Penultimate Draft. Published Version available at *Philosophia*:
http://link.springer.com/article/10.1007/s11406-013-9510-x

Answerability, Blameworthiness, and History

Daniel Miller

Introduction

While most theorists agree that we can be responsible and blameworthy for non-voluntary states such as attitudes, evaluative judgments, and character traits, many hold that responsibility and blameworthiness for the non-voluntary is always derivative in nature; that is, that it is always traceable to responsibility and blameworthiness for prior voluntary actions. Accounts that maintain this view have been called "volitional" accounts.[1] In contrast, non-volitional accounts deny that all responsibility and blameworthiness is somehow rooted in voluntary actions.[2] According to non-volitional accounts, then, we may be responsible or blameworthy for an attitude or a belief without being responsible or blameworthy for some prior action. In requiring that blameworthiness for the non-voluntary be somehow rooted in prior action, volitional accounts maintain that blameworthiness for attitudes and values is history-sensitive; that is, a person's blameworthiness for such things depends in part on how she came to acquire them. Because non-volitional accounts do not require that blameworthiness for the non-voluntary be traceable to prior action, most non-volitional accounts (to their detriment, I will argue) lack historical conditions on blameworthiness for such things.

This paper focuses on a non-volitional account that has received a good deal of attention recently, Angela Smith's rational relations view. I do not take a stance here on the debate between volitionism and non-volitionism. Instead, I argue that without historical conditions on blameworthiness for the non-voluntary non-volitionist accounts like Smith's are (i) vulnerable to

[1] McKenna 2008a uses the term "voluntarist" to refer to this type of account. For more detailed discussion on volitional accounts of responsibility and blameworthiness see Smith 2005.
[2] McKenna 2008a uses the term "non-voluntarist" to refer to such accounts, and Levy 2005 uses the term "attributionist." For examples of such accounts, see Adams 1985, Scanlon 1998, and Smith 2005.

manipulation cases and (ii) fail to make sufficient room for the distinction between badness and blameworthiness. Towards the end of the paper I propose conditions aimed to supplement these deficiencies. The conditions that I propose are tailored to suit non-volitional accounts of blameworthiness; unlike some volitional historical conditions on blameworthiness, the conditions that I propose do not require that the person have exercised voluntary control (e.g., via choices or decisions) over the acquisition of her attitudes or values.[3]

I. Smith's Rational Relations View and Unsheddable Values

Angela Smith maintains that all responsibility is answerability, and that answerability underlies a number of key elements of our moral practices (2012).[4] Such moral practices include the blaming attitudes, such as resentment and indignation, as well as overt expressions of such attitudes. On Smith's view, a person is blameworthy for Φ (where Φ is some action, attitude, or evaluative judgment) iff she is answerable for Φ and she has violated some moral norm or obligation via Φ (2007, p. 477).[5] Call this claim (B). In section III I offer a manipulation case that constitutes a counterexample to (B). If the case is successful, there are broad implications that extend beyond Smith's own view. First, the success of the case would help settle a recent dispute between Smith and Neil Levy concerning the distinction between bad and blameworthy agents (Levy 2005 and Smith 2008). Second, contrary to what Smith and other authors have maintained, the success of my case would show that blameworthiness for attitudes and values is history-sensitive; that is, a person's blameworthiness for such things depends in part on how she came to acquire them.[6] In light of this, I propose historical conditions on blameworthiness for attitudes and values that are amenable to non-volitional accounts.

On the rational relations view, a person is answerable for something (e.g., a psychological state or an action) only if it reflects the person's evaluative judgments (Smith 2005, p. 237). For this

---

[3] Mele 2005 and Haji and Cuypers 2007 offer historical conditions similar to the ones that I will offer.

[4] This claim is controversial. Watson 1996 draws a distinction between two "faces" of responsibility: attributability and accountability. *Roughly*, the distinction is as follows: an agent is responsible for something in the first sense when it is attributable to her as a basis of moral appraisal, and an agent is responsible for something in the second sense when she is an appropriate target of responses meant to reward or penalize on the basis of it. Many philosophers have followed Watson in making this distinction, including Darwall 2006 and Fischer and Tognazzini 2011. Shoemaker 2011 acknowledges this distinction and recognizes Smith's notion of answerability as a third conception of responsibility. I will remain non-committal on this issue in this paper.

[5] Smith uses the term "culpable", but it is clear that she uses it interchangeably with "blameworthy." Shoemaker 2011 also characterizes Smith's view about blameworthiness in this way.

[6] For examples of non-historical views of responsibility and blameworthiness, see Adams 1985 (p. 19), Frankfurt 2002 (p. 27), and Smith 2005.

reason, when someone is answerable for something she is, in principle, open to demands that she cite the reasons that she takes to justify it. This account, then, allows that we may be responsible for things that are not under our voluntary control, such as the attitudes that we bear towards other people or the evaluative judgments that we hold. Smith offers the following example: "A cruel person . . . is someone who judges that the fact that something will cause pain or suffering to another is no reason to avoid it . . . It is that judgment, as reflected in her actions and attitudes, for which we consider her answerable. . ." (2008, pp. 389-390). In this way, Smith's rational relations view is non-volitional; Smith denies that all responsibility is somehow rooted in responsibility for voluntary actions.[7]

As stated above, Smith holds that a person is answerable for an attitude only if it reflects her evaluative judgments. An alternative way that Smith articulates the conditions on answerability for attitudes reveals a prima facie tension with this claim: "In order for a creature to be responsible for an attitude, on the rational relations view, it must be the kind of state that is open, in principle, to revision or modification through the creature's own processes of rational reflection" (2005, p. 256). This articulation of a necessary condition on answerability places emphasis on the Scanlonian notion that the attitudes that are attributable to us are those that are judgment-sensitive. However, it seems quite possible for someone to have an attitude that reflects an evaluative judgment without being able to revise or modify the attitude (or judgment) upon reflection. For example, a person may hold an evaluative judgment that it is good to help people in need and an attitude that reflects this judgment: a general desire to help people in need. What if the person holds this judgment so strongly that no amount of reflection could cause her to give it up? One might be skeptical that anyone is so constituted (though I think this sort of constitution rather commonplace), but it is certainly possible for a person to be so constituted. Let us call such a person Kate. Kate's evaluative judgment (and the attitude that reflects it) is attributable to her in such a way that it serves as a basis of moral appraisal. Kate's attitude reflects on her in a morally positive way; it is the kind of attitude that is partly constitutive of a morally good person.

