THE ELUSIVENESS OF DOXASTIC COMPATIBILISM

Benjamin Bayer

February 16th, 2015

ABSTRACT: This paper evaluates recent proposals for compatibilism about doxastic freedom, and attempts to refine them by applying Fischer and Ravizza's moderate reasons-responsiveness compatibilism to doxastic freedom. I argue, however, that even this refined version of doxastic compatibilism is subject to challenging counter-examples and is more difficult to support than traditional compatibilism about freedom of action. In particular, it is much more difficult to identify convincing examples of the sort Frankfurt proposed to challenge the idea that responsibility requires alternative possibilities.

1. INTRODUCTION

While many moral theorists believe that they need a defensible compatibilism to preserve the concept of moral responsibility in case determinism is true, it is curious how little attention epistemologists have given to the parallel project in their own discipline. Just as assigning moral responsibility is a precondition for praising and blaming agents for their actions, so it seems that assigning epistemic responsibility is a precondition for praising and blaming believers for their beliefs.

The view that an assessment of epistemic justification presupposes an assignment of epistemic or doxastic responsibility is, at the very least, a view cherished by internalist theories of justification (Bonjour 1985, p. 8), and also by recent versions of virtue epistemology (Zagzebski 2001). Although some philosophers are not comfortable with the idea that we choose our beliefs, many find it more natural to say that beliefs can be freely formed, and more still think they can be responsibly formed.[1] Philosophers who take the concept of epistemic responsibility and the prospect of determinism seriously need a form of doxastic compatibilism.

Of course we might dispense entirely with the requirement that knowing subjects need to be epistemically responsible in any significant sense. This paper addresses itself mainly to those who, following the internalists and virtue epistemologists, do think that the assignment of epistemic responsibility is a serious precondition of the attribution of justification and knowledge.[2]

Lately there has been more interest in developing a defensible form of doxastic compatibilism.[3] A recent and especially thorough formulation of the view is offered by Matthias Steup (2000, 2001, 2008,

2011). Steup wants to show that if there is a workable compatibilism for freedom of action or moral responsibility, there should also be a workable version for doxastic freedom or epistemic responsibility. He offers this in response to doxastic involuntarists such as William Alston (1988) and Richard Feldman (2001) who claim that while it is highly plausible to regard our actions as free, it is implausible to regard our beliefs as such. He sees the most plausible version of doxastic compatibilism as one derived from *reasons-responsiveness* compatibilism, as defended most prominently by John Martin Fischer and Michael Ravizza (1998).[4]

Although Steup claims that doxastic compatibilism is just as defensible as standard compatibilism, he does not claim to endorse compatibilism as such. Nonetheless, his formulation of what a workable doxastic compatibilism *might* be is a good starting point for any who would seek to defend the theory. I will argue that in its unqualified form, Steup's formulation faces serious objections that typically apply to unrefined versions of compatibilism. Because Steup's formulation of a reasons-responsiveness compatibilism is tentative and somewhat rough, I will next describe a more refined formulation of the thesis, building on the specifics of Fischer and Ravizza's proposal. But I will contend that this refined version also suffers from significant flaws. After showing that a plausible formulation of doxastic compatibilism is difficult to come by, I will then show that a serious difficulty arises for even the most plausible formulation of the thesis: Frankfurt cases, which lend support to compatibilism by challenging the link between moral responsibility and the ability to do otherwise, turn out to be more elusive for doxastic freedom than for freedom of action.

I am not interested in attacking compatibilism in order to undermine the notion of doxastic freedom or epistemic responsibility. Rather, I sympathize with the view that epistemic responsibility requires doxastic freedom, and that our beliefs are freely formed in a robust way. So while my arguments might be of use to incompatibilist determinists, they might just as easily be used to lay the groundwork for an incompatibilist libertarianism about doxastic freedom.

2. STEUP'S INITIAL FORMULATION OF DOXASTIC COMPATIBILISM

To defend the thesis that compatibilism applies to doxastic freedom just as easily as it does to freedom of action, Steup surveys a variety of formulations of compatibilism for the freedom of action, and then evaluates analogous versions of doxastic compatibilism. The first of these formulations, *classic compatibilism*, is worth noting briefly, as Steup himself does, as problems for it will set the tone for the development of more nuanced versions of compatibilism:

> *Classic compatibilism*
> S's Φing is free iff (i) S Φs, and (ii) S wants to Φ.
>
> *Classic doxastic freedom*
> S's doxastic attitude A toward p is free iff (i) S has attitude A towards p; (ii) S wants to have attitude A toward p.

