

## Article

## Theme: AUTHOR MEETS CRITICS

## Responses to John Martin Fischer and Dana Nelkin

Derk Pereboom  
Cornell University  
Email: dp346@cornell.edu

Let me begin by expressing my gratitude to Gregg Caruso for inviting this symposium, and for writing such a clear, accurate and valuable précis.

## Response to John Fischer

I first want to thank John Fischer for his generous appraisal of the book, and for his astute and challenging comments on my treatment of the manipulation argument in Chapter 4. Fischer's core strategy for resisting this argument is (at least initially) a soft-line reply. Soft-liners claim that in some manipulation cases the agent is not morally responsible, and in others he is. A corollary of the soft-line reply is that there is a plausible compatibilist condition on moral responsibility that has not been met in some of the cases. One prominent response of this sort is that a key condition on basic desert moral responsibility (the notion of moral responsibility I will assume in this discussion of manipulation arguments) is the absence of intentional manipulation or causal determination (Lycan 2007, Feltz 2013, Murray and Lombrozo forthcoming). This is not the line Fischer takes. In the response he proposes, intentional causal determination is compatible with moral responsibility in Case 2, that is, in what he calls an initial design case, in which the intentional causation, as in Leibniz's theological determinism, is confined to the beginning of the agent's life. But in Case 1, where the intentional manipulation is direct and immediate, the agent is not morally responsible. Fischer's thought is that Plum's non-responsibility in Case 1 can be accounted for by the conditions on mechanism-ownership, a feature of the compatibilist account of moral responsibility he and Ravizza (1998) developed, which, he correctly points out, I neglect in favor of the reasons-responsiveness component.

I do provide a criticism of the mechanism-ownership feature of the account in *Living without Free Will* (2001), but my main argument there is that the agent can be manipulated or intentionally caused to satisfy the conditions of mechanism ownership, and the resulting intuition that the agent is not responsible would then undermine the compatibilist value of that component of the account. Fischer rightly points out, however, that he and Ravizza acknowledge that an agent can be manipulated or intentionally caused to satisfy the conditions of mechanism-ownership. I would at this point press the contention that this fact is a threat to their compatibilist proposal, but let's set that aside. I also won't pick at Fischer and Ravizza's specific account of mechanism ownership. Enough of that has been done, and Fischer agrees that the account needs some work. But I think that we should allow that the general notion of ownership or endorsement of one's powers of agency, set out along the broad outlines of the Fischer-Ravizza account, is an intuitively promising element of a compatibilist view. So instead of contesting that, specifically, I'll argue that there is intuitively as much mechanism-ownership in Case 1 as in Case 2 – or enough for someone with Fischer's compatibilist sensibilities. So at least for the manipulation argument as I set it out in the new book, mechanism-ownership won't lend support to a soft-line reply. If what I have to say is right, this should push Fischer toward a hard-line account, which, in general, incompatibilists will prefer to a soft-line reply.

The version of a local manipulation case that most easily falls prey to the absence of the agency-ownership objection is one in which the neuroscientists are not merely able to manipulate the agent's neural states directly at any time, but actually do in fact directly

manipulate him constantly. One might argue, for instance, that such an instance of Plum does not exhibit the sort of causal profile required for him to be a genuine agent (Fischer 2004, Mele 2005, Baker 2006, Demetriou (Mickelson) 2010). In addition, imagine that the manipulators directly intentionally cause mental states that would ordinarily realize the satisfaction of the mechanism-ownership requirement. In such circumstances it will be intuitive that the agent hasn't really come to own her agency.

I point out in the book that several of the concerns raised for a continuous and direct manipulation case highlight the fact that in an effective version of Case 1 two desiderata must be secured at the same time: the manipulation must preserve satisfaction of intuitive conditions of agency, and it must render it plausible that Plum is not morally responsible. To this let me now add that the manipulation must preserve the satisfaction of intuitive conditions on agency ownership. Accordingly, the account I develop in the book should be augmented in the following way. Agency, and agency ownership is regularly preserved in the face of certain direct, temporally localized external influences. There are many such influences on agency that a compatibilist, in particular, a compatibilist of Fischer's sort, who argues that moral responsibility does not hang on a thread, would affirm to be compatible with agency and ownership of agency. It's a commonplace that hearing disappointing news of events that are beyond one's control has a tendency to result in bad behavior. And recent studies indicate that we are likely to assess a person or an action more negatively when our disgust reaction is triggered, by, for instance, a bad smell or a dirty visual display. Such influences are so commonplace that if one's compatibilism is even minimally resilient, one would need to agree that agency and agency-ownership are preserved under such conditions. At this point one might caution that the states that realize agency-ownership must not be caused in this way, but instead, in accord with the Fischer-Ravizza model, result from what is intuitively the agent's own reflection on her agency.