Can Kate be answerable for her judgment and attitude that reflects it, given that no amount of reflection upon them could cause her to give them up?[8] It depends upon what work the "in principle" clause of Smith's claim is doing. It would help to further clarify Smith's claim that an

[7] Smith 2005 explicitly distinguishes her account from volitionist accounts of responsibility.
[8] The phrase "give them up" may be misleading here. I do not intend it to mean that agents have direct voluntary control over what they value or what attitudes they have. An alternative way of putting this point is that no amount of reflection could cause Kate to cease to have that judgment.

attitude (or judgment) for which a person is answerable must be, in principle, revisable upon reflection. For an attitude *A* of a person *S* (that reflects one of *S*'s evaluative judgments *V*) to be open, in principle, to revision the following conditional must be true:

(1) If *S* were to cease to have *V*, either *A* would be altered or else *S* would cease to have *A*.

Smith holds that we are also answerable for our evaluative judgments. For the purposes of this paper, it will suffice to say that in order for an evaluative judgment *V* to be open, in principle, to revision the following conditional must be true:

(2) If *S* judged that some reason *R* was a sufficient reason to give up *V*, *S* would cease to have because she so judged.[9]

Following Alfred Mele, I will call the kind of evaluative judgment that Kate has an *unsheddable value*. According to Mele, an unsheddable value is one that an agent is practically unable to shed.[10] An agent *S* is practically unable to shed some value *V* during some during some temporal interval *t* when (i) *S*'s psychological constitution precludes his or her voluntarily bringing about the conditions necessary to shed *V* during *t* and (ii) the obtaining of those conditions independently of *S* during *t* is "not in the cards", as it were (Mele 1995, p. 153).

So, while it may be true of Kate that, *if* she judged that some reason was a sufficient reason to give up her judgment that it is good to help people in need, then she would cease to have that judgment, Kate's psychological constitution precludes her from seeing any reason as a sufficient

---

[9] In a personal correspondence Smith confirmed that this adequately characterizes her view on this matter, with a caveat: Smith wants to leave room for cases in which an agent may have a conscious belief that some consideration is a sufficient reason to give up a certain evaluative judgment *V*, but at the same time subconsciously judge that she has sufficient reason to retain *V*. In such cases, Smith maintains that the evaluative judgment may still be open, in principle, to revision upon rational reflection. This is because it may still be the case that, if the subconscious judgment about her reasons were to change, so would *V*. A more adequate characterization of Smith's view would involve the following somewhat more complicated conditional about the conditions on in-principle revisability for evaluative judgments:

(2') If *S* judged that some reason *R* was a sufficient reason to give up *V*, and if *S* did not hold any other judgments (conscious or subconscious) that conflicted with *R*, *S* would cease to have *V* because of *R*.

I use the simpler conditional above, since what I have to say in this paper should not hinge on the difference between the two conditionals.

[10] To "shed" some attitude or value, according to Mele, is to eradicate or significantly attenuate it.

reason for giving up that judgment. The upshot of this discussion is that the rational relations view accommodates the intuition that Kate can be answerable for her unsheddable value and for the attitude that reflects it. This result is integral to the case that I offer in section III.

Before continuing I should say something to help distinguish between evaluative judgments and attitudes. According to Smith, evaluative judgments are "not necessarily consciously held propositional beliefs, but rather tendencies to regard things as having evaluative significance" (2005, p. 251). In this way, evaluative judgments are dispositional; we are disposed to respond or react in particular ways in particular situations (e.g., dispositions to act certain ways, to have certain emotions with respect to certain things, etc.) in virtue of the evaluative judgments that we hold. Among attitudes, Smith includes desires, emotions, and what are commonly called the reactive attitudes (e.g., resentment, indignation, anger, etc.). It may be that attitudes have belief components of one sort or another; a fundamental difference between evaluative judgments and attitudes, though, seems to be that evaluative judgments have a belief component of a special kind; that is, a belief, conscious or conscious, about what one has reasons to do, what things or states of affairs are good, etc. In the remainder of this paper I use "evaluative judgment" interchangeably with "value." By "value," then, I mean to refer to what Smith refers to when she uses the term "evaluative judgment."

II. Smith on Manipulation

One prima facie worry for the rational relations view is that it is vulnerable to manipulation cases. If individuals can be responsible for involuntary psychological states, as Smith holds, then can an individual be responsible for an attitude induced by manipulation? Smith responds to this worry:

> . . . [I]t seems that an attitude 'implanted' by a mad scientist, or one induced through posthypnotic suggestion, would also fail to meet the rational relations condition I have described. . . Since these attitudes are, by hypothesis, detached from a person's own rational assessment, it would be inappropriate to demand that she defend them, or to take them as a basis of rational or moral criticism. They do not really 'belong' to her in a way that would make it possible to draw an inference about the evaluative judgments she accepts (2005, pp. 261-2).[11]

[11] Scanlon 1998 takes a similar line with respect to manipulation cases:

An attitude induced via manipulation would be free-floating, as it were; it would not be grounded in (or reflective of) the person's own beliefs and values. As such, it would not, according to the rational relations view, be something for which the person is answerable. Smith continues to point to a sort of coherence condition on responsibility:

> I see no other way of giving content to the expression 'the agent's own' here, however, except in a way which makes reference to the very network of beliefs and attitudes which I am suggesting ground our attributions of responsibility. . . A reasonable account of the conditions of responsibility should preserve our sense of the rational interrelations among our attitudes . . . (2005, p. 262).

These considerations, along with considerations from the previous section, can be joined to form a more comprehensive condition on answerability. In order for a person *S* to be answerable for an attitude *A*, on Smith's view, the following conditions must obtain: (a) *A* must reflect an evaluative judgment of *S*, (b) *A* must be, in principle, revisable upon *S*'s rational reflection, and (c) *A* must (to some extent) cohere with *S*'s network of beliefs and attitudes.[12]

III. A Counterexample

Recall Smith's (B): a person is blameworthy for Φ (where Φ is some action, attitude, or evaluative judgment) iff she is answerable for Φ and she has violated some moral norm or obligation via Φ. In this section I will offer what I take to be a counterexample to (B). The ingredients for this counterexample correspond to Smith's remarks concerning manipulation. Smith claims that a person is not answerable for an attitude that is induced via manipulation because the attitude will not reflect the agent's evaluative judgments. It's possible, though, that an agent be manipulated to hold certain evaluative judgments. Smith also seems to be committed to a sort of coherence condition on answerability, such that an agent is answerable for some

---

"What distinguishes cases like hypnosis and brain stimulation is thus not that they involve causal influences but rather the fact that these causal influences are of a kind that sever the connection between the action or attitude and the agent's judgments and character. . . This category of excuses might be called 'innocent agent' cases, since in these cases it is claimed that some agent. . .cannot be judged on the basis of the action in question, since it does not reflect that person's judgment-sensitive attitudes" (p. 278).

