Philosophy and Phenomenological Research
Vol. LXI, No. 2, September 2000

Replies

JOHN MARTIN FISCHER
University of California, Riverside

MARK RAVIZZA
Jesuit School of Theology, Berkeley

I. Reply to Mele

As the result of spending perhaps too much time in London, Phil has become a reflective and committed hard determinist. Nevertheless, he does not see himself or others as "out of control" in some important sense. Thus, "Phil now believes that moral responsibility is a very useful myth." He still believes that the reactive attitudes have some useful functions: they help to motivate socially productive behavior and to discourage socially unproductive behavior. Also, Phil continues to care about "pleasure, pain, well-being, and the like—both his own and others," and to act accordingly.

Mele points out that we contend that someone who is genuinely morally responsible must satisfy certain "subjective conditions": he must see himself as morally responsible in order to be morally responsible. Mele seems to suggest that although Phil does not meet these subjective conditions, he nevertheless applies the reactive attitudes, as well as being an appropriate candidate for them. Certainly, he is not an entirely passive bystander to the movements of his body, nor does he appear to opt entirely out of the moral community.

Mele has usefully focused on what is no doubt one of the most contentious parts of our overall view (its subjective nature), and perhaps the one for which it is most difficult to argue. We shall begin with the apparent suggestion that Phil continues to apply the reactive attitudes. Phil's situation is that he sincerely and after considerable reflection does not believe that the reactive attitudes are "metaphysically justified." And yet he—apparently—continues to apply these attitudes to himself and others, and Mele contends that he is an apt target for them. Now we can see how Phil might believe that the reactive attitudes are not "metaphysically justified," in some narrow sense, but nevertheless "practically justified" and thus justified "all-things-considered." What is somewhat puzzling, however, is that Phil should believe that all-things-considered the reactive attitudes are not justified, and yet never-

theless genuinely continue to have them. (By "metaphysically justified" we mean justified all-things-considered: p. 236.)

Take, for instance, the attitude of respect. How can Phil *genuinely* have respect—as opposed to merely mouthing certain words and behaving as if he had respect—when he sincerely and reflectively believes that this attitude is unjustified? We have no doubt that Phil can *act as if* he respected someone, or that Phil can have an attitude like respect in many ways, except lacking the full content of this attitude.[1] But how can Phil genuinely respect someone (another or himself), while also sincerely believing that this attitude is all-things-considered unjustified (especially in light of metaphysical considerations which appear to threaten the idea that individuals *deserve* certain responses)? Similarly, how can Phil really have the attitude of resentment, while also sincerely believing that it is, all-things-considered, unjustified? (Note that this is not a general worry about the possibility of weakness of the will, but a worry that issues from the specific nature of the reactive attitudes.) We can see how Phil could have many of the same sorts of emotional responses, and behave just as he might, if he genuinely had resentment; but, as with all the other reactive attitudes, resentment is something more than a set of affective responses and mere behavior (verbal and non-verbal). In a way that is perhaps difficult to articulate precisely, an individual who is really resentful or respectful, who really loves or hates, or who really punishes or rewards (rather than merely engaging in conditioning of even a sophisticated sort), must not believe that these attitudes are (on balance) not defensible. Otherwise such an agent is just going through the motions.

But perhaps Mele would not contend that Phil genuinely has the reactive attitudes. Perhaps his contention is that Phil's attitudes (however they are properly characterized) and behavior indicate that he does not deem himself entirely out of control, and whether or not he actually has the reactive attitudes toward others, he can be deemed an apt target for these attitudes (applied by others). It is difficult to present a decisive reply to this important challenge. We offer the following (admittedly incomplete) sketch of a response.

