinism that the typical excusing conditions for responsibility—such as acting in ignorance, or accidentally, or without forethought, or when psychologically abnormal, or morally undeveloped—exempt all causally determined agents from responsibility. For deter-

minism does not entail that all our actions are done out of ignorance, accident, and so forth. In Strawson's summation, "it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition" (1962: 68). So determinism would not provide us with any reason to modify, eschew, or suspend our reactive attitudes.

A concern against this strand of the rationalistic strategy is that reflection on one's developmental history can and does affect our reactive responses. For example, Watson (1987b: 267–80) has argued, drawing on Robert Harris's case, that appreciation of the causal influences of an abusive childhood on one's subsequent actions often does have the effect of mitigating degrees of blame. But if this is so, why can't reflection on the causal effects of determinism on our actions also serve to affect our reactive responses?

A second concern is that Strawson incorrectly construes pessimists as claiming that if determinism is true, then everyone would be abnormal, and that the abnormal cannot be held responsible. The pessimists' worry is really that determinism would incapacitate everyone in some way that undermines responsibility. Pessimists, for instance, might charge that no one can exercise contracual causal freedom if all events are determined, but this sort of freedom is required for responsibility. (This criticism is raised in Russell 1992).

A second strand of the rationalistic strategy suggested by Strawson and developed by Jonathan Bennett (1986) is that, if we assume the truth of determinism, even if we did have choice over whether to abandon the reactive attitudes, it would be rational to retain them and our other practices of holding people responsible, in that the cost of abandoning these things would be too high. Given the connection between responsibility and the reactive attitudes, and given that these attitudes are inextricably associated with our interpersonal relations, to stop holding people responsible for forsaking the reactive attitudes would be to sacrifice all interpersonal relationships. This would greatly impoverish human life.

Contesting this strand, Derk Pereboom (1995; 2001: ch. 4) has argued that although determinism would undermine judgments of responsibility like those of praiseworthiness and blameworthiness, abandoning responsibility in the face of determinism would not threaten good interpersonal relationships. In Strawson's view, some of the attitudes most important for interpersonal relationships are anger, resentment, forgiveness, gratitude, and mature love. Pereboom proposes that a number of these attitudes would have "analogues" that would survive determinism and would, hence, underpin or foster the interpersonal relationships that so concern Strawson.

Galen Strawson (1986) has raised a different sort of objection to this second strand of the rationalistic strategy. Against the elder Strawson's view that we would have overriding reasons to retain the reactive attitudes no matter what theoretical reasons like those advocated by incompatibilists were advanced against them, Galen Strawson argues that we have a commitment to truth, a propensity to live in
according with the facts. But then if determinism is indeed incompatible with responsibility, it is not obvious that it would be rational to abandon our commitment to truth in favor of our commitment to sustaining a way of life in which our reactive attitudes figure prominently.

To turn briefly to the naturalistic strategy underlying P. F. Strawson's position, its point is that our human commitment to the reactive attitudes is so "thoroughgoing and deeply rooted" in our nature, that, pure the pessimists, it would be psychologically impossible to give them up or entirely abandon them even if we had theoretical grounds for doing so. It is thus "useless" and "idle" to ask whether or not it would be rational to suspend or abandon our reactive attitudes if determinism were true.

Critiquing this position, Paul Russell (1992) suggests that Strawson fails to distinguish between "type-naturalism" and "token-naturalism." The type naturalist plausibly holds that we have a general disposition to adopt the reactive attitudes which is partly constitutive of human nature, and consequently that this disposition is insulated from theoretical discoveries. In contrast, the token-naturalist says that we are prone to adopt particular tokens of, say, gratitude or resentment, under particular circumstances, and that we are so naturally constituted that we cannot abandon or suspend such tokens of specific attitudes under the relevant circumstances. Russell proposes that whereas type-naturalism is plausible, it cannot be used to discredit the pessimists' view that, confronted by adequate theoretical reasons, we would be able to abandon our reactive attitudes. For the pessimists' view is most charitably construed as the thesis that, in light of theoretical considerations, we are able to abandon tokens of, say, resentment, under certain (responsibility-undermining) circumstances; and that if determinism is true, we are all in such circumstances.

Finally, Strawson's theory can be assessed according to whether it responds adequately to the putative threats of determinism to control and agency outlined previously. We have seen that Strawson holds that being morally responsible is nothing more than being a recipient of the reactive attitudes and a participant in the associated practices. The notion of responsibility is to be understood as largely the expression of our concerns and demands about our treatment of one another. Further, Strawson insists that responsibility is not a function of holding propositions "external" to the practice to be true, such as the proposition that responsibility requires contracausal freedom. So, similarly, one might think that Strawson's response to the putative first threat of determinism would be to say that the proposition that responsibility requires freedom to do otherwise is "external" to the germane practice and hence is not relevant to our understanding of responsibility. Similarly, in response to the putative second threat, he might say that the proposition that decisions that issue from sources over which we have no control are ones for which we cannot be responsible is, once again, "external" to our practices. These responses, though, to the two threats are suspect and we can appreciate this if we further explore the relevant practice.

If practice is construed narrowly to denote simply reactive responses to, or treatment of, others, then Strawson's account is objectionable. For then it would not be possible to explain why some of our reactive responses, like expression of blame, would be misplaced. I might believe that Jack is an appropriate candidate for blame with respect to a particular deed, and react accordingly; but unaware of certain pertinent facts, I might be mistaken. Presumably, though, Strawson would handle this sort of worry by construing practice broadly to include excusing and exempting circumstances. But our practice of responsibility ascriptions includes clarifying the excuses and exemptions themselves, partly, by trying to discern the common features (if any) that they share so that we can frame, for example, legal policy. But a careful look at our practice reveals that several of our excuses tend to cluster around epistemic or control considerations. With respect to the latter, incompatibilists of a certain bent (for example, Edwards 1958) might insist that it is not improbable to assume that such considerations share lack of freedom to do otherwise. Further, incompatibilists of a different bent might argue that a number of the exemptions, such as psychological abnormality or systematic behavior control, share the feature that the agent is not in ultimate control of her decisions or actions; they are, in a transparent sense, the product of sources external to the agent (see, for example, Kane 1996a: ch. 5). But then, again, it is not far-fetched, these incompatibilists might say, to regard determinism as threatening responsibility in an analogous way: if it is true, then no one is in ultimate control of his or her decisions or actions.

2. R. Jay Wallace's Account

R. Jay Wallace (1994), who also develops a Strawsonian account, meets the objection regarding alternative possibilities head-on. Like Strawson, Wallace takes the question about the conditions under which persons are morally responsible to be tantamount to the question of the conditions under which it is appropriate to hold persons responsible; and holding persons responsible involves susceptibility to the reactive attitudes in relation to them. Unlike Strawson, Wallace restricts the reactive attitudes pertinent to responsibility to resentment, indignation, and guilt. These attitudes, he says, "hang together as a class," in that they are linked by related propositional objects: each is "explained exclusively by beliefs about the
violation of moral obligations (construed as strict prohibitions or requirements), whereas other moral sentiments are explained by beliefs about the various modalities of moral value" (ibid.: 12).