[12] This third condition is admittedly vague, but it does seem to capture something of what Smith is concerned about in the above passage. I take it that (a) – (c) are individually necessary and jointly sufficient for blameworthiness, on Smith's view.

psychological feature only when it (to some extent) coheres with the person's network of beliefs and attitudes. It is possible, though, to induce (via manipulation) a coherent network of evaluative judgments in a person. Consider the following case.[13]

Jason is a good-natured and caring young man who has devoted his life to charity because he strongly values helping people in need. Jason's values lead him to feel compassion towards those in need (e.g., the sick, the poor) and to desire to do whatever he can to help them. Overnight, a group of evil neuroscientists sneak into Jason's room and tinker with his brain, implanting a new set of values and eradicating any of Jason's pre-existing values (or attitudes) that might conflict with any of the values in the new implanted set.[14] The neuroscientists do not remove or alter any of Jason's memories. Jason wakes up the next morning, quite surprised to find that he no longer values helping people in need. He remembers how much he used to care about helping people, but he finds that he now values "social evolution" and believes that the world would be better without the "weaklings" that he used to care so much about.[15] Jason is not sure how to explain this radical change, and supposes that he must have finally come to realize that people in need are weak and that it is best to leave them to die out. As a result of these new judgments, Jason takes on an apathetic (and sometimes even a hostile) attitude towards people in need. The neuroscientists programmed Jason such that his new evaluative judgments are unsheddable for a year after his programming.[16] Although it is true of Jason that, if he were to see some reason as a sufficient reason for giving up the evaluative judgment that the world would be better without the "weaklings" that

---

[13] The case I present is closely modeled after manipulation cases presented in Mele 2006, but carefully adapted to apply to the rational relations view.

[14] Some of Jason's values will remain unchanged. For example, it may be that Jason's love of such things as french fries and film carry over after the manipulation (of course, what *types* of film Jason values may change quite drastically).

[15] Clearly, the sort of "social evolution" that Jason values may differ in content from what others may think "social evolution" consists in (e.g., the furthering of global humanitarian efforts, etc.).

[16] I want to be careful about what I'm saying the neuroscientists are doing. Recall (i) and (ii) from my above discussion of Mele's usage of unsheddability. The neuroscientists ensure that (i) is satisfied. That (ii) is also satisfied is a stipulation of the case. One might want a more detailed story that explains how neuroscientists ensure that (i) is satisfied. Here is one way to tell such a story. The neuroscientists eradicate any of Jason's original evaluative judgments that might conflict with the new set of implanted evaluative judgments. That being the case, Jason can only reflect on and assess his evaluative judgments by assessing them in the light of other evaluative judgments that he holds. All of the other evaluative judgments that he holds support each other. More specifically, the new network of evaluative judgments includes one central evaluative judgment from which all of the others flow and with respect to which they rationally relate to each other: the judgment that the world would be better without "weak" people (the poor, the sick, the elderly, etc.). This central judgment has such weight that the only way for Jason to give up any of his evaluative judgments is by way of giving up the central judgment, and given the strength of the central judgment (we may assume that the neuroscientists can induce values of varying strengths), no competing reason can be seen by Jason as a sufficient reason for giving up his central evaluative judgment during the stipulated temporal interval.

he used to care so much about, then he would cease to hold that judgment, his psychological constitution precludes him from seeing any reason as a sufficient reason for giving up that judgment.

Smith's view implies that Jason is answerable for his evaluative judgments and the attitudes that reflect them. Jason's apathetic and hostile attitudes towards other people in need reflect his evaluative judgments about social evolution and progress. Jason's having these attitudes is, by hypothesis, *not* detached from his rational assessment.[17] Because of this, Jason is open to demands that he cite the reasons that he takes to justify his attitudes. Jason's attitudes and evaluative judgments are, in principle, open to revision upon rational reflection. Further, his network of beliefs (including his values) and attitudes are coherent: they all fit nicely into a framework that is structured around his central judgment that the world would be better without the "weaklings" that he used to care so much about.

Is Jason blameworthy for his values and attitudes shortly after the manipulation? Would it be appropriate for anyone to be resentful or indignant towards him on the basis of these attitudes and values?[18][19][20][21] There are good reasons to think not. It is true that, *if* Jason were to see some

---

[17] I say this to contrast it with some of Smith's remarks on manipulation that I quote above. My case can also be contrasted with her claim that, in manipulation cases, the implanted attitudes "do not really 'belong' to [the person] in a way that would make it possible to draw an inference about the evaluative judgments she accepts." For, in my case it is not attitudes that are implanted, but evaluative judgments. The attitudes, of course, result from the implanted values. Such attitudes *do* belong to Jason in a way that allows us to draw an inference about the evaluative judgments that he accepts.

[18] I am working with a conception of blameworthiness according to which a person is blameworthy *only if* it would be *pro tanto* appropriate for someone to blame her, where blaming involves having negative reactive attitudes such as anger, resentment, indignation, and, in cases of self-blame, guilt. According to Smith 2007, "active blame," (which goes beyond the mere *judgment* that an agent is blameworthy) involves these reactive attitudes (pp. 476-7). Following Smith 2007, I hold that "one can actively blame a person simply by feeling resentment, indignation, or anger toward her, without ever expressing these emotions in any way"(p. 477). As I mentioned in footnote 4, Smith 2007 uses "culpable" instead of "blameworthy."

[19] It should be noted that Smith's view of blame has evolved in recent years. Smith 2013 understands blame as moral protest. In addition to judging that an agent is blameworthy, to blame another on this view is "to modify one's own attitudes, intentions, and expectations as a way of protesting (i.e., registering and challenging) the moral claim implicit in her conduct…" (p. 43). In this way, Smith's view of blame has become more encompassing.

Smith holds that "the reactive attitudes are not necessary for blame, though they may well capture better than any other reaction the sort of moral protest I think is the crucial element of blame" (p. 41). Though the evolution of Smith's account of blame is both significant and interesting, it should not affect my project here. For, even on Smith's current view, an agent's blameworthiness for his attitudes and actions can still make resentment and indignation appropriate. On Smith's current view, if Jason is blameworthy for his values and attitudes, then it would be appropriate for at least someone to be resentful or indignant towards him on the basis of them. Smith's view implies that Jason is blameworthy, but as I argue here, it would not be appropriate for anyone to have these attitudes towards Jason.

reason as a sufficient reason for giving up his evaluative judgments, he would cease to have them. But there is, by hypothesis, nothing that Jason can experience, do, or think during a year after his programming that would alter his evaluative judgments. Any reflection upon his new values during this period of time will only result in endorsement of them. Furthermore, Jason's values post-brain manipulation are not the product of his values pre-brain manipulation; Jason's rational capacities have been bypassed, leaving him with a set of values and attitudes that are (in a very strong sense) ineliminable.[22] According to Smith's characterization of answerability, Jason is answerable for his evaluative judgments and attitudes, and many of these violate moral norms. However, these considerations strongly suggest that Jason is not blameworthy for them. The result is that (B) is false: being answerable for a norm-violating attitude or evaluative judgment is not sufficient for being blameworthy for it.