So here is how I now want to set up Case 1. We commonly suppose that acting in a way that is influenced by direct, temporally localized external influences is compatible with moral responsibility. However, we can imagine an egoism-enhancing momentary influence that preserves agency but intuitively rules out moral responsibility (cf. Shabo 2010). Suppose that by

way of neural intervention the manipulators enhance Plum's disposition to reason self-interestedly at the time he contemplates killing White, so that the manipulators know that as a result it is causally ensured that he will decide to do so. Like the effect of smelling something disgusting, this intervention would not undermine Plum's agency. But intuitively enhancing the egoistic disposition in this specific way does render him non-responsible for his decision.

Here is the full description of the case:

**Case 1:** A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms to his second-order desires. In addition, his process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire – the neuroscientists do not induce a desire of this sort.

First, it's my intuition that Plum is not morally responsible in this case. McKenna (2008) and other hard-liners will disagree, but I'll assume for now that Fischer, as a soft-liner, is provisionally on board with this intuition.

Second, Fischer expresses the concern that it may

be assumed, contrary to what I specify, that Plum in this case acts on an irresistible psychological factor. In response, let me add the following specification. The neuroscientists calculate the minimal egoistic strength of the reasoning process required to causally determine the decision to kill White in these exact circumstances. Suppose that it turns out to be 15 egoistic strength units (ESUs). The production of additional ESUs is expensive, so they induce precisely 15 ESUs in this situation. We can now imagine a nearby possible world in which Plum has just received news that a paper of his was accepted at a fine journal, or one in which he has just received a higher-than-expected salary increase, in which it would take 16 ESUs to causally determine the decision to kill. So in nearby possible worlds in which we hold fixed the actual egoistic strength of the reasoning, Plum does not choose to kill. On Fischer's account, the presence of such nearby possible worlds should rule out Plum's acting on an irresistible causal factor.

In addition, Plum would now also appear to meet the requirements for guidance control (Fischer 1994, Fischer and Ravizza 1998). That is, in Case 1 thus specified, Plum's decision involves his rational consideration of reasons relevant to the situation, and in some alternative circumstances in which there are sufficient reasons for him to do otherwise than he actually does, he would be receptive to these reasons and would have chosen and done otherwise by the efficacy of the same deliberative mechanism that actually results in the action. In the world in which he receives the salary-increase news, there are sufficient reasons for him to refrain from deciding to kill, and he would be receptive to those reasons and would have done otherwise by the ordinary mechanism of rational deliberation, which is also efficacious in the actual situation. Perhaps Fischer will now have the intuition that Plum is responsible in Case 1. But I would emphasize that in this situation the neuroscientists correctly calculate that in the actual situation, inducing 15 ESUs guarantees a decision to kill, and they then induce those 15 ESUs.

Let me note and respond to a related concern raised by Eddy Nahmias (in a talk he gave at the University of Fribourg in the summer of 2013). Those of us who judge Plum non-responsible in Case 1 might be assuming a relevant and important difference between Case 1 and Case 4—the ordinary deterministic example. In Case 4 we are likely to suppose that if Plum's

initial altruism were stronger, he might or would refrain from making the decision to kill. But in Case 1, we're likely to think that if Plum's initial altruism were stronger, the neuroscientists would increase the induced ESUs to achieve the result they desire. After all, manipulators tend to be intent on achieving their objectives, and this leads us to expect that they would modify the manipulation as needed. So the assumed causal profile of nearby possible worlds in Case 1 and Case 4 differs, and this is what explains the variation in intuitions of moral responsibility. To correct for this, we can specify that the technology the neuroscientists are using is limited, and they can't induce more than 15 ESUs, and in this type of situation the only decision-producing technology at their disposal involves producing ESUs. If Case 1 is embellished in this way, Cases 1 and 4 will not feature a relevant difference in causal profiles of nearby possible worlds. If at this point one's intuition that in Case 1 Plum is not responsible wavers, I would again emphasize again that in this example the neuroscientists have correctly calculated that in the actual situation inducing 15 ESUs guarantees the decision to kill, and they then induce those 15 ESUs.