On Wallace’s view, a person is morally responsible for some action if and only if it would be appropriate to hold her responsible for that action (ibid.: 91). Because moral norms of fairness, Wallace contends, set the standards of appropriateness for responsibility ascriptions, this view is to be construed as the normative one that a person is morally responsible for an action if and only if it is fair to hold her responsible for that action.

Consonant with this interpretation, Wallace regards incompatibilists as claiming that "it would be unfair (and hence wrong) to hold people responsible if determinism is true" (ibid.: 110). Elaborating, Wallace attributes to incompatibilists the "generalization strategy" according to which what unifies the standard excusing and exempting conditions is that they are all different ways of showing that the agent could not have done otherwise. Because determinism effaces alternative possibilities, it would, consequently, never be fair to hold people responsible. To meet the incompatibilists’ charge, Wallace advances principles of fairness that include, as excusing and exempting conditions of responsibility, the very ones that we acknowledge, but that do not entail that what unifies these conditions is the generalization that agents are unable to do otherwise.

When excuses (such as coercion, physical constraint, duress, mistake, accident, and inadvertence) apply, Wallace proposes that what we find is that the agent has not violated a moral obligation. He explains that something is a moral obligation for an agent only if it is susceptible to direct influence by reasons (ibid.: 131). Mere bodily movements cannot be so influenced but only intentions or decisions, or, as Wallace prefers to say, “choices” of the agent (ibid.: 132). When, for example, a person harms another out of ignorance, the harm is not intentional, and so in most cases does not result from a cause to cause the harm; as such, the harm fails to issue from violation of a moral obligation. Wallace then advances the “no blameworthiness without fault” principle that it is unfair (because undeserved) to blame someone unless he has violated a moral obligation. Surely, Wallace adds, it is “doubtful in the extreme that . . . determinism would entail that people never act on choices that violate moral obligations we accept” (ibid.: 133).

With the exemptions—psychopathy, behavior control, stress, insanity, addiction, childhood, hypnotism, deprivation, or torture—Wallace theorizes that the agent has been deprived of the ability to grasp and apply moral reasons, and the ability to control his behavior by such reasons (ibid.: 155–66). But if holding an agent responsible involves judging it appropriate to sanction her morally, and to sanction her as a result of the violation of moral obligations for which there are moral reasons, then it will be unreasonable to do this if she lacks the ability to apply the moral reasons that sustain the obligations violated. Another plausible principle of fairness is that it is unfair (because unreasonable) to blame someone for violating a moral obligation if she lacks the power to recognize and act on the moral reasons supporting that obligation. Again, Wallace indicates that determinism does not entail that people never have the power to recognize and act on moral reasons that underpin obligations. He concludes that the incompatibilists’ generalization strategy fails, and the Strawsonian view developed vindicates the incompatibilists’ stance that determinism is no threat to responsibility.

I confine critical discussion of Wallace’s rich account to the "no blameworthiness without fault" principle. The principle implies the following:

**No Fault** A person is blameworthy for something (a choice or action, for example) only if that thing is morally wrong.

Wallace, it seems, accepts the Kantian ought implies can principle (ibid.: 231–25), which says:

\[ K \ S \text{ has a moral obligation to perform [not to perform] } A \text{ only if } S \text{ can perform [not perform] } A. \]

A noncontroversial deontic principle that links obligation and wrongness is

\[ OW \text{ Agent } S \text{ ought to do [not to do] } A \text{ if and only if it is morally wrong for } S \text{ not to do } [to do] A. \]

K and OW enjoy both intuitive support and theory-based support—the latter as both are theorems within some of our best theories of the concept of moral obligation (for example, those of M. Zimmerman 1966 and Fred Feldman 1986). These deontic principles, in turn, entail that there is a requirement of alternative possibilities for overall wrong actions:

\[ \text{WAP } \text{It is morally wrong for } S \text{ to do [not to do] } A \text{ only if } S \text{ can refrain from doing [do] } A. \]

Now if determinism eradicates genuine alternatives, and wrong actions require such alternatives, then the truth of determinism implies that no actions (where action is broadly construed to include choices) are wrong. But this result, in conjunction with No Fault, would saddle Wallace with the result, anathema to a compatibilist like him, that no one is blameworthy for anything in a determined world. In summary, if No Fault is true, then an incompatibilist might well exploit this principle, in conjunction with the fact that determinism subverts deontic wrongness, to argue for the view that no one can ever deserve blame for anything in a determined world. Such an argument need not hinge on any considerations...
of fairness. Rather, it would emphasize that alternative possibilities are required for deontic wrongness and that determinism undermines such wrongness.

Suppose, though, that No Fault is false, as many have argued (for example, Brandt 1953; M. Zimmerman 1988; Haji 1996). Then Wallace's account of why excusing conditions subvert responsibility straightforwardly falls because it invokes the "No fault without blameworthiness" principle, and No Fault is a false implication of that principle.

3. Hierarchical Accounts: Frankfurt and Dworkin

We now turn to an altogether different strategy for showing that freedom and responsibility are compatible with determinism. This strategy develops the idea that the sort of freedom required for responsibility is essentially a function of an appropriate "mesh" or connection between an agent's choices or actions and her other actional constituents like desires and preferences. We will start with hierarchical theories and then consider other types of "mesh theory" (a useful term introduced by Fischer and Ravizza 1998: 185).

The hierarchical theories of Gerald Dworkin and Harry Frankfurt are internalist. According to such theories, a person is responsible solely by virtue of facts internal to his psychology. Internalists hold that responsibility also requires that a person's motivations that give rise to choice or action be caused, in an appropriate way, by factors in the external world. Unlike externalists, internalists stress that the causal history of our springs of action are crucial to responsibility assumptions. Internalism appears to enjoy at least one theoretical advantage over externalism: if responsibility does not depend upon history, then compatibilism appears to enjoy at least one theoretical advantage over externalism.

In a seminal article, Frankfurt (1971) asserts that one essential difference between persons and other creatures is to be found in the structure of persons' wills. Unlike simpler animals and young children, are able to form second-order desires and "have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires" (ibid.: 7). Similarly, Dworkin's theory (1988) of personal autonomy, which attempts to capture the core idea of autonomy as self-government, appeals to the distinction between second-order and first-order desires. Unlike a first-order desire, whose object is an action or a state of affairs, a second-order desire's object is another actual or possible desire of the agent whose desire it is.

Dworkin proposes that a person is autonomous "if he identifies with his desires, goals, and values, and such identification is not influenced in ways which make the process of identification in some way alien to the agent" (1988: 61). Roughly, Dworkin conceives of identification as an agent's reflecting critically on a first-order desire and giving "higher order" approval of that desire. In early renditions of his theory that advance necessary conditions of responsibility, Frankfurt invokes both the idea of conformity between one's will (the first-order desire that moves one to action) and a second-order volition (a second-order desire that some first-order desire be one's will), and identification with one's will. Again, the key idea is that responsibility requires that we assess our first-order desires and form second-order preferences as to which first-order desire should move us to action.