We believe that the attitudes involved in moral responsibility (on the Strawsonian model)—the reactive attitudes—are *reactions* to or *responses* to agents whose behavior has a certain quality. The reactive attitudes (as opposed to other, related attitudes) do not seem appropriate, unless this quality is present. This quality is subjective (having to do with the way the agent sees the behavior), and it is difficult to define. We would suggest that the quality in question is a kind of "self-engagement." When an addict sees that he has an irresistible urge to take a drug, and he goes ahead and acts to satisfy his desire

---

[1] For a careful and detailed development of such an approach, see Derk Pereboom, *Living Without Free Will* (Cambridge: Cambridge University Press, forthcoming).

for the drug because he knows that it is futile to resist, it is natural to say that the addict is in some important sense "passive" with respect to his behavior. It is Frankfurt's intuition, and ours, that this sort of addict is not morally responsible for taking the drug (given that he is not morally responsible for having acquired the addiction in the first place). Similarly, a sailor who knows that his rudder is broken (through no fault of his own) becomes "passive" with respect to the direction of his sailboat. In these cases the actions of the agents do not "implicate the self" in the way required for moral responsibility; the agents are not "active", and they do not engage in a distinctive kind of "self-expression."[2] We would suggest that these agents lack the kind of "self-engagement" required for moral responsibility.

Now we grant that Phil is not like the addict or the sailor in all respects, but there is a crucial way in which he is similar: we would claim that his behavior does not exhibit the relevant kind of self-engagement. He does not see his behavior as the exercise of freedom, and thus he does not think he is an apt target for the reactive attitudes. He is thus similar to the addict and sailor in lacking a certain kind of self-engagement in his behavior, understood subjectively; there are different "routes" to the failure of self-engagement. Insofar as the reactive attitudes are a response to behavior with this particular sort of quality, Phil is not an apt target for them, even if he is a perfectly appropriate target for a range of similar attitudes, sanctions, and incentives.

It is hard to know how to *argue* for the contention that the reactive attitudes are responses only to behavior that has this subjective characteristic. But it is an intuition shared by others. As Galen Strawson, with whom (according to Mele) Phil had those somewhat disorienting conversations while on sabbatical, puts it, "Why should lack of explicit awareness of *A* be supposed to have as a consequence lack of *A* itself, as lack of any sense or conception of freedom [responsibility] seems to have as a consequence lack of freedom [responsibility] itself? Well, that is the question. But it does seem to be so.... We are free (truly responsible), if we are, partly because we see ourselves and our action in a certain way—as free (or truly responsible)."[3]

Mele intriguingly asks whether "a hypothetical, isolated community of emotionless beings—hence, beings with no reactive attitudes—include morally responsible agents? Such beings do not take themselves to be fair targets of the reactive attitudes.... But might they, even so, have desires for the welfare of others, [and] beliefs about what they themselves morally ought to do...." Assuming that Dr. Spock (of "Star Trek") has no emotions, could he nevertheless be morally responsible?

---

2 For a discussion of the relevant notion of self-expression, and its connection to moral responsibility, see John Martin Fischer, "Responsibility and Self-Expression," *Journal of Ethics* 3:4 (1999), pp. 277–97.

3 Galen Strawson, *Freedom and Belief* (Oxford: Clarendon Press, 1986), pp. 250–51.

Mele's question points to the differences among various explications of the concept of moral responsibility. We begin the book by pointing out that there are various analyses of the concept of moral responsibility (pp. 1–8). Although we find the Strawsonian analysis (in terms of the reactive attitudes) natural and plausible, and we adopt it as a working hypothesis in the book, we are not certain that it is the correct analysis of our culture's shared concept of moral responsibility. One could adopt a "ledger" view of the concept of moral responsibility, according to which an individual is morally responsible insofar as he has a moral ledger (or less metaphorically, insofar as he is an apt target for judgments of right and wrong, permissibility and impermissibility, and so forth). In *Responsibility and Control*, we were more concerned with laying out the conditions of application of the concept of moral responsibility, rather than arguing for a particular specification of this concept. Mele's question about emotionless beings points to a way in which the different candidates for the specification of the concept of moral responsibility differ. But note that whichever candidate analysis one selects, there will be a suitable subjective condition. So, for example, on the ledger view, the subjective approach would require that an individual view herself as an apt target of certain moral judgments (rather than the reactive attitudes or those specific moral judgments presupposed by them).