Returning to the question of mechanism-ownership, I propose that from Fischer's compatibilist point of view, Plum in Case 1 specified in this embellished way should satisfy any plausible ownership condition on the practical-reasoning mechanism. In the situation in which someone acts badly after hearing that the home team lost we would want to say that mechanism-ownership remains in place, as we would when someone behaves more generously after hearing that her paper has been accepted. One might reply that it's not Plum's practical reasoning mechanism that's producing the action in Case 1. But then it seems one would need to say that in any situation in which some external momentary factor immediately influences the agent, mechanism ownership is compromised. This would lead to a seriously non-resilient compatibilism, to which Fischer would be averse. Moreover, while I've argued that in general the Fischer-Ravizza account does not fall prey to the objection that mechanism-individuation is apt to be rigged to yield the intuitively correct moral responsibility judgments (Pereboom 2006), it would risk occasioning this objection if one were to claim that it's not the practical reasoning mechanism that produces the action in Case 1 but it is in the case of generous behavior upon hearing good news.



I agree with Fischer about the dialectical role of manipulation arguments. It would be best for incompatibilists and compatibilists to alike take a step back from the position they endorse when forming intuitions about the cases, and to approach a neutral stance. True, those of us who are committed to either compatibilism or incompatibilism are not particularly adept at actually realizing full neutrality, but we can attempt to approximate this ideal. The participants in the discussion who actually stand the best chance of being influenced by the argument are those who haven't yet taken a stand, especially those who are just now becoming familiar with the debate. For such a context, I will reiterate a point I make in the book (which largely accords with Patrick Todd's (2011) general assessment). Most of those who enter into the free will debate begin with the assumption that human agents are blameworthy in the basic desert sense when they knowingly do wrong. Some people are natural compatibilists, and for them the prospect that whenever we act, we are causally determined by factors beyond our control wouldn't change this assumption. Others are natural agnostics about compatibilism and incompatibilism, and for them the responsibility assumption would be challenged but not defeated by the prospect of causal determination. Incompatibilists believe that these two types of reaction fail to face up to the implications of causal determination. Manipulation arguments aim to remedy this presumptive shortcoming by first devising a case of deterministic intentional manipulation with the hope that it will be more successful at eliciting a non-responsibility intuition than does causal determination alone. The next step is to argue that non-responsibility is preserved even when the intentional manipulation is subtracted, for the reason that there is no responsibility-relevant difference between the intentional manipulation case and the ordinary deterministic one. The salient common factor is causal determination by factors beyond the agent's control, or on Carolina Sartorio's (2014) formulation, by factors beyond her causal reach, and this feature will therefore be sufficient for non-responsibility.

The intention is that a deterministic intentional manipulation case would help reveal that ordinary causally determined agents lack the sort of free will required for basic desert moral responsibility. In this way the manipulation argument aims to persuade the natural compatibilist and the natural agnostic that their resistance to incompatibilism is best given up.

Manipulation cases are designed to sway intuitions in a dialectical situation in which prior reasoning has not broken the impasse. To this let me add that the argument, as I conceive it, is not rationally coercive. The incompatibilist can hope that the intuitions about the cases will go his way. He can strengthen his argument by pointing out, for instance, that in the dialectical situation it's not appropriate to privilege intuitions about ordinary cases. But he cannot justify claiming that when the compatibilist's intuitions don't go his way, she is irrational.