Frankfurt has offered different accounts of the notion of identification. So, for example, in one work (1987), he proposes that one identifies with a first-order desire when one has an unopposed second-order volition to act in accordance with it, and one judges that any further deliberation involving other higher order desires about the matter would result in the same decision. In recent essays, Frankfurt appeals to a distinction between passivity and activity with regard to one's desires in order to explain identification. He says that the desires with which a person has identified are "wholly internal to a person's will rather than alien to him; that is, he is not passive with respect to them" (1992b: 8). Further, "inssofar as a person's will is affected by considerations that are external to it, the person is being acted upon. To that extent, he is passive. The person is active, on the other hand, inssofar as his will determines itself" (1994b: 457).

To facilitate discussion, the basic structure of hierarchical theories can be summarized in this way: The hierarchical component is captured by the condition that an agent, S, identifies with a first-order desire, D, to perform some action, A, only if S has a second-order volition, V (relative to D), that D be S's motivating desire. Depending upon how the notion of identification is filled out, the condition will expand to a necessary and sufficient one for identification. We can then add that agent, S, exercises hierarchical control over action, A, if and only if S does A, and A is caused (in a nondeviant, appropriate way) by a desire with which S identifies. Finally, the proponent of an hierarchical theory can be taken to be committed to this principle:

HT A person is morally responsible for performing an action only if she exercises hierarchical control over that action.

This sort of theory encounters problems that have invited illuminating and continuing discussion. Some of the more prominent of these include the following. First, the theory generates some counterintuitive results. For example, an addict
can identify with an irresistible desire for taking some drug. Assuming other conditions of responsibility (like epistemic ones) are satisfied, HT yields the result that the addict is morally responsible for taking the drug, even though, it seems, he is a slave to his relevant desire. Another example concerns akratic action that is free, intentional action contrary to one's best judgment. Assume that Mickey judges it best that he ought not to eat the pie, and he identifies with the desire to refrain from eating the pie. But suppose, succumbing to weakness, he akratically eats the pie. Developing the case in a certain fashion, critics have charged that though Mickey is blameworthy for eating the pie, HT yields the result that he is not, in that he does not identify with the desire to eat the pie. (Compare Mele 1992a for an illuminating discussion of akratic action and HT-type theories.)

A second sort of worry concerns a possible infinite regress of volitions. For an agent to be morally responsible by virtue of the conformity between his will and a second-order volition, V2, V1 must be freely willed. But for V1 to be freely willed, there must be conformity between V1 and some higher order volition, V3, which must in turn be freely willed and so requires a yet higher order volition, and so forth (Watson 1975; Shatz 1985; Friedman 1986, Christian 1991; Zimmerman 1981; and Capers forthcoming). Indeed, the different notions of identification developed by Frankfurt reflect, in part, his efforts to meet this regress objection.

Third, some critics have claimed that HT rests on the unwarranted assumption that the agent's "real self" is to be identified with the cluster of her higher order volitions and those lower order elements selected by them (Thalberg 1980; Berofsky 1993).

Finally, there is the concern that identification can be "engineered" in such a fashion that the agent, contrary to the implications of HT, is not morally responsible. After all, the very components essential to reasoning, including one's values, desires, and beliefs, which are also central to identification, can be acquired via means (like unsolicited, direct electronic stimulation of the brain) that undermine responsibility (Slote 1980; Watson 1987a). The underlying worry here is that hierarchical theories are too internalist; they are insufficiently sensitive to how one acquires one's springs of action.

How might theorists like Frankfurt respond to this last objection? On their behalf, I would urge distinguishing between concerns of responsibility-grounding control and those of ultimacy—roughly, those factors by virtue of which an agent's springs of action are truly an agent's own. The theorists could then be viewed as offering accounts of control and not ultimacy. To elaborate briefly, it has been customary to treat cases involving various sorts of manipulation as test cases for control. But they can also be used to cast doubt on accounts according to which the satisfaction of control (and epistemic conditions) is sufficient for responsibility. That these conditions are insufficient, I believe, is one of the deep morals to be drawn from appropriate cases of manipulation. This recommendation, if heeded, will help to refocus various debates concerning responsibility and control. For example, barring other difficulties with it, one would be misguided to charge that Frankfurt's theory is inadequate by virtue of its control dimension falling prey to suitable manipulation cases, on the assumption that the hierarchical machinery invoked by Frankfurt specifies the control and not ultimacy (or "ownership") requirement of responsibility.

To sum up, how well do HT-like theories cope with the two threats of determinism with which we began? First, to meet the charge that responsibility requires alternative possibilities, Frankfurt (1969a) argues directly against the principle of alternate possibilities ("an agent is morally responsible for something only if he could have done otherwise") by invoking the now famous "Frankfurt-type example" (adumbrated in the discussion that follows). I think this intriguing kind of example is by and large convincing. He then, of course, offers a hierarchical theory that does not imply that the sort of freedom required for responsibility entails the existence of genuinely open alternatives. Theories like HT, however, do not respond well to the second threat of determinism concerning agency. Indeed, I believe that such theories need to be supplemented by an account, roughly, of "authentic agency," or an account of what makes one's springs of action "truly one's own."

Let us now examine further varieties of "mesh" theories, which require some suitable connection between the agent's action and other elements in the external world like values or reasons.

4. FURTHER MESH ACCOUNTS: WATSON AND WOLF

Gary Watson's account (1975), unlike Frankfurt's, is nonhierarchical but it is still a mesh account because it assumes that responsibility requires a suitable connection between the agent's behavior and her values. Reviving certain Platonic themes, Watson claims that our desires or preferences have different sources. Our "vul­

Gual “motivational preferences,” by contrast, have their source in "appetite." Mickey, for example, may most want to eat the pie in the sense of "most want" that amounts to having the causally strongest desire for something yet, given his medical condition, reason does not sanction that desire but the opposed one to
refrain. Watson suggests that we should identify free action not with an agent's causally strongest desire, but with her value judgment about what she ought to do. His view can be summarized in this way:

Watson1 Agent, S, does A freely at time, t, if S most values doing A at t, S does A at t, and S's doing A at t issues in a nondeviant fashion from what S most values then.

Watson says that free agents have the capacity to translate their values into action; their actions flow from their "evaluational system." This (and other claims) suggest that Watson also endorses the following.

Watson2 If an agent acts against what she most values at a time, then she acts unfreely at that time.

One problem with Watson's views lies with Watson2. There appear to be cases in which though a person acts against what she most values at a time, she then acts freely. Mickey's acratically eating the pie seems to be a case in point. Kadri Vihvelin (1994) gives another example: Person X lies to Customs officials. When questioned later, X agrees that breaking the law is wrong but candidly explains that he did it because he did not want to spend hundreds of dollars on import duties. X acted contrary to his judgment about what he ought to do; so he acted against what he most valued. But there is little reason to suppose that he acted unfreely.