Mele also raises a worry about our contention that weak reactivity to reasons (together with a more robust sort of receptivity to reasons) is sufficient for the sort of reasons-responsiveness required for moral responsibility. He points out that "even in extreme cases of phobia or addiction we can usually imagine *some* reason such that if a troubled agent who did *A* had possessed that reason for doing *B*, he would have done *B* for that reason rather than doing *A*." For example, "Fred's agoraphobia is so powerful that he has not ventured out of his house in ten years, despite his family's many attempts to persuade him to do so and the many incentives they have offered him." Fred recently failed to go to his daughter's wedding because of his crippling problem. (By the way, we can't help but think that Phil would have benefited from a bit of this agoraphobia!) But even so, "in some possible world with the same laws, there was a raging fire in Fred's house on his daughter's wedding day. Fred, it turns out, is even more afraid of raging fires than of leaving his house." So Fred leaves his house and "heroically" attends the wedding in this world, and all by the right sort of causal sequence. Mele points out that Fred seems to meet our conditions for reasons-responsiveness, and yet Mele worries that Fred is not morally responsible, insofar as he is not morally responsible for his agoraphobia in the first place, and his fear is "so debilitating that it would take something as frightening as a raging fire to move him to decide to leave his house or to leave it intentionally...."

Cases involving fears, phobias, and addictions are delicate and difficult. Here it is crucial to recall the distinction between moral responsibility and

(for example) moral blameworthiness (or praiseworthiness). (pp. 7–8) One can be morally responsible for a morally neutral act. Thus, being morally responsible does not entail being blameworthy or praiseworthy. Further, on the Strawsonian approach we adopt, to be morally responsible for a bit of behavior is to be an *apt target* for the reactive attitudes on the basis of this behavior; it does *not* follow that all-things-considered the reactive attitude in question ought to be applied in the context.

We are particularly concerned in *Responsibility and Control* to mark a distinction between moral responsibility and the lack of it. This still leaves much important work to be done; for example, it still leaves the development of the theory of praiseworthiness and blameworthiness, and the specification of the conditions in which it would be correct, all things considered, actually to apply the reactive attitudes. But we hope that it will have been useful to present an approach which at least helps to mark the distinction between those agents who are morally responsible and those who are not (i.e., those agents who are *apt targets* for the reactive attitudes and those who are not). It may even be the case that this approach can help to structure future developments of the theories of praiseworthiness and blameworthiness, and the application of the reactive attitudes.

Obviously, there are various different sorts of fears, phobias, and addictions. At the risk of considerable oversimplification, we would sketch our views here as follows. We shall assume that the individuals in question are not responsible for the phobias and addictions in the first place; this of course may not be accurate in a particular case, especially in the context of addiction. (If an individual is indeed responsible for the phobia or addiction, then we would address the case via the "tracing" component of our theory.) Now if the individual really is not responsible for the phobia or addiction, and it issues in genuinely irresistible urges, then intuitively the individual is not morally responsible for the relevant behavior, and our theory has precisely this result. On the other hand, if the individual is like Fred, then his urges are strong but not irresistible, and we are inclined to say that he is interestingly different from someone whose urges are genuinely irresistible (or, for example, someone whose brain is being significantly manipulated to ensure that he behave in a similar way). The way we mark this distinction is to say that Fred is morally responsible, whereas someone subject to significant manipulation, or someone with irresistible urges, is not morally responsible. Whatever language one uses for the distinction, there is some important distinction here to mark. Someone who is directly manipulated in significant ways lacks control and is like a marionette (controlled by someone else), and someone subject to genuinely irresistible urges similarly lacks control (although he is

not necessarily controlled by someone else). But Fred is importantly different from both sorts of individual.[4]

Fred, as opposed to the others, is morally responsible, on our approach. But it does *not* follow that he is (say) blameworthy for staying inside and not going to his daughter's wedding. This is a separate judgment, one we would not be inclined to make, given Mele's description. (So in a sense we are in agreement with Mele, even though we want to insist that there is an important difference between Fred and (say) someone who clearly lacks control because of systematic electronic stimulation of the brain.) Indeed, we would suggest that the fact that the sort of incentive needed to cause Fred's "heroic" efforts to attend his daughter's wedding is so drastic can help to structure a more nuanced development of a theory of blameworthiness. In general, careful attention to reasons-receptivity and reasons-reactivity profiles might usefully structure our understanding of praiseworthiness, blameworthiness, and the appropriate conditions of application of the reactive attitudes.