IV. Scanlon's Diachronic Condition

One way for Smith to avoid the above counterexample is to strengthen the conditions on answerability. Here I will consider whether a diachronic condition that T.M. Scanlon offers on responsibility might serve as a helpful amendment to Smith's rational relations account:

> Being a rational creature is a matter of having a coherent psychology of a certain kind: of there being the right kind of stable and coherent connections between what one says, does, and how things seem to one at one time, and what one says, does, and how things seem to one at later times. . . (1998, p. 278).

What this diachronic condition amounts to is not so clear. One reason for this is that it is unclear what the "right kind of stable and coherent connections" are. However, Scanlon's claim

---

[20] Because the appropriateness I have in mind is pro tanto in nature, I hold that a person's being blameworthy or culpable for something may be a moral reason for blaming her without its being all-things-considered appropriate to do so, for there may be other moral considerations that count against blaming her. For more on this, see Smith 2007.

[21] A further question here concerns how appropriateness (or lack thereof) should be analyzed. We might understand appropriateness in terms of fairness, such that saying it would be inappropriate to hold such attitudes towards Jason means that it would be unfair to hold them. We might also understand appropriateness in terms of desert, or some other notion. While I will leave the notion of appropriateness unanalyzed here, it does seems right to me to say both that it would be unfair to hold these attitudes towards Jason and that such attitudes are undeserved. The plausibility of both of these claims serve to support the claim that it would be inappropriate for anyone to have these attitudes towards Jason, even if *this* claim is not reducible to either of the other two. For more on how to understand the appropriateness of certain reactive attitudes, see Wallace 1994, pp. 92-109.

[22] Presumably Jason's new values could be altered, but only by bypassing his rational capacities. The work of neuroscientists or of a supernatural being (e.g., God) would do the trick.

seems to imply that Jason is not responsible for the newly acquired values and attitudes since they are not in any way connected (causally or otherwise) with Jason's original set of values and attitudes.[23] Jason came to have his new values and attitudes quite abruptly, and they fail to cohere with many of his pre-manipulation values (e.g., helping those in need). Perhaps, then, the addition of Scanlon's diachronic condition to the other conditions Smith places on answerability would be an improvement to the rational relations view. It does allow Smith to avoid the counterexample presented above. However, a slight revision to the case will be problematic for Smith even if she were to amend her account to include Scanlon's diachronic condition. In a slightly revised version of the case, the neuroscientists program Jason such that his new evaluative judgments are unsheddable for fifteen years after his programming. In this new version of the case Jason satisfies Scanlon's diachronic condition: there is a stable and coherent connection between Jason's values and attitudes directly after the brain manipulation and Jason's values and attitudes fifteen years later. Since Jason is psychologically incapable of ridding himself of any of the implanted values during that time (since they are unsheddable during that interval), the passing of fifteen years should make no difference to whether or not Jason is blameworthy for those values or the attitudes that reflect them. So, if Jason is not blameworthy shortly after the manipulation, then he is not blameworthy fifteen years later.

A defender of Smith's view might maintain that Jason is, in fact, blameworthy after the fifteen years have passed. Here it might be revealing to ask a question of someone who would take up this line: *Is Jason blameworthy for the implanted values (and the attitudes that subsequently reflect them) shortly after the brain manipulation*?[24] I have a difficult time seeing how anyone would think that he is. And, if I am right about this, then it seems that the fifteen-year time lapse is doing some work in eliciting the intuition that Jason is blameworthy after fifteen years have passed. For those who hold this view (i.e., that Jason is not blameworthy shortly after the manipulation but is blameworthy fifteen years later), I offer the following challenge: *In what way does the passing of fifteen years make a difference to whether or not Jason is blameworthy for the implanted values and*

---

[23] I may be misinterpreting Scanlon here. Scanlon may be setting out a condition on being rational that can hold even in cases of manipulation. That is, a person is rational only if, *assuming we hold fixed their psychological structure*, there is "the right kind of stable and coherent connections between what one says, does, and how things seem to one at one time, and what one says, does, and how things seem to one at later times." If this is true, then it may still be true of Jason that he meets this condition. I am grateful to an anonymous reviewer from *Philosophia* for pointing this out. The diachronic condition as I present it, though, does have some intuitive pull; some may think that the reason why massively manipulated agents like Jason are not blameworthy is because the manipulation causes there to be a sharp break from their previous character. For this reason, I think it is worth exploring this idea whether it belongs to Scanlon or not.

[24] I say "shortly after the brain manipulation" because I assume it that it might take some time for the relevant attitudes to result from the implanted values.

the attitudes that reflect them, given that the values are unsheddable during that time? The defender of Smith's view might maintain that Jason is not blameworthy shortly after the manipulation because at that point in time the implanted values do not cohere with the rest of Jason's values and attitudes. That is, Jason is not blameworthy for the implanted values shortly after the manipulation because he is not answerable for them, and he is not answerable for them because the coherence condition on answerability is not satisfied. The time lapse allows for the implanted values (and the attitudes that reflect them) to be incorporated, as it were, into the rest of Jason's network of values and attitudes. Because of this, the defender might explain, Jason is not blameworthy shortly after the manipulation but is blameworthy fifteen years later.