Another worry concerns Watson2. One's values (or at least components of them), according to Watson, appear to be a subset of one's beliefs; they seem to be beliefs about what is right or wrong, or what one ought to do, or what is good and evil. If desires can be implanted in one against one's will or unbeknownst to one, so can beliefs, and hence so can values on Watson's conception of values. If an agent's behavior results from values surreptitiously implanted in her, then it is not clear, contrary to Watson, that she is responsible for that behavior. Whether Watson's theory has the resources to cope with this objection, I suspect, will turn largely on refinement of the pertinent view of values.11

Susan Wolf's theory (1990) is another mesh theory that implies that responsibility is a function of an appropriate fit between an agent's behavior and values. A striking feature of her view (which she calls the Reason View) is that responsibility requires that agents have the normative ability to appreciate the "True and the Good" and to do the right thing for the right reasons:

According to the Reason View, . . . responsibility depends on the ability to act in accordance with the True and the Good. If one is psychologically determined to do the right thing for the right reasons, this is compatible with having the requisite ability. . . . But if one is psychologically determined to do the wrong thing, for whatever reason, this seems to constitute a denial of that ability. For if one has to do the wrong thing, then one cannot do the right thing, and so one lacks the ability to act in accordance with the True and the Good. (ibid.: 79)

As Wolf herself indicates, her Reason View is committed to the curious asymmetry thesis that whereas one can be responsible for a right (or good) action that one could not have avoided performing, one cannot be morally responsible for a wrong (or bad) action that one could not have avoided performing.

Wolf motivates her asymmetry thesis largely by contrasting examples of seemingly right and wrong (or good and bad) actions that are unavoidable. With respect to one part of her thesis, in one of the examples, a woman on an uncrowded beach sees a young boy struggling in the water in obvious need of aid. Wolf assumes that it is literally impossible for the woman to refrain from saving the child "because her understanding of the situation is so good and her moral commitment so strong" (ibid.: 82). For a woman with her moral character, leaving the child to drown is either "unthinkable" or simply not a thought that can be taken seriously (ibid.: 59). Wolf concludes that even though the woman is not free to do otherwise, she is praiseworthy for saving the child. Wolf finds support for the other part of her asymmetry thesis in a class of cases involving kleptomania, drug addiction, hypnosis, and deprived childhoods, in which agents are apparently not blameworthy for performing seemingly wrong (or bad) actions even though they could not have done otherwise. Wolf claims that the judgment that they are not blameworthy stems from the fact that the agents could not do the right thing for the right reasons.12 Such examples lead Wolf to conclude that the freedom requirement for moral responsibility is asymmetrical.

One shortcoming of the Reason View turns on Wolf's insistence that "it is only the ability to do the right thing for the right reasons . . . that is required for responsibility" (ibid.: 81). It appears that a person could deliberately harden his heart to the supplications of morality and so not be able to act in accordance with the True and the Good and still be fully responsible for his actions. As an illustration, suppose Glaucon, who is well aware of moral restrictions, has decided to develop his character in such a way that his true guiding principle is one of self-interest. Imagine, now, that having "freely" and willfully developed his character in this direction, Glaucon is literally unable to refrain from pocketing a gold coin that he spies on the otherwise deserted stretch of road. When he pockets the coin, Glaucon acts in conformity with his character. His options are constrained by his earlier deliberate efforts to thwart requirements of the True and the Good in favor of maximizing self-interest. But although Glaucon cannot now act in accord with the True and the Good, it is clearly possible that he is blameworthy for pocketing the coin. His manner of conduct is relevantly analogous to what it would be
he to consent to being hypnotized for the purpose of ensuring that he pockets the coin. In the latter case, there would be little doubt about his culpability, even though he could not thwart the (let us suppose) powerful posthypnotic suggestion. Wolf might respond to this case by saying that Glaucus is indirectly responsible for pocketing the coin because he was indirectly capable of acting according to reason. She might say, Glaucus freely decided, when he did have the capacity to act in accordance with the True and the Good, to shape his character in a certain way; consequently, he must be responsible for at least some actions that "issue from" his acquired character.

This reply, however, is not fully satisfactory. Wolf’s view does support the reasonable verdict that Glaucus is responsible for acquiring his "self-interested character." But it fails to support the further plausible verdict that he is also responsible for pocketing the coin, a deed of his that issues from this character, if "it is only the ability to do the right thing for the right reasons..." that is required for moral responsibility.

A second problem with Wolf’s view is that it, too, just like HT theories, seems to succumb to the problem of responsibility-undermining manipulation, on the reasonable supposition that an agent’s ability (or inability) to do the right thing for the right reasons could be, for instance, electronically induced.

A third difficulty lies with Wolf’s asymmetry thesis. Apparently Wolf runs together two different sets of concepts, one having to do with deontic rightness and wrongness and the other with deontic value. We can say that an act is overall bad if and only if, were it performed, it would produce more intrinsic evil than intrinsic goodness. Similarly, an act is overall good just in case its performance would produce more intrinsic goodness than intrinsic evil. It seems perfectly possible for an agent to be in a situation in which all her options are overall bad. Still, if we allow for even modest consequentialist considerations, it is not unreasonable to assume that of those options, the one that is least overall evil is the one the agent ought morally to do. Hence, what is overall bad need not be wrong and what is overall good need not be right or obligatory. So even if—and this is contentious—there is an asymmetry in alternative possibility requirements for moral responsibility for overall good and overall bad actions, it does not follow from this asymmetry that there is an analogous asymmetry for right and wrong actions.

Still, Wolf might reasonably claim that in the range of cases that interest her—the case of the upright mother, the agent with a depraved childhood, and so on—it does seem plausible to suppose that the overall good or bad actions performed are also, respectively, right or wrong. But even then, as John Fischer and Mark Ravizza (1992a) have argued, a deep worry plagues her asymmetry thesis. Fischer and Ravizza generate a Frankfurt-type case ("Villain") in which a vile character, Joe, is blameworthy for performing a bad action although he could not have done otherwise. Joe decides (for his own perverse reasons) to push a child off a pier for the purpose of causing her to drown in the violent surf. Had Joe shown any sign of not acting on his decision to kill the child, a device in his brain would have caused him to acquire the decision and to act on it. But Joe acts on his own, independently of any interference from the device. In this sort of case, it seems sensible to suppose that Joe is deserving of blame, and hence is morally blameworthy, for performing an overall bad action even though Joe could not have done otherwise. Fischer and Ravizza conclude that Wolf’s asymmetry thesis is false; good and bad actions are symmetric with regard to the lack of a requirement of alternative possibilities for moral responsibility. In turn, if one (mistakenly) thinks that overall bad actions are wrong and overall good actions right, Villain should impel Wolf to reconsider her asymmetry thesis.

Let us next, examine one of the most recent incarnations of a mesh theory.

5. Bok’s Account

Toward developing her version of a mesh account of free will, Hilary Bok distinguishes two standpoints, theoretical and practical, from which human actions can be regarded. The theoretical standpoint involves scientific (including psychological and historical) description and explanation of the phenomena in the universe. This standpoint, which is the determinist’s, pictures the world as a place in which everything, human behavior included, is governed by natural laws. Now an agent has ultimate control over, say, his making some decision, when, roughly, there are no conditions "external" to him that are minimally causally sufficient for his making that decision. From the theoretical perspective, it appears that the libertarian’s desideratum of ultimate control cannot be satisfied. The practical standpoint is the one we use to reason about what we should do and how we should live (1998: 62-63). Aligning herself with Kant, Bok contends that our ascriptions of freedom and responsibility are based on the requirements of the practical standpoint and that when we occupy that standpoint, we have every reason to regard ourselves as free and to hold ourselves morally responsible for our actions (Ibid: 52).