## Response to Dana Nelkin

Let me begin by thanking Dana Nelkin for the kind words about my book, and for the perceptive and valuable comments about the part on the practical implications of the free will skepticism I propose. I agree with Nelkin that if causal determinism undermined basic desert blameworthiness, it would also threaten moral obligation. In fact, given that I advance a source rather than a leeway view, backed by what I consider to be a successful Frankfurt example, the agential requirements implicated by obligation will be more stringent than they are for basic desert blameworthiness. If an agent ought not have done A, this would require, by the 'ought' implies 'can' principle (OIC), that she could have refrained from doing A. By contrast, the agent's basic desert blameworthiness for A would not require that she could have so refrained. But still, in my view both would be undercut by causal determinism, at least supposing the Consequence Argument is sound with respect to the notion of 'could have done otherwise' at issue in OIC. A further and related concern is that if obligation were undermined by the general causal determination of action, moral wrongness would also be undercut. Ishtiyaque Haji (1998) argues that S has a moral obligation to perform A just in case it is morally wrong for S not to perform A, and, similarly, that S has a moral obligation not to perform A just in case it is morally wrong for S to perform A. On his view, then, moral wrongness and moral obligation stand and fall together. Let's consider these two threats in turn.

In the book I argue that a response to determinism's threat to 'ought' judgments might take advantage of the plausible claim that 'ought' has a range of correct uses, and as C. D. Broad (1952) suggests, it may be that not all are linked to an OIC requirement, or at least to an OIC requirement that lends support to

the incompatibility of causal determinism and 'ought' judgments. First, Ruth Barcan Marcus (1966), Lloyd Humberstone (1971), and Gilbert Harman (1977) distinguish between an 'ought' that applies to action and one that applies to states of affairs. An 'ought to do,' Harman contends, "implies that an agent has a reason to perform an action, while an 'ought to be' evaluates a state of affairs and does not by itself imply that any particular agent has a reason to contribute to bringing about that state of affairs" (Harman 1977: 87; cf., Humberstone 1971, Manne 2011). Similarly, Mark Schroeder (2011) distinguishes between the action-related *deliberative* sense of 'ought', and the *evaluative* 'ought', as in 'Larry ought to win the lottery' where Larry has been subject to a series of undeserved misfortunes. James Hobbs (2012) persuasively argues that certain 'ought to be' claims will have a loose kind of implication for reasons agents have to act. He offers the following account: if X *ought to be* the case, then agents to whom the relevant considerations apply have a reason to act in ways that respect the value of X. But he argues that the satisfaction of this condition does not imply a route that is accessible to the agent to the realization of what ought to be. However, if instead an agent *ought to do* something, then she has a reason to do it, and there is such a route to what she ought to do. Nelkin (2011: 111) contends, and I agree, that 'ought' propositions that specify what an agent ought to do are essentially action-directed, so that if 'S ought not do A' is true, then as a matter of the meaning of 'ought' judgments, or of the essential nature of obligation, S is thereby directed to refraining from A, and this entails that S can refrain from A (see Copp 2008 and Stapleton 2010 for similar views). One way to think about this distinction is that an 'ought to be' is an 'ought' of *axiological evaluation*, or sometimes of *axiological ideality*, which does not entail a 'can' claim, whereas an 'ought to do' expresses a demand made of an agent in a particular circumstance, which does entail that the agent can perform the specified action (cf., Humberstone 1971; Manne 2011). We can call this second type an 'ought' of *specific action demand*.

I think it's plausible that if causal determination precludes alternative possibilities for motivation and thus for action in a sense relevant to the issues under consideration, then the specific action demand notion of 'ought to do' will be in jeopardy. If I know that in this relevant sense an agent could not have avoided lacking the motivation required for refraining from performing an immoral action, it would be unfair, and I

think, mistaken, to claim that she ought not to have performed that action at that time. However, as Kate Manne (2011) contends, plausibly to my mind, the 'ought' of axiological evaluation applies not only to non-agential states of affairs, but also to actions. Thus one might say of the husband who can't control his abusive behavior that it ought to be the case that he not behave abusively. And it might well not be mistaken or unfair to recommend to him that he refrain from such behavior in the future, supposing that it's reasonable to believe that he will come to have the ability to consistently refrain from abusive behavior, and that it's epistemically open that he will acquire the requisite motivation, and in particular if it's reasonable to believe that making this recommendation would contribute causally to bringing about the motivation.