Finally, Mele helpfully points out that there are some omissions which should not be treated straightforwardly as being constituted by "bodily movements," where the bodily movement in question is an action. So, for example, Betty's omission to mow the lawn by 9:00 may be constituted by her body's keeping still during the relevant time (given that she is a deep sleeper!), but this is not plausibly thought of as an *action* of hers. We agree. Although we do not have space to elaborate here, we would be inclined to think of this "keeping still" as a *consequence* of some previous behavior of hers (such as setting the alarm for 9:30), where the relevant behavior can indeed be understood in terms of some action. A more refined account of moral responsibility for omissions would need to be adjusted accordingly.

II. Reply to Bratman

We wish to begin by thanking Bratman for noting an error in the statement of our position on page 238 of *Responsibility and Control.* As he points out, we say that "taking responsibility is a genuinely historical notion." The sentence in question should have read, "On our account, moral responsibility is a genuinely historical notion, and its structure is similar to other important historical, recursively defined notions...." We believe that moral responsibility is genuinely historical insofar as it requires a prior process of taking responsibility. We are grateful to Bratman for this correction.

Bratman points out that we give two sorts of examples which, we contend, help to motivate the claim that moral responsibility is a historical notion: tracing and manipulation examples. He says that it seems to him that

---

[4] We believe in the doctrine that "reactivity is all of a piece," which we discuss on pp. 73–74. Thus, we would say that Fred's mechanism *can* respond even to the actual-sequence reason. This helps to distinguish Fred from an individual subject to significant manipulation or who develops an irresistible urge in some other fashion.

"tracing examples do no work here. The issue is whether there are cases of moral responsibility in which the global history is irrelevant.... Tracing cases only show that there are cases in which conditions at the time of action would not on their own suffice for responsibility, and yet the agent is responsible because of the history of those conditions. Appeal to such cases cannot show that there are not as well cases in which conditions at the time of action *do* on their own suffice for responsibility."

We grant that consideration of tracing cases cannot decisively establish that moral responsibility is a historical notion (and thus that there are no cases in which conditions at the time of action—say, responsiveness characteristics—suffice for moral responsibility). The sorts of tracing examples we discussed, involving drinking too much alcohol and subsequently being out of control, are cases in which two agents equally fail to be in the requisite way responsive, and yet one is morally responsible (for the relevant thing) whereas the other is not. Further, it was our contention that the reason why reflective people would discriminate between the two agents is that the (relevant) histories are different; it seems that in such cases, ascriptions of moral responsibility do not supervene on snapshot properties (or responsiveness profiles), but depend importantly on how those snapshot properties (or responsiveness characteristics) were "put in place." Now it is possible that in cases of lack of responsiveness, history matters, but in cases of responsiveness, history does not matter. So certainly Bratman is correct that the tracing examples do not decisively establish that history matters to ascriptions of moral responsibility. But we are not inclined to say that "tracing examples do no work." They at least show that there are certain contexts in which moral responsibility ascriptions do not supervene on responsiveness profiles, but depend crucially on history. This can make it at least plausible that *in general* such ascriptions depend on history, especially in light of the fact that it may seem odd that history matters only in certain cases (where there is lack of responsiveness) but not in others (where there is responsiveness). On what basis could it be thought that there would be this sort of asymmetry in the relevance of history?

Turning to the manipulation cases, Bratman points out that we criticize "mere mesh theories" insofar as the selected mesh (between higher- and lower-order preferences, or between values and motives) may be induced via responsibility-undermining histories. He contends however that this point establishes only that a "mere time-slice mesh" is not sufficient for moral responsibility. Bratman asks us to "suppose that a Frankfurtian hierarchy unequivocally favors a course of action, and that this hierarchy is itself responsive, in a strong sense, to reasons for and against. Perhaps all this is not necessary for moral responsibility; but the crucial question is whether it is sufficient." Bratman goes on to imagine that without his knowledge or consent a neurosurgeon operates on his brain "in a way that newly results in

a certain meshing hierarchy of desires, and in the fact that the continued presence of this hierarchy is strongly responsive to reasons. There are, of course, moral criticisms to be made of the neurosurgeon. But it does not follow that she has not caused there to be a responsible agent."