Next, Bok argues that the claims of theoretical and practical reasoning do not conflict (Ibid:ch.2), and that the freedom and moral responsibility presupposed by the practical standpoint is ultimately reconcilable with all our behavior’s being completely determined by the natural laws. Central to establishing the consistency of freedom of the will and determinism is Bok’s distinction between theoretical possibility (or possibility tout court) and practical (or narrow) possibility. Her view
is that when we reason practically, all we assume is that our various alternatives are possible in the sense that we would perform them if we chose; we assume, in other words, a conditional or hypothetical conception of freedom to do otherwise. Given certain facts about deliberation (ibid.: 108), we do not assume in the practical standpoint that available alternatives are possible tout court, that is, we do not assume that, given the natural laws and the past, a person could have done other than what he or she in fact did.

Not surprisingly, on Bok's view the notion of practical possibility is relevant to free will. She writes:

From the practical point of view... our use of this general [conditional] conception of possibility, as opposed to the narrower possibility tout court, is both unavoidable and rational. It is unavoidable because while we deliberate we cannot possibly employ a conception of the alternatives that are available to us that is narrower than the set of actions that we would perform were we to choose to do so. It is rational because, for the purposes of deliberation, we must regard the question what we will choose to do as open and because, if we regard that question as open, we should not regard the various actions that we would perform if we chose as differing with respect to their possibility, since any of them would be possible tout court if we chose to perform it. Moreover, to determine whether or not a particular action is one which we would perform if we chose is to determine whether or not we can regard it as a possible object of choice: an action about which question whether or not we have reason to perform it can legitimately be raised. (Ibid.: 117–18)

Armed with the conception of practical possibility, Bok introduces her conception of freedom—the sort that is required for normative appraisals:

A person is free if she is capable of determining her actions through practical reasoning; such an agent is free to choose among all those acts that she would perform if she chose to perform them, and she is free to perform a given action if she would perform it if she chose to do so. (Ibid.: 120)

Bok remarks that freedom of the will has traditionally been claimed to involve two conditions: first, our wills are free only if we can choose among genuine alternatives (ibid.: 118); and second, freedom of the will involves stepping back and asking ourselves whether or not we should act on our various alternatives and desires (ibid.: 119). Both these conditions, she says, are met by her view of freedom: the first because when we engage in practical reasoning, constraints on knowledge dictate that we should regard all those actions that we would perform as genuine alternatives; and the second because determining our conduct through practical reasoning requires that we evaluate our motivations and decide which we have most reason to activate. In that her account satisfies this second feature, Bok claims, it is a "hierarchical" or, more aptly, a mesh account. She indicates that her account differs from those of, for example, Watson or Frankfurt in interpreting freedom as consisting in our ability to determine our conduct through practical reasoning, and not in our ability to act on our values or second-order volitions (ibid.: 119).

Bok then relates her conditional account of freedom to moral responsibility by advancing a particular conception of responsibility. Her view is that we are responsible for those actions that reveal the quality of our character and our will (ibid.: 139, 140, 152, 180). When my conduct reflects my will, I can appropriately hold myself accountable for it; it can legitimately be attributed to my charge (ibid.: 152). I can ask what such conduct reveals about me and my character and about ways in which it might be improved. As for blameworthiness, when, through exercise of my will—that is, through determining my actions by engaging in practical reasoning—I freely and knowingly violate my standards by performing some action, I am blameworthy for that action (ibid.: 167, 168, 192).

Let us begin evaluation of Bok's theory by focusing on ultimate control. Addressing the libertarian's worry that an action or choice that is ultimately caused by events outside an agent's control cannot be one for which the agent can be responsible, Bok writes:

[O]ur main interest is in evaluating our contributions to the process whereby we came to act as we did. We want to discover whether we did anything that we think wrong: whether we were attentive to all those considerations we think we should have taken into account, whether we thought hard enough and seriously enough about what we were about to do, whether we made the right choices, and, if we did not, what explains our failure. It may be that it was determined that not only our act but the process whereby we came to perform it would fail to conform to our standards; but that fact is not relevant to the question whether or not it did fail to conform to them. And it is that question that we must answer if we wish to decide whether or not the fact that we did something that we regret indicates the existence of some fault in our wills that we can and should try to correct. (Ibid.: 156)

Now it is true that the contributions that we make to the process whereby we come to act as we do are relevant to responsibility. But surely so are contributions of the past. In the case of Robert Harris discussed by Watson, when we learn, for example, about the parental abuse he suffered, and the neglect and the soiled conditions under which he spent his formative years, it is difficult to hold him fully responsible for his later vicious criminal behavior. If such past conditions mitigate responsibility—Harris would not be responsible to the same degree as he would be if his past were not so depraved—then contributions of the past, in addition to one's contributions to the process, which fuel one's actions, do bear on responsibility. But if we admit that certain conditions of the past over which one had no control can duly influence responsibility, then the compatibilist must squarely face the challenge of explaining why other conditions of the past over which one could not have had control—for example conditions that existed long
before one was born—are not relevant to responsibility. And I believe this libertarian challenge has not been adequately addressed by Bok.

It might, in connection with this problem of past influences on responsibility, be noted that Wallace’s normative conception of responsibility has a definite advantage over Bok’s. Wallace, as we have seen, proposes that an agent is responsible for something insofar as it is fair to hold him responsible for that thing. In a case like Harris’s, presumably, the normative approach would explain mitigation of responsibility by invoking considerations of fairness: given Harris’s past, it is not fair to hold him responsible to the degree to which he would be held were it not for his unfortunate past.

Other compatibilist approaches relevant to past influences involve developing a conception of “authentic” springs of action, roughly, springs of action that are “truly one’s own” (Fischer and Ravizza 1998: ch. 8; Haji 1998, chs. 6, 7). The strategy here is that with some such conception in hand, we could infer that, in a case like Harris’s, some of Harris’s springs of action, perhaps the values that he applies to evaluative best judgments that, in turn, give rise to decisions, are not truly, or are not fully truly, his own. Yet other approaches involve careful attention to the actual causal pathways that culminate in action with an eye toward isolating responsibility-undermining factors that might be present in some but not all of the pathways.

Another concern with Bok’s mesh account is that it is susceptible to the problem of clandestine manipulation that also plagues othermesh accounts like Wolffe. If responsibility is a function of an appropriate fit between the choices an agent makes and her practical reasoning, and the very “inputs” of such reasoning can be surreptitiously controlled by an external manipulator, it appears that Bok’s mesh condition can be satisfied even though the agent is not intuitively responsible.