The above response to my challenge is misguided, though. For, by hypothesis, the new values implanted in Jason's brain are implanted as a coherent set, and any of Jason's prior values or attitudes that might conflict with the new set are eliminated by the neuroscientists. Because of this, the implanted values cohere with the rest of Jason's values and attitudes directly after the manipulation. This being the case, the above response cannot support the position that Jason is not blameworthy shortly after the manipulation but is blameworthy fifteen years later. In the absence of a principled reason to support this position it is reasonable to conclude that, if Jason is not blameworthy shortly after the manipulation, then he is not blameworthy fifteen years later.[25]

V. A Worry about Personal Identity

A further objection concerns personal identity. Some may worry that Jason does not survive the brain manipulation; that is, that the resulting person is not personally identical to the pre- manipulation person. That this is so is not obvious. First, all of Jason's memories remain intact through the manipulation. Second, some of Jason's attitudes and values may survive the manipulation; the ones that do not are those that might conflict with the newly implanted set (they are eliminated by the neuroscientists). Either way, though, one who takes up this objection would do well to explain how it relates to the success of the counterexample, for it is unclear how it is an objection to my claim. My claim that the post-manipulation person is not blameworthy does not

[25] This discussion lends plausibility to the claim that Scanlon's diachronic condition collapses into a synchronic coherence condition. I am not entirely sure how to assess this claim.

rely on any claim about his being personally identical to the pre-manipulation person.[26] A successful counterexample to Smith's claim about the conditions on blameworthiness simply requires that Smith's set of jointly sufficient conditions for blameworthiness is satisfied in some possible scenario by a person who is not blameworthy. If "post-manipulation Jason" is answerable for the implanted values (and the attitudes that reflect them), but not blameworthy for them, then the counterexample succeeds.

VI. A Hard Line Reply[27]

Not everyone agrees that globally manipulated agents like Jason are not responsible or blameworthy. Some theorists maintain that, as long as an agent satisfies a certain set of conditions on responsibility (or blameworthiness) he is responsible, manipulation notwithstanding. Harry Frankfurt is perhaps the most well known theorist who takes this position:

> A manipulator may succeed, through his interventions, in providing a person not merely with particular feelings and thoughts but with a new character. That person is then morally responsible for the choices and the conduct to which having this character leads. We are inevitably fashioned and sustained, after all, by circumstances over which we have no control. The causes to which we are subject may also change us radically, without thereby brining it about that we are not morally responsible agents. It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through deliberate manipulative designs of other human agents (2002, 27-28).

For Frankfurt, an agent's being responsible for his actions depends not on how he came to have them, but rather on whether he "identifies with the springs of his actions" (1988, 54). McKenna 2008b has dubbed the sort of reply that Frankfurt offers to manipulation arguments a "hard-line reply." A theorist who offers a hard-line reply to a manipulation maintains that, since the agent satisfies the theorist's preferred conditions on responsibility, the agent is responsible, despite being manipulated to satisfy those conditions.

---

[26] If the post-manipulation person is not personally identical to the pre-manipulation person, then my case is similar to creation or original-design cases, e.g., Cases 1 and 2 of Pereboom's Four-Case Argument Four Views on Free Will, pp. 94-96.

[27] I am grateful to an anonymous reviewer from *Philosophia* for urging me to address the hard-line reply to global manipulation cases.

Although Frankfurt is unmoved by manipulation cases, many theorists hold that (at least some) globally manipulated agents are not responsible or blameworthy. Because of this, a number of compatibilists (including Fischer and Ravizza 1998, Mele 2006, and Haji and Kuypers 2007) embrace what Mele 2006 calls "history-sensitive compatibilism." Their position is that agents in such cases are not morally responsible or blameworthy because they came to acquire certain values or attitudes in the wrong sort of way. These theorists adopt historical conditions on moral responsibility and blameworthiness that are meant to rule out the sort of manipulation that occurs in these cases.

Those who, like Frankfurt, deny that facts about an agent's history are relevant to his or her responsibility (and blameworthiness) may be called non-historical theorists. Smith is one of these. Because the conditions that Smith places on blameworthiness for the non-voluntary are satisfied in the case that I present, it seems that she must choose between two options. First, Smith can choose to take the position that Jason is not blameworthy and adopt historical conditions on blameworthiness for the non-voluntary. Second, Smith can take the hard-line reply, maintaining that Jason is, in fact, blameworthy for his newly acquired values and attitudes since he satisfies the relevant conditions on responsibility (i.e., answerability) and blameworthiness.

I think that Smith should take the first option. The reason that I think this is, of course, because I believe that Jason is not blameworthy for his newly acquired values and attitudes. Some may disagree. At the very least, though, everyone should acknowledge that the claim that Jason is blameworthy is, on the face of it, counterintuitive. Indeed, it has been argued that the hard-line reply involves bullet biting. As Mele 2013 explains it, "biting the bullet" amounts to three things (172). Here I apply these to the non-historical theorist: First, the claim that globally manipulated agents (like Jason) are blameworthy is counterintuitive. Second, the non-historical theorist recognizes that this is so. Last, the non-historical theorist makes this claim (at least partly) because her non-historical position commits herself to it, and not simply because it is the claim that they find most plausible.

McKenna 2004 takes on the task of defending the non-historical position. Although I do not have the space to offer a thorough summary of McKenna's defense, I will point out two ways in which he tries to soften the bullet for the non-historical theorist. First, McKenna hypothesizes that part of our reluctance to treat manipulated agents like Jason as being responsible and blameworthy

may be that we believe that our judging them to be so involves a failure to acknowledge the fact that such agents were seriously wronged in being manipulated. McKenna points out correctly that non-historical theorists can maintain that agents like Jason are responsible and still acknowledge that the he or she was seriously wronged in being manipulated (183-4). Second, McKenna distinguishes between responsibility for acquiring certain attitudes or values and responsibility for having them (183). The non-historical theorist can make a distinction between globally manipulated agents and non-manipulated agents that helps to explain our differing intuitions about them: globally manipulated agents like Jason are not responsible for the *acquisition* of the implanted values or attitudes, while other non-manipulated agents may be. Smith 2008 makes this distinction as well: "I think we would do well to distinguish two different questions: the question of one's responsibility for becoming a certain sort of person, and the question of one's responsibility for the judgments expressed in one's actions and attitudes" (389).

McKenna does not try to argue that the non-historical position is true. Rather, his aim is to show that there are defenses available to the non-historical theorist. McKenna grants that, in certain global manipulation cases, the agents at least *seem* not to be morally responsible (189) and McKenna admits that some considerations "lean in favor of a historical conclusion" (186). Nevertheless, McKenna thinks that the considerations are indecisive, and he remains agnostic about whether there is a historical condition on responsibility and blameworthiness.

Smith maintains that whether or not a person is answerable (and, therefore, blameworthy) for a norm-violating attitude or a value is not a matter of how a person came to have it, but rather a matter of whether the item in question reflects the person's evaluative judgments (2005, p. 267) When confronted with the issue of manipulation, however, Smith doesn't seem to be inclined towards the hard-line reply; i.e., that agents in manipulation cases are responsible and blameworthy. As I explained in section II, Smith holds that the agents in the sorts of manipulation cases she considers don't satisfy the conditions on answerability. However, Smith is responding to a different type of manipulation case, in which the sorts of things that are induced in the agent's psychology are attitudes. Since such attitudes do not reflect any of the agent's evaluative judgments, the agent cannot be answerable (or blameworthy) for them. I have offered a case in which Smith's conditions on answerability and blameworthiness are satisfied. I am not exactly sure how Smith would respond to a manipulation case in which the agent does satisfy the conditions that she places on answerability and blameworthiness.