Bratman further suggests that one's reaction to this sort of case may depend on one's view of the nature of reasons for action. He says, "Suppose one sees practical reasons as essentially grounded in the agent's contingent ends. This might encourage the suspicion that the fact that the hierarchy is strongly responsive to reasons is not enough to override the fact that the ends themselves have been manipulated. Suppose, in contrast, that one sees practical reasons as objective in a way that avoids essential relativity to the agent's contingent ends. This might help support the view that strong responsiveness to such reasons suffices for responsibility." Bratman concludes that one's views about various manipulation cases may well depend on one's view about the nature of reasons for action, and that our attempt to develop a general theory of moral responsibility which is neutral with respect to particular conceptions of reasons for action may have prevented us from coming to a proper analysis of the cases.

We think that manipulation cases are very complicated, and one's intuitions can be somewhat protean and unstable. Sometimes subtle differences in the ways the examples are presented or "framed" can elicit changes in one's intuitive judgments about the cases. We are however inclined to say that in Bratman's example of the surgical implantation of a strongly responsive Frankfurtian hierarchy (in which the individual is unaware of the surgery and thus obviously has not consented), the individual is *not* morally responsible, at least not at first. (This leaves it open that at some point in the future—perhaps even soon after the surgery— the agent "take responsibility" for the new surgically implanted mechanism and *become* morally responsible. This might involve coming to find out about the surgery and accepting the new strongly responsive hierarchy.)

It is hard to know exactly how to argue for the above intuition. Perhaps cases in which certain central values or "ends" have been changed elicit our intuition more effectively. So, for example, imagine that Mary has always been a very strong proponent of a woman's right to choose abortion. Unbeknownst to her, a surgeon implants a Frankfurtian hierarchy of preferences that is "pro-life." She now finds herself, somewhat mysteriously, strongly opposed to abortion. Further, imagine that the new structure of preferences is appropriately responsive to reasons; for example, holding fixed the relevant mechanism, if Mary were to come to recognize good reason to support abortion, she would do so. Of course she *actually* does *not* recognize such reasons. Because of the way her responsive hierarchy was implanted, our intuition is that Mary is not morally responsible. Our explanation is that she has not (yet) taken responsibility for the mechanism that issues in her behav-

ior; this is of course consistent with her doing so subsequently. On our view, one makes the relevant mechanism "one's own" by taking responsibility for it. Intuitively, though Mary's choice may well track reasons, it is not (in a sense that is perhaps difficult to specify) *hers*.

Bratman's claim that views about manipulation cases depend on particular theories of reasons for action is suggestive but puzzling. He seems to suggest that taking different approaches to specifying what reasons an agent has will lead to different conclusions about the moral responsibility of an agent such as Mary. But it does not appear to us that different views about (say) the objectivity of reasons should lead to different views about Mary's moral responsibility. So, for example, one might suppose that there are objective moral reasons, and one could demand responsiveness to these reasons for moral responsibility. Still, our basic intuition would emerge here as the view that someone could be "connected" to these reasons via responsibility-undermining histories and thus *not* be morally responsible. Mere connection to objective reasons is not enough for moral responsibility, if the connection is forged via (say) manipulation. Whether reasons are "objective" or not is one issue, but how one latches onto these reasons is another; it seems that intuitions about moral responsibility are driven at least in part by the latter consideration. It seems to us that our intuitive reaction to a case such as Mary's is based primarily on the intuitive demand that one come to be connected to reasons via the right sort of history—not on implicit assumptions about the nature of reasons.[5]

Consider, now, Judith, who has had a moderately reasons-responsive mechanism implanted in her (without her knowledge or consent). We don't see how this case is importantly different from the case of the implantation of a strongly reasons-responsive mechanism in Mary, with respect to the attribution of moral responsibility. Judith's unfortunate transformation from peaceful to violent parallels Mary's transformation from pro-choice to pro-life. In both cases our intuition is that the agents are not morally responsible because, in some sense difficult to specify, the choices are not the *agent's own*. Our theory seeks to capture this intuition by requiring that the choices (and behavior) issue from the relevant agent's own mechanism, and it holds that a mechanism becomes the agent's own only when the agent takes responsibility for it. Of course, just as with Mary, Judith may *become* morally responsible subsequently; on our approach, this would be in virtue of subsequently taking responsibility for the new (surgically implanted) mechanism.