Finally, there are concerns with Bok’s views that are associated with “deontic morality.” Moral responsibility requires freedom or control. Similarly, other moral appraisals, like deontic ones of rightness, wrongness, and obligatoriness, presuppose freedom. This is perhaps most evident in the case of moral obligation. The verdict that the paraplegic ought to have walked across the lawn (or that he did wrong by failing to walk across the lawn), when ought denotes moral obligation or requirement, appears to be conceptually inconsistent. There is a control principle with intuitive and theoretical appeal associated with moral obligation, to wit, principle K, which is, recall, the ought implies can principle.

I want to evaluate Bok’s suggestion that, from the standpoint of practical reasoning (ibid.: 62–65), the relevant conception of freedom presupposed by various moral appraisals (like those of moral responsibility or deontic ones) states that we are free only when we choose among courses of action that are truly open to us: in short, if we freely performed some action, we had genuine alternative possibilities. For Bok, the pertinent notion of possibility is the practical notion: when we engage in practical reasoning, we assume that various alternative acts are possible in the sense that we would perform them if we chose (ibid.: 104–9). For purposes of deciding what one should do, then, it seems that Bok favors a conditional analysis of can:

CC  S can do A =df. S would do A if S chose to.

There are, however, problems with such an analysis. J. L. Austin (1961) described cases in which, though it is true that some person could do something, it is false that she would do that thing if she chose to. In one of his best known examples, Austin imagined himself a golfer standing over a three-foot putt. Suppose he had made such puts in several but not in all relevantly similar situations in the past. Then it seems perfectly consistent to say, Austin proposed, that he could make the putt on this occasion, although he might miss (on this occasion). So his ability to make the putt does not entail that he would make the putt whenever he chose to do so. Keith Lehrer (1968) offers a different sort of counterexample against CC. There can be cases, Lehrer claims, in which although a person cannot do something, it is true that she would do it if she chose to. Suppose, for instance, that as a result of the traumatic experience of having been bitten by a snake in her youth, Leila has developed an extreme psychological aversion to snakes. Her aversion renders her unable to choose to pick up the harmless snake in her biology class. And (pardonly) because she cannot choose to pick up the snake, she cannot pick up the snake. Yet, Lehrer argues, it might be true that if she were to choose to pick up the snake, she would do so. So whereas Leila cannot pick up the snake because of her aversion, the conditional analysis yields the result that she can so act.

Deontic appraisals pose another challenge to Bok’s conditional notion of freedom. To explain this challenge, we start by noting that there is a requirement of alternative possibilities for deontic acts. An outline of the argument for this requirement follows: We have seen that K (ought implies can) and principle OW (OW says that an agent S ought to do [not to do] A and only if it is morally wrong for S not to do [to do] A) entail that overall wrong actions require alternative possibilities. Setting aside one complication, it is straightforward to show that under conditions in which an agent lacks alternative possibilities, the agent’s actions, besides not being overall wrong, are neither overall obligatory nor overall right. For if some action, A, is obligatory for some person, then failing to do A—an omission—is wrong for that person. But it is false that any action (or omission) in a world in which a person lacks alternative possibilities is wrong for any person, given WAP (it is morally wrong for S to do [not to do] A only if S can refrain from doing [do] A). So it is false that failing to do A is wrong for that person. Hence, in such a world, no action would be morally wrong or obligatory for that person.
There is no analogous way to show that rightness also requires alternative possibilities, for there is no analogous principle like OW that will allow us to infer that right implies “can refrain.” But it is highly plausible to maintain that right does imply “can refrain.” Otherwise, insofar as obligatoryness and wrongness do require alternative possibilities, we would have to contend with the unpalatable view that it is morally right for one to do whatever heinous acts one cannot avoid doing.

An alternative route to the conclusion that the primary deontic properties of rightness, wrongness, and obligatoriness require alternative possibilities assumes, very reasonably, that, just like ought, right and wrong also imply can. Then OW and the assumption that wrong implies can entail that there is a requirement of alternative possibilities for obligatory acts; OW and K entail that there is such a requirement for wrong acts; and we add to this the result that right, too, requires alternative possibilities.

Now consider a standard “Frankfurt-type” case in which though an agent intentionally performed some action on “his own,” he could not have refrained from performing it given the particular circumstances in which he found himself. In such cases, a counterfactual intervener or controller keeps vigil over the agent, ensuring that if the agent, say, Jones, shows any sign whatsoever of not performing the relevant act (for example, killing White), the controller will intervene and cause Jones to choose to act in the desired fashion. In his situation (assuming that he is not incapacitated to choose to kill White when he so chooses), Jones is unable to refrain from killing White. We have established that no one is able to prevent a successful exercise of the relevant ability or abilities. This fact and the pertinent ones of Jones’s case entail that Jones’s killing White lacks any of these moral statuses.

Bok’s conditional notion of freedom yields the right results in this case. If one is able to do something, then it must be true that one has the opportunity to do that thing. In standard Frankfurt-type cases, although Jones has the ability to refrain from killing White, the counterfactual controller ensures that he lacks the opportunity to refrain from killing White. As it is false that Jones would have refrained from killing White had he so chosen, Bok’s conditional notion of freedom correctly implies that Jones could not have refrained from killing White.

Suppose, now, that determinism is true. Then there is exactly one pathway for each person into the future. Just as the counterfactual intervener in standard Frankfurt-type cases ensures that Jones could not have refrained from doing what he in fact did, so, if determinism is true, the natural laws and the past ensure that each agent cannot do other than what she in fact does.

It is important here to understand the pertinent parallel between Frankfurt-type cases and a deterministic world. An agent is able to do something only if she has both the ability and the opportunity to do that thing. There are relatively weaker and stronger senses of ability. A relatively weak sense of ability entails merely that the agent has the germane capacity, whereas a relatively strong sense of ability entails, in addition, that the agent have pertinent skills and perhaps know-how. So, for instance, even if Fritz, who is unfamiliar with computers, succeeded in turning one on, his turning it on would be something of a fluke. As he did turn it on, he did have the weak ability—the capacity—to turn it on, but we would not want to say that he had the pertinent skills or know-how, he lacked the strong ability to turn it on.

As for opportunity, by and large advocates of the principle that ought implies can recommend that opportunity is to be understood broadly to include all circumstantial factors that would enable someone to exercise a pertinent ability. On this view, someone lacks the opportunity to perform an action just in case there is something (a bolted door or a state of unconsciousness, for example) that would prevent a successful exercise of the relevant ability or abilities.

In a standard Frankfurt-type case, the counterfactual intervener does not determinally affect Jones’s ability to refrain from killing White; rather, the intervener ensures that Jones lacks the opportunity to refrain from doing other than what he does. Similarly, I agree with Kadri Vihvelin’s (n.d.) insight that, once we realize that to have an ability is just to have a certain kind of capacity or skill, it should be uncontroversial that the possession of abilities, including unexercised ones, is compatible with deterministic as well as indeterministic laws. For any ability we might think relevant to the question of freedom and moral responsibility—such as mental abilities like reasoning and deliberating concerning possible actions, making evaluative judgments, forming intentions, and so forth—there is no reason to think that deterministic causal laws (or the past) would deprive us of this ability. Rather, determinism deprives agents of the opportunity to do anything other than what they in fact do. But if this is so, then, as advocates of K insist, because an agent can do something only if he has the opportunity to do that thing, determinism ensures that an agent cannot do other than what she in fact does.