Thus I'm proposing that given determinism and that determinism precludes alternatives, when one tells an agent that he ought to refrain from performing an action of some type in the future, it's not the 'ought' of specific action demand, but rather the 'ought' of axiological evaluation that's legitimately invoked. This use of 'ought' proposes as morally valuable a state of affairs in which the agent refrains from performing the action and recommends that he not perform it. I call this the 'ought' of *axiological recommendation*. By contrast with the 'ought' of specific action demand, it is not an 'ought' of moral obligation. Supposing the general causal determination of action and that such determination rules out the relevant alternative possibilities for action, telling someone that he ought not have performed some bad action, using the 'ought' of specific action demand, would not be correct or fair. But under these circumstances telling that person that he ought not perform actions of this sort, using the 'ought' of axiological recommendation, would not be similarly undermined.

Nelkin asks whether relationships such as friendship are possible supposing that the 'ought' of moral obligation is ruled out. On a plausible proposal, what sets friendship apart from mere acquaintances, for example, is a structure of obligations. If A and B are friends, and if A is in dire need, then there is a strong *prima facie* moral obligation for A to help.

In response, I suggest that there is a near functional equivalent for the structure of obligations in friendships that does not appeal to the notion of obligation, nor to any notion threatened by free will skepticism.

ticism. We might begin with a thought from Kant's *Groundwork of the Metaphysics of Morals* (1784/2009). In Kant's conception the moral law does not pertain to God in the form of principles of moral obligation. The moral law does, however, apply to us in this form, but only for the reason that we are subject to inclinations, motivations that are apt to conflict with morality. The moral law applies to us in an action-directive form because it's possible for us to act in ways that do not conform to it. This account suggests that in Kant's view the moral law's obligatory aspect is in a sense contingent on the nature of the human agents, and other similarly imperfect possible agents. Even though Kant does not say this explicitly, given this perspective the essence of the moral law is independent of the notion of moral obligation.

One might thus look for an essential feature shared by the way in which the moral law applies to God and to us. The tradition gives us a candidate: love in the sense of charity, that is, *caritas* in Latin, *chesed* in Hebrew, and *agapē* in Greek, sometimes translated as loving-kindness, which one might think of as good will towards beings with moral standing generally. This virtue is arguably insufficiently circumscribed to count as an alternative to the notion of obligation in characterizing friendship. But, plausibly, this type of love generates specific *commitments* in particular contexts. For instance, in the case of friendship loving-kindness would generate a commitment on the part of each friend to provide help in dire need. True, such a moral commitment can be understood as holding oneself to moral obligation. But, plausibly, the notion of commitment isn't restricted to this sense. One can instead understand a commitment as a recommendation to oneself that one resolves to honor.

Still, we human beings are subject to moral frailty, and so one might wonder how to conceptualize the interpersonal moral relationship between friends A and B in a situation in which there is a threat that A will not come to B's help in time of dire need, and thus not to honor his own commitment. Here it's natural for A to think of himself as morally obligated to help B, and for B to make a demand of moral obligation that A help her. But we've provisionally relinquished the notion of moral obligation and the correlative notion of moral demand. Instead we can frame B's expectation in terms of the notion of moral wrongness: A can think that it would be wrong for him not to help B, and B can communicate to A that it would be wrong

of him not to help her. However, this brings us to our second threat, according to which judgments of normative wrongness are undermined by causal determination. Haji (1998, 2002) argues that due to the tight connection between moral obligation and moral wrongness, the threat posed to judgments of moral obligation extends to those of moral wrongness. Crucial to his argument is the following principle:

S has a moral obligation to perform [not to perform] A if and only if it is morally wrong for S not to perform [to perform] A.

If this principle is true, then if judgments of moral obligation are ruled out by causal determination, judgments of moral wrongness will be excluded as well. In the book I argue that although this biconditional principle may be attractive considered in the abstract, my sense is that it captures only one aspect of the complex notion of moral wrongness. This can be made plausible by reflecting on the fact that while the left-to-right half of this biconditional

If S has a moral obligation to perform [not to perform] A then it is morally wrong for S not to perform [to perform] A,

is evidently true, the right-to-left half, i.e.,

If it is morally wrong for S not to perform [to perform] A, then S has a moral obligation to perform [not to perform] A

is less clearly secure. There's likely no example in which it's at all credible that an agent has a moral obligation not to perform an action while it is not morally wrong for him to perform it. But there are cases in which it's intuitive that performing an action would be morally wrong for an agent, while it's at least less clear that he has a moral obligation not to perform it. Imagine that a psychopath could not have avoided an act of deception due to his psychological disorder. OIC provides an intuitive basis for denying that he was morally obligated to refrain from deceiving, while it remains evident that his action was morally wrong.