---

[5] Bratman claims that a particular view about reasons "might encourage the suspicion that..." or "might help support the view that...." These are rather weak claims, and we don't know exactly what to say about them. What is more important to us is to try to isolate what seems to be doing the *primary* work in our intuitive reactions to manipulation cases.

Bratman usefully wonders how exactly we can get the result we want here. That is, we claim that Judith is not morally responsible to the extent that she has not taken responsibility for the "manipulation mechanism." Bratman wonders how we are individuating mechanisms—a difficult but fair question. He points out that it will not work to specify the manipulation mechanism as "the actual motivational workings of the desire" or "the earlier causal history of implanting the desire." We do not have a well worked-out theory of mechanism individuation. Our methodological approach is to rely on what we take to be intuitive and natural ways of individuating mechanisms, trying to be consistent throughout the book. Then we ask the reader to check to see if we have indeed been consistent, and to evaluate the overall theory in light of its ability to resolve difficult puzzles and challenges about moral responsibility. More specifically, we have in mind a way of individuating mechanisms which does not require that the agent know everything about the causal origins of its inputs or its inner workings. We think that it is tolerably clear that ordinary practical reasoning is in some way interestingly different in kind from practical reasoning in which crucial inputs have been implanted through direct manipulation, or the inputs are being processed through direct electronic manipulation. Note that this is the basic intuition that drives our whole analysis of the Frankfurt-type cases. That is, we contend that in the Frankfurt-type cases different mechanisms operate in the alternative scenario and the actual sequence: whereas ordinary practical reasoning operates in the actual sequence, a manipulation mechanism (however precisely one specifies this) operates in the alternative scenario.

Bratman points out that the scientists have implanted in Judith a "strong but not irresistible" desire to be violent. He concludes that when Judith gives in to this urge we have a case of "culpable irresponsibility, not of non-responsibility." But it just seems that Bratman is being too hard-hearted and severe here. Recall that Judith has always been a peaceful person—violence of any sort, much less attacking a friend who simply commits the crime of meeting her for a cup of coffee, has always been abhorrent to her. The scientists completely change this without Judith's consent. Now she has an extremely strong urge to hit her friend, although (holding her actual mechanism fixed) she would resist if (say) that were the only way to save the earth from nuclear annihilation. Nothing remotely like these imaginary scenarios obtains, and she acts on her extremely strong urge. We think it is unduly severe to suppose that Judith is morally responsible here.[6]

---

[6] In our view, Judith is thus different from agents with certain phobias, such as Fred (in Mele's contribution to this symposium). Even if Fred is not morally responsible for developing the phobia, its development in him can be assumed to take a certain course which makes it plausible to say that the mechanisms which issues in his behavior are "his" in a way in which the manipulation-mechanism is *not* Judith's. Thus, Fred can be said to be morally responsible for not going to his daughter's wedding, even though he may well not

III. Reply to Stump

The Direct Argument purports to show that causal determinism is incompatible with moral responsibility, quite apart from any considerations pertaining to alternative possibilities. It typically employs some sort of transfer of non-responsibility principle, such as Transfer NR: If (1) p obtains and no one is even partly morally responsible for p; and (2) If p then q obtains, and no one is even partly morally responsible for the obtaining of this conditional; then (3) q obtains and no one is even partly morally responsible for q.