One might worry that unlike Frankfurt-type cases, determinism has actual prior effects on the agent’s capacities, abilities, character, and motives, whereas a merely counterfactual (but nonactual) intervener in Frankfurt-type cases does not have any actual effects on these things. So determinism, one might claim, does affect both one’s capacities or abilities and one’s opportunities.

In response, deep concerns still remain about how deterministic causal laws can undermine relevant abilities. Think of the matter in this way. When giving a positive account of free will, sophisticated indeterminists can adopt many of the pertinent views of compatibilists, including for example, views about the causal connections among one’s values, best judgments, and intentions or decisions. Indeterminists of one brand (for example, Kane 1996a) typically postulate indeterminacy at some point or points in the pathway of intentional action. They theorize that some of the laws pertinent to the causal explanation of our choices or action
will be nondeterministic or probabilistic ones. But surely nondeterministic causal laws do not detrimentally affect our having abilities, including unexercised ones, required for free action. If they do not, it is a puzzle why it should be thought that deterministic causal laws would detrimentally affect the appropriate abilities.

In any event, granting that determinism negatively affects both one’s abilities and opportunities to do otherwise simply strengthens the case for the view that determinism expunges alternative possibilities.

One moral we can now draw is that if Jones’s act of killing White is not wrong (or right, or obligatory) in a Frankfurt-type situation, and it is not wrong because he lacks the opportunity to do other than kill White, then Jones’s act of killing White should also not be wrong (or right, or obligatory) in a deterministic world, as in such a world he would, again, lack the opportunity to refrain from killing White. Indeed, for any act performed (or intentionally not performed) by any agent in such a world, the agent would lack the opportunity to refrain from performing (or performing, in the case of an intentional omission) that action. So in such a world, no act could instantiate a primary deontic property.

Bok might object that one feature of Frankfurt-type examples that is absent from deterministic worlds and bears significantly on whether acts in such worlds can instantiate primary deontic properties: in Frankfurt-type examples, it is not the case that the agent like Jones would have done otherwise had he so chosen. But in a deterministic world this is not so; rather for example, Jones would have refrained from killing White had he so chosen. So the conditional notion of freedom, Bok might argue, yields this result: whereas in Frankfurt-type examples, Jones could not have refrained from killing White, Jones could have so refrained in a deterministic world. And, further, she might propose, merely the conditional notion of freedom to do otherwise is presupposed by deontic acts. Hence, she might rejoin, agents in deterministic worlds can perform acts that instantiate primary deontic properties.

But I believe that this objection is suspect. First, as we have seen, there are serious problems with the conditional analysis of can. (See, also, Berofsky’s essay in this volume, ch. 8) More important, Bok’s verdict that acts can be right or wrong or obligatory in deterministic worlds is predicated on her conditional notion of freedom. If this verdict is correct, then the view that an agent can perform an act only if she has the opportunity to perform it must be rejected by Bok, if opportunity is to be understood in the broad sense explained previously. We saw that in this broad sense of opportunity, an agent lacks the opportunity to perform an action if something prevents or would prevent a successful exercise of the relevant ability or abilities.

In a deterministic world, the natural laws and the past prevent or would prevent a successful exercise of each person’s ability to do other than what he or she in fact does. An agent can lack the opportunity to do other than what he in fact does even if it is true that he would have done other than what he in fact did, had he been chosen to do so. So, for instance, if S does A in situation U, S can lack the opportunity to do other than A in U even if it is true that S would have done other than A in U if S had chosen to.

Suppose Jones is in a deterministic world and he kills White. Then determinism ensures that he lacks the opportunity to refrain from killing White. But if he lacks this opportunity, he cannot, in his situation, refrain from killing White. As it must be true both that Jones can kill White and that he can refrain from killing White if his killing of White is to have a primary deontic property, his killing of White cannot have such a property if determinism is true. But if Bok believes otherwise, if she believes that Jones’s killing White can have such a property in a deterministic world and she accepts WAP, then it seems that she must reject the view that one can perform an act only if one has the opportunity to do so. However, this view ought not to be rejected; it is not false. Hence, it follows that Bok’s verdict is incorrect. Indeed, it follows that Bok’s conditional concept of freedom is not the concept that is presupposed by deontic acts.

I believe that the discussion of Bok’s conditional notion of freedom and deontic acts yet again casts doubt on her proposal that the practical and the theoretical standpoints can be isolated from one another when we think about freedom and normative appraisals.

As we have seen, Bok argues that when we engage in practical reasoning, we must use the conditional concept of freedom; and that when we act freely in this sense, we choose among genuine alternatives, and we should hold ourselves responsible for the acts we freely perform (Bok 1998: 205). As we can act freely in the conditional sense of free even if determinism is true, determinism, she believes, is no threat to responsibility. But this sort of “insulation argument” won’t go through with deontic acts. Even granting that practical reasoning requires that we use the conditional notion of freedom, it seems that this notion of freedom is not presupposed by deontic acts as reflection on Frankfurt-type examples suggest. So even if we can act freely in the conditional sense of free in a deterministic world, acting freely in this sense will not sustain the conclusion that agents in such a world can perform acts that instantiate one or more of the primary deontic properties.

NOTES

Many thanks to Robert Kane for his comments and suggestions.

1. Arguments for the incompatibility of determinism and freedom to do otherwise have been advanced, for example, by Fischer (1983a, 1994), Ginet (1983, 1990), van Inwagen (1983, 1989), Warfield (1996), and Wiggins (1973). Critical discussion of these sorts of arguments can be found in Lewis (1986), Slote (1983), and Vihervuori (1990, 1991). See the
essays in this volume by van Inwagen (ch. 7) and Kapitan (ch. 6) for further discussion of these arguments.

2. Saul Smilansky (2000) defends the following interesting dualistic position in the free will debate. Smilansky claims that what he dubs the "Core Conception" captures the "basic intuition" (ibid.: 2) underlying the free will problem. This is the intuition that, with respect to things like moral responsibility, desert, justice, and punishment, the question of control or lack of it is morally crucial (ibid.: 2). This deep intuition, in turn, Smilansky proposes, has its roots in the obligation to respect persons (ibid.: 14–22).

Smilansky argues that compatibilist forms of control are inadequate; if we have only such forms of control, "we are in serious difficulty in ethical and personal terms" (ibid.: 3). Similarly, he reasons that libertarian free will that demands "ultimate control" cannot be sustained. However, Smilansky's dualistic position, while not denying the hard determinist's lessons (the hard determinist rejects the possibility of responsibility in a determined world), especially the lesson that no one is "ultimately responsible" for one's choices and actions if determinism is true, also maintains that in order to be, for example, just, we have to be partial compatibilists. Compatabilism must be "taken into account" (ibid.: 93), he claims, "not only because of the pragmatic need for certain social arrangements [in which, for instance, considerations of desert are central], but also in a fundamental non-consequentialist moral sense, which stems from the Core Conception" (ibid.).