An alternative non-deontological notion of moral wrongness, one that isn't biconditionally linked to moral obligation, accommodates this intuition. The core of Alastair Norcross's (2006) proposal for a purely axiological ethics involves specifying for each action-rele-



vant situation the pertinent options for acting ranked in order of value of the consequences realized, without an obligation to maximize value or even an obligation to satisfice (see Slote 1984). As Nelkin points out, the type of consequentialism that fits with the views I defend in the book is one according to which respect for rights, for instance, counts as a good consequence. For example, the claims I make about criminology invoke the right to liberty and the right to self-defense and defense of others, and respect for and absence of violation of such rights count as good consequences.

In the book I propose that an option for acting would be counted as morally wrong when its value is low enough in the ranking for it to be morally justified for a relevantly positioned interlocutor to protest the type of action (cf. Hieronymi 2001), partly on the basis of the largely forward-looking reasons set out in Chapter 6 of the book – moral formation, protection, and reconciliation in relationships. This would not be a moral justification for a deontological claim, but rather for the high moral value of consequences of protest in this case. This proposal does not characterize wrongness independently of when it is appropriate to protest for these reasons, and as a result it cannot ground the appropriateness of protest in wrongness. But it does have the consequence that it's appropriate to protest a type of action only when his acting for some reason would be morally wrong, and in thus it satisfies another important intuition we have about this relationship.

This proposal also yields an account of rights without corresponding duties, a further issue on which Nelkin has invited me to respond. Suppose we agree that by virtue of the friendship, A has a right to B's help in time of dire need. In Nelkin's view, an essential feature of friendship that becomes salient in this situation is B's moral obligation to help A. On the view I'm proposing, what's salient is rather that given the nature of friendship, (i) B has a commitment to help A, and (ii) it's appropriate, on consequentialist grounds, for A to protest the threat of B's not helping her. Thus the role of moral obligation in a relationship has a near functional equivalent whose constituents are commitment and the appropriateness of protest in cases in which there is a threat that the commitment will not be honored.

A challenge Nelkin raises for the kind of consequentialism on offer is that there are possible situations in which the theory would tell us that it would be better to violate a right to prevent even more violations, and

this is in a way quite radical or at least revisionary. In response, the consequentialist can point out that in standard examples for illustrating this problem there are resources she can use to secure the ordinary moral intuitions. For instance, in the case in which the magistrates can secretly frame an innocent person to prevent a riot in which many will be killed, she can point to the extremely bad consequences that would result from the scheme's being uncovered, notably general loss of confidence in the judicial system. Because this sort of thing is almost always is found out, high expected disutility would result. And if the number of people expected to be killed in the riot is high enough, our sense that framing an innocent person is impermissible begins to falter. Once considerations of this kind are acknowledged, the difference between ordinary moral intuitions and what the consequentialist recommends may well not be all that different. But in the end consequentialists will have to recognize trade-off between some degree of revisionism and the theoretical virtues of their position.

Finally, note that I characterize wrongness in terms of the justifiability of protest, and not in terms of the appropriateness of resentment or indignation (cf. Strawson 1962). These reactive attitudes are essentially backward looking insofar as they presuppose or are closely associated with a presumption of basic desert. Such a presumption would be ruled out by free will skepticism, but also by an independently motivated, largely forward-looking conception of morality (Pereboom, forthcoming). The notion of protest, in my view, is positive and forward-looking in just the right sort of way.

## Acknowledgements

Thanks to John Fischer and Dana Nelkin for valuable comments on an earlier draft.