In *Responsibility and Control* we argue that the Direct Argument does not succeed. We contend that the validity of the Transfer Principle employed in the argument has not been established by its proponents. As Stump points out, we use certain cases involving preemptive overdetermination, and other cases involving simultaneous overdetermination, to cast doubt on Transfer NR (and some related principles). We call such cases "two-path" cases. In two-path cases there can be one path along which it is (relatively) uncontroversial that the relevant agent exercises control and thus is morally responsible for bringing about the result, even though the other path contains some responsibility-undermining factor. We argue that in virtue of the existence of the responsibility-conferring path, the agent can be deemed morally responsible, even though the Transfer NR Principle, seizing on the responsibility-undermining path, would have it that the agent is not morally responsible for the obtaining of the relevant state of affairs.

In her challenging and thoughtful comments, Stump agrees with our objections to Transfer NR, but suggests that there is another transfer of non-responsibility principle which would be immune to our worries and also potent in a direct argument for incompatibilism. She says, "It might be that any counterexample to Transfer NR has some feature which can't be found in any case involving responsibility and determinism. In that case, Fischer and Ravizza's strategy for undermining Transfer NR couldn't be employed in the cases relevant to moral responsibility and causal determinism. If Transfer NR is given a restricted formulation which excludes the Fischer and Ravizza cases of overdetermination, then, unless cases of moral responsibility can be assimilated to such cases, the direct argument for incompatibilism which relies on Transfer NR will still be sound." Stump is essentially contending that if causal determinism is assumed to obtain, then all cases will be one-path cases, insofar as any choice and subsequent behavior will be the result of a deterministic causal chain. This entails that any path in which one supposes that one has found moral responsibility will in fact have the (purportedly) responsibility-undermining feature—causal determination. Thus, if one

---

be blameworthy. In contrast, Judith is not even morally responsible for her violent attack. This is, admittedly, a subtle discrimination—one whose elaboration is an intriguing challenge that issues from Bratman's probing critique.

restricts Transfer NR to one-path cases in some suitable way then obviously two-path counterexamples will be irrelevant, and it would also appear as if the incompatibilistic result will have been achieved. Stump concludes, "...the Fischer and Ravizza strategy fails. Their counterexamples to Transfer NR all require a certain feature for their success—namely, an alternative pathway in which someone is clearly responsible—which isn't in the cases involving moral responsibility and causal determinism."

What is behind Stump's worry? Think of a case in which an agent deliberates in the "normal" way—no phobias, addictions, direct manipulation of the brain, subliminal advertising, irresistible urges, and so forth. It might be thought that the path leading to the individual's decision and behavior confers moral responsibility. But if causal determinism is true, then on that very path is causal determination (by past states of the universe and the natural laws). The determination is in the very path we had previously supposed to be responsibility-conferring; the causal determination "works through" the deliberation, and is thus in some crucial sense inseparable from it. Thus we don't have two separate paths, but one. So moral responsibility gets no opening against the onslaught of the responsibility-undermining factors fixed on by Transfer NR.

Stump evidently supposes that the only way our "strategy" can succeed is by decisively showing Transfer NR, and all its relevant "restrictions" (or, perhaps, all related transfer of nonresponsibility principles) to be invalid. We do not believe that we can do this, nor do believe that we need to.

Imagine, for a moment, that we live in a self-contained (peace-loving and unmeddlesome) community in which there is some belief *B* that is central to our ways of thinking about ourselves and others, and is the basis of a set of practices that is also central to our way of life. We are content with this belief and these practices. But one day someone from another community visits us, and proclaims, "I can prove to you that your belief *B* is false, and thus your whole way of life is a cruel delusion." We are willing to listen, and he proceeds to produce an argument against *B*, an argument which employs as an indispensable ingredient, a principle *P* (alleged by the visitor to be valid). The visitor holds that *P* indeed captures something deep in our own views. Rejecting *P* (and thus saving our central belief B) would thus be to reject important parts of our commonsense understanding of the world.

But then we press the visitor about *P*. It turns out—and the visitor concedes this—that *P* in its "unrestricted" form is *not* valid: it does not capture our considered judgments in a wide variety of contexts. Whatever evidence is invoked by the visitor on behalf of *P* (i.e., contexts in which it looks like *P* yields intuitively the correct answer), then, is inconclusive—it cannot establish that *P* is valid, because there is counterevidence from other contexts. Thus, at this point, we are perfectly within our rights to tell the visitor that we are not yet convinced by his argument, since he has not

employed a valid principle to get to his skeptical result (a result we have good reason to decline to embrace). Our commonsense understanding of the world does *not*—or certainly has not been shown to—require acceptance of *P*.