5. Smilansky (2000: ch. 9) distinguishes two versions of Strawsonian (or "reactive") naturalism. Revisionist naturalism "seeks to change the perception that there is a theoretical need to justify common attitudes and practices, holding that there is no need for general grounding and that the reactions themselves provide all the (self-grounding required)" (ibid.: 222). Weaker nonrevisionist naturalism holds that "in practice common attitudes and behaviours remain constant, whatever the theoretical case may be" (ibid.: 224). On this view, our reactive attitudes, for example, cannot be threatened by the realization that (if Smilansky is right) there is no libertarian free will. Smilansky argues against both these versions of naturalism.

6. Watson (1987b: 259–60) usefully distinguishes between excusing and exempting conditions. Excusing conditions (like coercion or ignorance) inhibit responsibility locally; they don't imply that the agent is not a fit subject of responsibility. Exemptions (like insanity), in contrast, block responsibility globally; they do imply that the agent is not an appropriate candidate for responsibility ascriptions.

7. This sort of distinction between interstitial and externalist theories is drawn by Fischer and Ravizza (1998: 252).


David Zimmerman (2000) has recently developed a different sort of criticism of Frankfurt's hierarchical view. The criticism builds on the idea that a person who, for example, has fallen victim to manipulation at the hand of another party, can "take re-

sponsibility" for externally induced desires simply by bringing herself to identify with them wholeheartedly and decisively, despite their causal origins. Zimmerman argues that this view of taking responsibility commits Frankfurt to the morally repellent Stoic thought that "resignation to necessity is a path to liberation" (ibid.: 38). In addition to developing this objection, Zimmerman also proposes modifications to Frankfurt's hierarchical view that Zimmerman believes meet this objection.

10. Building on Frankfurt's 1976 essay, "Identification and Wholeheartedness," Michael Bratman (1996) provides an account of identification that, among other things, tries to capture the sense in which one's motivations are fully one's own. Briefly, Bratman's approach develops Frankfurt's relatively early idea that the concept of decision is central to identification. The primary components of Bratman's suggestive account include the following. One treats one's desire as reason-giving when one treats it as setting an end that can, to some extent, justify performance of relevant means and/or relevant preliminary steps (ibid.: 9). For example, you treat your desire to pursue money as reason-giving if you treat that desire as setting, say, the end of amassing economic power, and this end justifies performance of relevant steps like careful investment. One is satisfied with a decision to treat some desire as reason-giving when that decision does not conflict with other standing decisions and policies about which desires to treat as reason-giving (ibid.: 11). An unwilling or grudging addict might treat her desire for the drug (to which she is addicted) as reason-giving, but this desire might conflict with a general policy of hers against treating her desire for the drug as reason-giving. In this case, the grudging addict is not satisfied with her decision to treat her desire for the drug as reason-giving. To identify with a desire, (1) one decides to treat that desire as reason-giving in some of one's practical reasoning and planning (ibid.: 8–9); (2) one is satisfied with that decision (ibid.: 10–11); and (3) one either treats that desire as reason-giving or is fully prepared to treat it as reason-giving were a relevant occasion to arise (ibid.: 11–2).

11. Watson could be interpreted as offering the following account of identification. Roughly, an agent, A, identifies with his desire to A, and so, if that desire moves S to A, with S's A-ing, if the desire to A is part of S's evaluational system and this desire causes S to A. But Watson renounces this view in "Free Action and Free Will" (1987). He thinks that there can be cases in which one's act is fully one's own but the act is not caused by components of one's evaluational system. For example, it seems possible to identify with an action one does not think to be best or care most about. Watson says: I might fully "embrace" a course of action I do not judge best; it may not be thought best, but it is fun, or thrilling; one loves doing it... Call such cases, if you like, perverse cases... There is no estrangement here. One's will is fully behind what one does. Of course, a person's evaluational system might be defined just in terms of what that person does, without regret, when it comes right down to it, but that would be to give up on the explanation of identification by evaluation. (1987: 196).

12. The cases involving drug addiction and hypnosis are discussed by Wolf in Wolf (1987).

13. It is notable that Bratman's account, as presented above, is about purely internal identification. However, even if Bratman's account is about purely internal identification, Bratman's account can still be applied to external identification. For, in Bratman's account, the concept of decision is central to identification. And since the concept of decision is central to identification, the concept of decision is central to external identification as well. Therefore, even if Bratman's account is about purely internal identification, Bratman's account can still be applied to external identification. For, in Bratman's account, the concept of decision is central to identification. And since the concept of decision is central to identification, the concept of decision is central to external identification as well. Therefore, even if Bratman's account is about purely internal identification, Bratman's account can still be applied to external identification.
locutor makes sense only if people treat themselves as having belief-forming and desire-forming capacities that satisfy three conditions: there are norms governing what they should believe and desire; they are capable of recognizing the demands of these norms; and by and large they have the capacity to respond appropriately to the demands they recognize (ibid.: 433–40). Pettit and Smith then propose that to regard an interlocutor as capable both of recognizing and of responding to these norms is to regard her as a subject who can be held responsible for what she believes and what she desires and does (ibid.: 440–44). They theorize, and here the salient resemblance to Wolf’s view surfaces, that the sort of freedom required for believing freely and desiring freely is the ability of the agent, in the event of getting things (beliefs or desires) wrong, to get them right (ibid.: 444–47); that is, an agent’s “beliefs and desires are free to the extent that they are the product of an ability, in the event of his being wrong, to get them right” (ibid.: 448n.21). On their view, the ability to believe or desire otherwise is “inherently attractive . . . only so far as it is the person’s ability for anything that is not rightly believed or desired always to have believed or desired otherwise” (ibid.: 444). Responsibility in belief and desire, then, is a function of a suitable fit between an agent’s beliefs and desires and certain yardsticks or demands (as dictated by the relevant norms) of right belief and right desire. Finally, Pettit and Smith propose that the natural position for them to take with respect to free choice is that “a subject’s choices are free to the extent that they are the product—product, no doubt, ‘in the right way’—of beliefs and desires that are themselves free” (ibid.: 448n.21). It would seem that this interesting approach regarding freedom in belief and desire (and choice) would succumb to one prominent sort of criticism that Fischer and Ravizza advance against Wolf’s Reason View: Frankfurt-type examples, it appears, could be constructed in which though an agent “freely” believes or desires wrongly on his own, he could not have believed or desired otherwise because of the counterfactual intervener.

14. Some terminology will be helpful. Use the label “primary deontic properties” to refer to the moral properties of rightness, wrongness, and obligatoriness, and call any act that instantiates one or more of these properties a “deontic” act. The set of deontic acts comprises “deontic morality.”

15. For further discussion of these problems, see the essay in this volume by Berofsky (ch. 8).

16. Briefly, the complication is this: one requires principles additional to OW and K to show that intentional emissions cannot be wrong when one lacks alternative possibilities. I discuss this issue in Hajj (1999a).

17. If there is a requirement of alternative possibilities for deontic acts, then we have further reason to believe that Wolf is mistaken in claiming that right and wrong actions are asymmetric with respect to the requirement of alternative possibilities.

18. Kane (1999d: 50) also questions whether Bok can effectively shield the theoretical standpoint from practical questions about responsibility and guilt.