## References

- Baker, L. R. 2006. "Moral Responsibility without Libertarianism," *Noûs* 40, pp. 307–30.
- Broad, C. D. 1952. "Determinism, Indeterminism, and Libertarianism," in his *Ethics and the History of Philosophy*, London: Routledge and Kegan Paul, pp. 195–217.
- Caruso, Gregg. "Précis of *Free Will, Agency, and Meaning in Life*," *Science, Religion and Culture* 3: 178–201.
- Copp, D. 2008. "'Ought' Implies 'Can', and the

- Derivation of the Principle of Alternate Possibilities" *Analysis* 68, pp. 67-75.
- Demetriou, (Mickelson) K. (2010). "The Soft-Line Solution to Pereboom's Four-Case Argument," *Australasian Journal of Philosophy* 88, pp. 595-617.
  - Feltz, A. 2013. "Pereboom and Premises: Asking the Right Questions in the Experimental Philosophy of Free Will," *Consciousness and Cognition* 22, pp. 53-63.
  - Fischer, J. M. 1994. *The Metaphysics of Free Will*, Oxford, Blackwell Publishers.
  - Fischer, J. M., and M. Ravizza. 1998. *Responsibility and Control, A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
  - Fischer, J. M. 2004. "Responsibility and Manipulation," *The Journal of Ethics* 8, pp. 145-77.
  - Haji, I. 1998. *Moral Appraisability*, New York: Oxford University Press.
  - Haji, I. 2002. *Deontic Morality and Control*, Cambridge: Cambridge University Press.
  - Hieronymi, P. 2001. "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62, pp. 529-54.
  - Harman, G. 1977. *The Nature of Morality*, Princeton: Princeton University Press.
  - Hobbs, J. 2012. Dissertation, Department of Philosophy, Cornell University.
  - Humberstone, L. 1971. "Two Sorts of 'Oughts'," *Analysis* 32, pp. 8-11.
  - Kant, I. 1784/2009. *Groundwork of the Metaphysics of Morals*, New York: Harper Collins, tr. H. J. Paton.
  - Lycan, W. G. 1997. *Consciousness*, Cambridge: MIT Press University Press.
  - Manne, K. 2011. *Not by Reasons Alone*, MIT, PhD Dissertation.
  - Marcus R. B. 1966. "Iterated Deontic Modalities," *Mind* 75 580-82.
  - McKenna, M. 2008. "A Hard-Line Reply to Pereboom's Four-Case Argument," *Philosophy and Phenomenological Research* 77, pp. 142-59.
  - Mele, A. 2005. "A Critique of Pereboom's 'Four-Case' Argument for Incompatibilism," *Analysis* 65, pp. 75-80.
  - Murray, D. and T. Lombrozo. forthcoming. "Effects of Manipulation on Attributions of Causation, Free Will, and Moral Responsibility."
  - Nelkin, D. 2011. *Making Sense of Moral Responsibility*, Oxford: Oxford University Press.
  - Norcross, A. 2006. "Reasons without Demands: Rethinking Rightness," in *Blackwell Contemporary Debates in Moral Theory*, J. Dreier, ed., Oxford: Blackwell.
  - Pereboom, D. 2001. *Living without Free Will*, Cambridge University Press.
  - Pereboom, D. 2006 "Reasons Responsiveness, Alternative Possibilities, and Manipulation Arguments against Compatibilism: Reflections on John Martin Fischer's *My Way*," *Philosophical Books* 47, 2006, pp. 198-212.
  - Pereboom, D. 2014. *Free Will, Agency, and Meaning in Life*, Oxford: Oxford University Press.
  - Pereboom, D. forthcoming. "A Notion of Moral Responsibility Immune to the Threat from Causal Determination," *The Nature of Moral Responsibility*, Randolph Clarke, Michael McKenna and Angela Smith, eds., Oxford University Press, forthcoming.
  - Schroeder. M. 2011. "Ought, Agents, and Actions." *Philosophical Review* 120, pp. 1-41.
  - Sartorio, C. 2014. "The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will," in *Surrounding Free Will*, Alfred Mele, ed., New York: Oxford University Press.
  - Shabo, S. 2010. "Uncompromising Source Incompatibilism," *Philosophy and Phenomenological Research* 80, pp. 349-83.
  - Slote, M. 1984. "Satisficing Consequentialism," *Proceedings of the Aristotelian Society*, Supplemental Volume 58, pp. 139-63.
  - Stapleton, S. 2010. *Hard Incompatibilist Challenges to Morality and Autonomy* Dissertation, Department of Philosophy, Cornell University, Ithaca NY.
  - Strawson, P. F. 1962. "Freedom and Resentment," *Proceedings of the British Academy* 48, pp. 1-25.
  - Todd, P. 2011. "A New Approach to Manipulation Arguments," *Philosophical Studies* 152, pp. 127-133.