Now suppose the visitor says, "Ok, I grant that you are correct so far. But now consider a restriction of *P*, *P**, designed so that *P** does not apply to those contexts in which we have previously agreed that there are problems. Indeed, *P** only applies to the context of *B*, and it can be invoked in a valid argument which shows that *B* is false. The only way you would be justified in maintaining your comfortable belief *B* (and associated way of life) would be to show me that *P** is invalid (by producing a nonquestion-begging counterexample to it)."

We cannot see why the members of our peace-loving and unmeddlesome community should accede to the visitor's demand. Why do they have the obligation to prove in a nonquestion-begging way that *P** is invalid? After all, they were just peacefully minding their own business until the visitor came to town, and it was the visitor who said that *he* could *prove* that they were mistaken in holding *B*. (They did not venture into his community and say that they could prove his principle wrong, nor did they have any interest in doing so.)

Why shouldn't the members of the community reply to the visitor as follows? "We do not find any strong reason to accept your principle *P** (in the absence of which there is no reason to worry about our *B*). Granted, we cannot prove that it is invalid (in a nonquestion-begging way). But this is not something that it is reasonable to expect us to do. We are very willing to consider evidence for the principle *P**. The problem is that you began with an intuitively natural, plausible principle *P* which you alleged applies quite generally and also generates (in combination with other plausible ingredients) the surprising and—to us—disturbing result that *B* is false. You invoked evidence for this principle. But then when we considered the matter more carefully, we found that *P* is in fact invalid; we found there to be evidence from a set of contexts which you had ignored, and by reference to which it could be seen—even by you—that *P* is invalid. You then deliberately qualified *P* to produce *P**, a principle which only applies to a restricted set of contexts. Indeed, it seems *only* to apply to the context of *B*, which is the very belief in question! Now you claim that *P** is valid, but on what basis? It no longer is a natural, intuitive, *general* principle which gains support from yielding intuitively plausible results across a range of contexts, including those which are not the home of the very belief in question. Rather, it seems as if you are simply *insisting* that it is true.

We believe that the situation is (roughly speaking) the same with regard to our very central intuitive belief that we are morally responsible agents (even under the assumption that a consortium of physicists should in the future ascertain that the equations describing the universe are deterministic). A

skeptic may seek to employ Transfer NR to dislodge this central belief, which forms the basis for important features of our way of life. But the evidence invoked to support Transfer NR is not decisive, and evidence from certain contexts ignored by the proponents of Transfer NR shows (uncontroversially) that Transfer NR is invalid. One could (in some unspecified way) "restrict" Transfer NR so that it does not apply to the contexts of countervailing evidence, but only applies to the sort of context in which causal determinism would be true. But now what evidence can be adduced to support the restricted principle? Wouldn't one have reason to be skeptical about the plausibility of the principle, given that it has been generated by a restriction meant to rule out counterexamples?

It would seem that any evidence that could be adduced in support of the restricted principle would be question-begging, and any evidence that could be adduced to show that the principle is invalid would also be question-begging. This is a paradigmatic case of a Dialectical Stalemate.[7] But note that a compatibilist about causal determinism and moral responsibility need not prove the invalidity of the restricted transfer principle (unless of course he sets himself the task of decisively disproving any alternative theory). If there are strong reasons to accept the belief both that we are morally responsible and that this responsibility does not depend on a particular scientific theory of the universe (i.e., that the universe is indeterministic rather than deterministic), then isn't it enough to show why there is no strong reason to accept a principle inconsistent with this belief? Now some might yearn for a *decisive refutation* of such a principle, but it was not our strategy to seek such a refutation, nor do we think it reasonable to expect such a strategy to succeed.

---

7 For a discussion of Dialectical Stalemates, see John Martin Fischer, *The Metaphysics of Free Will* (Oxford: Blackwell Publishers, 1994), pp. 83–85.