The case I offer (and other cases like it) helps to motivate a historical condition, since the addition of a historical condition can accommodate the intuitive claim that Jason is not blameworthy for his new values and attitudes. It seems clear that the non-historical views that hard-liners are committed to involve bullet-biting, even if there are ways to soften the bullet a bit. Here I am concerned to point out that non-volitionists like Smith needn't bite the bullet at all; they can accommodate such cases by taking on a historical condition on blameworthiness and/or responsibility, and they can do this without giving up on their non-volitionism. This is why the historical condition(s) I propose in the final section of the paper are suited to non-volitional views. Before we look at these conditions, I want to argue that there is a further reason for Smith and other non-volitionists to adopt historical conditions into their accounts. Without historical conditions, these accounts fail to make sufficient room for the distinction between badness and blameworthiness. In the following section I will argue that this ought to be a welcome distinction.

VII. Bad and Blameworthy Agents

Neil Levy has criticized Scanlonian views of moral responsibility (such as Smith's) on the grounds that they fail to allow for a distinction between bad agents and blameworthy agents. The crux of Levy's criticism can be summed up as follows:

> "Volitionists agree that we can assess agents upon the basis of their morally relevant attitudes, as the attributionists claim. What they deny is that finding that an agent is morally flawed is necessarily to hold that agent responsible for her flaws; that all negative assessment is blame" (2005, p. 6).[28]

Before delving deeper into this issue, it may be helpful to see how Levy's criticism might be framed as a type of open question argument against accounts like Smith's: Even when we regard an agent as answerable for a morally bad attitude or judgment, whether or not she is blameworthy for it seems to remain an open question. Thus, it seems reasonable for someone to say, "Yes, I agree that the attitude (or judgment) is bad, and that it reflects badly on the agent. But, is she *blameworthy* for it?"

[28] In using the term "non-volitional" towards the beginning of the paper, I mean to refer to the same sort of accounts that Levy calls "attributionist."

Although Smith seems to think that it may be possible for an agent to be bad but not blameworthy, she is highly skeptical of this distinction, and argues that we should avoid invoking it whenever possible. Smith offers two reasons in defense of this position:

> (1) Maintaining the distinction between badness and blameworthiness would require us to regard some agents as the passive victims of their faulty judgments.
> (2) To regard a person as morally bad while denying that she is morally blameworthy is to deny her status as a moral agent (2008, pp. 390-1).

Let us first examine (1). Smith claims that maintaining the distinction between bad and blameworthy agents "would require us to regard some agents as the passive victims of their faulty judgments, as I was the passive victim of my faulty hearing." Because answerability is a kind of rational activity, a failure to be answerable for some psychological feature would imply a sort of passivity with respect to it. So, Smith's thought would seem to be that if a person is not blameworthy for her morally bad attitudes (and the morally bad evaluative judgments that they reflect) then she is not answerable for them, either. This thought, though, simply expresses the contrapositive of (and therefore it is logically equivalent to) (B). But if (B) is false, as I have argued, then it cannot be used to support (1). Notice that, in denying (1), I am not thereby committed to the view that anyone who isn't blameworthy for some mental item is merely the passive victim of her faulty judgments or attitudes. If an agent is answerable for something, then surely she is active with respect to it in a way in which she is not active with respect to something like faulty hearing. Indeed, my argument against (B) is precisely an argument to the effect that one can be answerable for some morally bad attitude or evaluative judgment (and therefore active in some way with respect to it) without being blameworthy for it.

Before we examine (2), it would help to consider the case of Robert Harris, which helps to illuminate the point of contention between Levy and Smith. Harris was physically and psychologically abused by both of his parents countless times from a very young age, and was shown little to no love or care from either of them. Over time, Harris became a hardened and cruel individual who carried out a brutal murder without a moment's hesitation. Harris's character was, no doubt, predominantly a result of his terrible upbringing. According to Smith's account, Harris was answerable for his actions and attitudes, since they reflected his own evaluative judgments. And, since his actions and attitudes clearly violated moral norms, Smith's account implies that Harris was blameworthy for them. But, given the influence of his upbringing upon his character

formation, this implication seems counterintuitive. While there is no difficulty in the judgment that Harris was a bad person, this doesn't seem to be the case with the judgment that Harris is blameworthy for who he was and what he did.[29] Smith offers the following response to this worry:

> In such cases, there is clearly a temptation to say, as Levy suggests, that Harris was ''morally flawed'' but that he was not responsible for his flaws. Yet when we consider more carefully the implications of that claim, I think it becomes considerably less appealing. For the ''moral flaw'' in this case was Harris's own judgmental activity, his own evaluation of the weight and significance of the claims presented by others. Thus *Levy's view would commit us to denying Harris responsibility for his own judgments—literally, for what he thinks—which is tantamount to denying him basic status as a moral agent* (emphasis added, 2008, 389).[30]

Here Smith offers us a second reason to avoid the badness-blameworthiness distinction: To regard a person as morally bad while denying that she is morally blameworthy is to deny her status as a moral agent. In order to assess this, we would do well to examine what conditions Smith places on being a moral agent. According to Smith, an individual is a moral agent only if she is answerable for her judgments (and presumably the attitudes and actions that reflect them). We can now understand Smith's reasons for holding (2). Smith is committed to (B): an individual is blameworthy for Φ iff she is both answerable for Φ and she has violated some moral norm or obligation via Φ. On this view, then, if an individual is not blameworthy for her (morally bad) judgments, then she is not answerable for them. And, according to Smith, if an individual is not answerable for her judgments, then she is not a moral agent. If this is correct, and if one's answerability for something that violates some moral norm is conceptually sufficient for being blameworthy for that thing, then denying that a morally bad person is blameworthy *is* to deny her status as a moral agent.

---

[29] It should be noted that the case of Harris is contested. Although some theorists find the view that Harris was blameworthy counterintuitive (e.g., Levy 2005, Wolf 2011), plenty of theorists accept this claim (e.g., Smith 2008, Scanlon 2008). Watson 1987, who originally drew attention to the case of Harris, says that, while we do not suspend our reactive attitudes when we hear of Harris's past, we are ambivalent about how to react to him in light of it (243).

[30] There seems to have been a misunderstanding here. As far as Smith is concerned, all responsibility is answerability. When Levy denies that Harris is responsible, then, Smith takes this to amount to a denial that Harris is answerable. It is important to note, though, that Levy argues that answerability is not an adequate notion of responsibility. Levy's view is that moral responsibility requires a kind of control that answerability does not require, and thus that being answerable for something is not sufficient for being morally responsible for it. So while it might be true that Levy's view would commit him to the claim that Harris is not responsible for what he thinks, Levy's view does not commit him to the claim that Harris is not *answerable* for what he thinks (Levy 2005, p. 5).

Let us grant that answerability for one's judgments is a necessary condition on moral agency. What follows is that (2) is true *if* (B) is true. That is, it's true that to regard a person as morally bad while denying that she is blameworthy is to deny her status as a moral agent *if* answerability for something that violates some moral norm is conceptually sufficient for being blameworthy for that thing. But, as I have argued, (B) is false. Therefore, it cannot be used to support (2).

All of this may be put more simply, though. I have offered a counterexample to (B). One upshot of this is that denying that a person is blameworthy for something needn't involve denying that she is answerable for it. So, whatever kind of activity that is entailed by answerability can be possessed by an agent without that agent's being blameworthy, and whatever kind of moral agency that is entailed by answerability can be possessed by an agent without that agent's being blameworthy. In showing that (B) is false, my case illustrates the importance of the distinction between moral badness and blameworthiness. For, being answerable for something that violates a moral norm is partly constitutive of being a morally bad person. Recall Jason. Jason is callous and cruel; he judges that people in need are weak and that it is best to leave them to die out. This judgment and the attitudes that reflect it are constitutive of a morally bad person. But, although Jason is a morally bad person, he is not blameworthy for the values and attitudes that constitute his moral badness. It would not be appropriate, in principal, for anyone to be indignant with Jason or to resent him on the basis of these attitudes and values.

This point can also be made with respect to a more general consideration about moral criticism. A judgment that a person is answerable for something that violates some moral norm (e.g., a hostile attitude towards other people) is a type of moral criticism. And it may be that this type of moral criticism of a person is made appropriate by the fact that the person's action, attitude, or judgment violates some moral norm and reflects her evaluative judgments. However, it doesn't follow from the appropriateness of *this* type of moral criticism that it would be appropriate to blame a person for the action or attitude in question, nor is it true that every moral criticism necessarily involves blame. To put this point more succinctly, the appropriateness of moral criticism is distinct from blameworthiness, and moral criticism is distinct from blame.

VIII. Towards a Historical Condition for Non-Volitional Accounts

As I mentioned in section VI, Smith maintains that whether or not a person is answerable (and, therefore, blameworthy) for a norm-violating attitude or a value is not a matter of how a person came to have it, but rather a matter of whether the item in question reflects the person's evaluative judgments (2005, p. 267). In order to accommodate the concerns presented in this paper, Smith may have to alter her view on this matter. And, insofar as other views of responsibility and blameworthiness share Smith's position, they may have to do the same. I would suggest that what is missing in these views is a certain historical condition on blameworthiness for attitudes and values. Further, I believe that such a condition is available. The case I have presented in this paper draws attention to the fact that attributions of blameworthiness for attitudes and values are sensitive to facts about how a person comes to have them. What might explain why manipulated agents (like Jason) are not blameworthy for the implanted values and the attitudes that reflect them? We may not need to look much further than what Smith herself says:

> What differentiates implanted attitudes from these others, in my view, is that such attitudes are not based upon the agent's own evaluative appraisal of her situation and surroundings *but are induced in a way that bypasses her rational capacities altogether* (emphasis added, 2005, p. 262).

It is curious that Smith has not endorsed a historical condition on either answerability or blameworthiness for the non-voluntary; as cited above, she explicitly denies that there is such a condition on answerability and blameworthiness. For, the italicized portion of the above passage seems to cite a failure to satisfy a historical condition as an explanation for the fact that a manipulated agent is not answerable for certain attitudes. Notice that what Smith says can be applied to agents like Jason. Jason came to have his new values in a way that bypassed his rational capacities completely. Further, nothing that Jason can experience, do, or think would alter his evaluative judgments (since they are unsheddable). They are not the result of any choice or deliberation on Jason's part, nor did they come about as a result of any other values or attitudes that Jason previously had. These considerations point us in the direction of some possible historical conditions on blameworthiness for attitudes and values. Here I will offer a rough idea of what some such historical conditions might look like.[31]

[31] In this section I offer various conditions on blameworthiness for attitudes and values. A non-volitionist might rather use them as conditions on responsibility, or more specifically, answerability. In doing so, one might just as well avoid the sorts of criticisms I offer in this paper. If so, one could substitute "responsibility" or "answerability" in place of "blameworthiness."

One lesson that might be drawn from the case presented in this paper is that a person is blameworthy for an attitude or a value only if it was not induced in a way that bypassed her rational capacities completely. When an attitude or value is induced in this way, the person's acquiring the attitude or value in question is not, in principle, judgment-sensitive. That is, the process by which the person acquired the attitude or value was not something that could have been preempted or blocked by any reasoning process of the person. Normally, when we acquire new attitudes and values the acquisition of these things is sensitive to and (at least in part) the causal result of other values that we already hold. Further, our coming to have new values and attitudes is usually reflective of other values that we have previously held. This is not so in Jason's case. Given these considerations, one might conclude that Jason is not blameworthy for the relevant values because of the fact that they were implanted in a way that bypassed his rational capacities completely. Consider the following condition on blameworthiness for attitudes and values:

(HB) An agent *S* is blameworthy for an attitude (or value) *A* only if it is <u>not</u> the case that *A* was acquired in a way that bypassed *S*'s rational capacities completely.

Although the fact that Jason's values were acquired in a way that bypassed his rational capacities completely is relevant to Jason's lack of blameworthiness, this view is incomplete. For, it may be that a person can be blameworthy for an attitude or a value of this sort (i.e., an attitude or value that was induced in a way that bypassed a person's rational capacities completely) if it is not unsheddable. If the attitude or value is not unsheddable and persists over some time, it may be that the agent is, after some time, blameworthy for it.[32] So, let's try this again:

(HB') An agent *S* is blameworthy for an attitude (or value) *A* only if it is <u>not</u> the case that (a) *A* was acquired in a way that bypassed *S*'s rational capacities completely <u>and</u> that (b) *A* is unsheddable.

This condition will not yield clear results for certain cases, though. To see this, suppose that an attitude or value is induced in a way that bypasses an agent's rational capacities completely, but is not unsheddable when it is induced. Suppose that, after a time, it becomes unsheddable (partly as a result of the agent's own rational activity). It is not clear to me that the agent is not blameworthy in

[32] Indeed, this is what both Smith 2005 (p. 261) and Scanlon 1998 (pp. 278-9) seem to suggest. If the psychological item (attitude or value) is not unsheddable, the person has the opportunity to, as Smith puts it, "reject or revise [it] in the light of her other beliefs and commitments." If the item persists over some time, it is reasonable to think that the person has allowed it to become incorporated into her psychology in a way that reflects her own rational activity.

this case. However, the above condition does not yield a clear judgment here. The condition is ambiguous about a case like this because it lacks temporal indices. Let's add some, then:

(HB'') An agent *S* is blameworthy for an attitude (or value) *A* at t_1 only if it is <u>not</u> the case that (a) *A* was acquired at some earlier time t_0 in a way that bypassed *S*'s rational capacities completely <u>and</u> that (b) *A* is unsheddable from t_0 (at least) up to and including t_1.[33]

I offer the above condition as a proposal for further consideration. This condition by itself, though, does not rule out Jason's being blameworthy for the attitudes that reflect the implanted values. This is because Jason's manipulators do not implant in him attitudes; rather, they implant values (the attitudes that Jason comes to have are, of course, a virtually inevitable result of the implanted values). So, how can we accommodate the intuition that, even after some time (say, fifteen years), Jason is not blameworthy for the attitudes that reflect the implanted values? The following condition seems to do the supplementary work needed:

(HB$_A$) A person is blameworthy for an attitude only if it reflects an evaluative judgment for which the person is blameworthy.

In this way, Jason is not blameworthy for these attitudes because he is not blameworthy for the corresponding values. The only support I can think of for this principle (besides the fact that it helps explain our intuitions about Jason) is to say that it would be quite odd if a person was blameworthy for an attitude but not blameworthy for the evaluative judgment that it reflects, since a person's attitudes are so closely tied to the evaluative judgments that they reflect. I am not entirely sure how this suggested principle might stand up to scrutiny, but it seems like a helpful place to start.

Although the majority of my discussion in this paper has focused on Smith's account, the additional conditions that I have proposed are tailored to suit non-volitional accounts of

---

[33] Mele 2005 (p. 171) suggests a condition much like this one. It should be added (as Mele himself does) that the bypassing of the person's rational capacities cannot have been set up by the person. Mele also suggests that there may be cases in which a person acquires an unsheddable attitude in a way that bypasses his or her rational reflection and yet still be autonomous with respect to it, owing to the fact that the person has some preexisting attitude or value that supports the induced one. I'm not quite sure what to say about this in relation to my proposal. One reason for this is that it seems there may be different conditions between autonomous possession of an attitude and blameworthiness for an attitude.

blameworthiness more generally. Unlike some volitional historical conditions on blameworthiness, the conditions proposed do not require that the person have exercised voluntary control (e.g., via choices or decisions) over the acquisition of her attitudes or values. Rather, the suggested conditions are expressive of concerns about judgment-sensitivity, i.e., the rational relations that newly acquired attitudes and values must bear to preexisting ones.

Acknowledgements: For helpful comments on this paper, I would like to thank to Randy Clarke, Alfred Mele, Kyle Fritz, Justin Capes, Mirja Perez de Calleja, and an anonymous reviewer from *Philosophia*. I owe a special thanks to Angela Smith both for a beneficial correspondence and for valuable comments on an earlier draft of this paper.

References

Adams, RM (1985) Involuntary Sins. The Philosophical Review 94:3-31.

Darwall, S (2006) The Second-Person Standpoint. Cambridge, MA: Harvard University Press.

Fischer, JM and Ravizza M (1998) Responsibility and Control: An Essay on Moral Responsibility. Cambridge: Cambridge University Press.

Fischer, JM, Kane, R, Pereboom, D and Vargas, M (2007) Four Views on Free Will. Cambridge, MA: Blackwell.

Fischer, J and Tognazzini, N (2011) The Physiognomy of Responsibility. Philosophy and Phenomenological Research 82:381-417.

Frankfurt, H (1988) The Importance of What We Care About. Cambridge: Cambridge University Press.

Frankfurt, H (2002) Reply to John Martin Fischer. In S. Buss and L. Overton (eds.) Contours of Agency (27-31). Cambridge, MA: MIT Press.

Haji, I and Cuypers, S (2007) Magical agents, global induction, and the internalism/externalism debate. Australasian Journal of Philosophy 85 (3):343-371.

Levy, N (2005) The Good the Bad and the Blameworthy. Journal of Ethics and Social Philosophy 1(2):1-16.

McKenna, M (2004) Responsibility and Globally Manipulated Agents. Philosophical Topics 32:169-192.

McKenna, M (2008a) Putting the Lie on the Control Condition for Moral Responsibility. Philosophical Studies 139:29-37.

McKenna, M (2008b) A Hard-line Reply to Pereboom's Four-Case Manipulation Argument. Philosophy and Phenomenological Research 77: 142–159.

McKenna, M (2012) Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism. Journal of Ethics 16:145-174.

Mele, A (2001) Autonomous Agents. New York, NY: Oxford University Press.

Mele, A (2006) Free Will and Luck. New York, NY: Oxford University Press.

Mele, A (2013) Manipulation, Responsibility, and Bullet-Biting. Journal of Ethics 17:167-184.

Scanlon, TM (1998) What We Owe to each Other. Cambridge, MA: Harvard University Press.

Scanlon, TM (2008) Moral Dimensions: Permissibility, Meaning, Blame. Cambridge, MA: Harvard University Press.

Shoemaker, D (2011) Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. Ethics 121:602-632.

Smith, A (2005) Responsibility for Attitudes: Activity and Passivity in Mental Life. Ethics 115:236-271.

Smith, A (2007) On Being Responsible and Holding Responsible. Journal of Ethics 11:465-484.

Smith, A (2008) Control, Responsibility, and Moral Assessment. Philosophical Studies 138:367-392.

Smith, A (2012) Attributability, Answerability, and Accountability: In Defense of a Unified Account. Ethics 122:575-589.

Smith, A (2013) Moral Protest and Moral Blame. In D.J. Coates and N.A. Tognazzini (eds.) Blame: It's Nature and Norms (pp. 27-48). New York, NY: Oxford University Press.

Wallace, RJ (1994) Responsibility and the Moral Sentiments. Cambridge, MA: Harvard University Press.

Watson, G (1987) "Responsibility and the Limits of Evil" in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. Ferdinand Shoeman. Cambridge, MA: Cambridge University Press.

Watson, G (1996) Two Faces of Responsibility. Philosophical Topics 24:227-248.

Wolf, S (2011) Blame, Italian Style. From Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon. New York, NY: Oxford University Press.