As Steup notes, the leading problem with the classic formulation of compatibilism about action is that people sometimes do things that, in spite of their wanting to do them, are nonetheless paradigmatically unfree. Standard examples discussed include compulsive behavior such as obsessive hand-washing and other more serious neuroses and psychoses. The same problems would apply to classical *doxastic* freedom: a desire to *believe* something might also be compulsive. This is especially important to note because of the way it helps show the dependence of our account of the freedom of action on a further account of the freedom of underlying psychological states. If our actions are not free unless there is a sense in which we control the desires that motivate them, then a full account of the freedom of action must include an account of what it means to have control over deeper aspects of our psychology. If (as I might argue) it turns out that desires and other emotion are themselves the products of certain beliefs, this points further to the need for a compatibilist account of doxastic freedom.[5]

In his ultimate formulation,[6] Steup considers the compatibilism he takes to be the most plausible, a form of "reasons-responsiveness compatibilism" inspired by Fischer and Ravizza (1998):

> *Reasons-responsiveness compatibilism*
> S's Φing is free iff (i) S Φs; (ii) S wants to Φ; (iii) S's Φing is the causal outcome of a reason-responsive mental mechanism.
>
> *Reasons-responsiveness doxastic freedom*

> S's attitude A toward p is free iff (i) S has attitude A toward p, and (ii) S wants to have attitude A toward p; (iii) S's having taken attitude A toward p is the causal outcome of a reason-responsive mental process.[7]

Very roughly, the formulation contrasts free and responsible action from compulsive behavior in the following way. A free and responsible agent might wash his hands because they are genuinely dirty, exhibiting responsiveness to a well-defined reason (the practical need to maintain hygiene, for example). But a compulsive hand-washer who washes his hands in the absence of evidence of any contamination does not seem to be acting on the basis of a reasons-responsive mechanism.[8] The compulsive hand-washer can be said to have a compulsive *belief* that his hands are contaminated when he will continue to believe this in a wide variety of situations, even when there is no evidence for its truth. We would not hold him responsible for this belief; we would refrain from blaming him for holding what would otherwise be an irrational belief.

Steup modifies his criterion for doxastic freedom somewhat further, but the core of the thesis, clause (iii), remains the same.[9] He points out that for his criteria to be plausible, reasons-responsiveness must be accounted for in terms of the "right kind of reasons," and he admits that explaining exactly what counts as the right kind of reasons-responsiveness is "not exactly an easy project" (2008, p. 380). Some ways of understanding "reasons-responsiveness" are clearly too weak. For instance, the compulsive hand-washer's belief in ubiquitous contamination could be the product of a mechanism that is responsive to accidental psychological "reasons" that do nothing to generate epistemic responsibility. His belief could result from poor self-esteem: were his self-esteem to be boosted, he would lose his compulsion. Perhaps in an attempt to strengthen clause (iii), Steup proposes "responsiveness to the subject's evidence" as the relevant criterion for reasons-responsiveness (2008, p. 380).

I will now argue that *unqualified* responsiveness to evidence is not much of an improvement over an unqualified responsiveness to reasons, and so not a viable criterion for doxastic freedom. To avoid the problem I will shortly describe, doxastic compatibilism needs a criterion in terms of a qualified form of responsiveness to evidence (a criterion to be presented and evaluated in the next section).

Steup cites the absence of "responsiveness to the subject's evidence" to explain why a belief counts as unfree: we can see why it might count as compulsive if one maintains it in a wide range of circumstances, regardless of the evidence. But understanding compulsion via this kind of unresponsiveness to the evidence runs the risk of implying that some cases of obviously *freely formed* but irrational beliefs are in fact *unfree*. Consider: Freely formed but irrational beliefs can exhibit the same kind of persistence as compulsive beliefs can in the face of evidence. A psychologically healthy person can maintain an irrational belief in the importance of constant hand-washing even when evidence suggests (e.g.) that too much hand-washing weakens the immune system and makes a person more likely to get sick. As confirmation holists have illustrated, it is possible to maintain any belief in the face of any evidence, as long as one adjusts one's auxiliary hypotheses accordingly. But it would be a mistake to count all such stubbornly maintained beliefs as unfree or non-responsible. We need to count many of them as free and responsible in order to blame the believer for being irrational. The first problem with Steup's criterion of freedom in terms of unqualified responsiveness to the evidence is that it counts as unfree these responsible but irrational beliefs.

At the same time, a second problem is that Steup's criterion of freedom in terms of unqualified responsiveness to the evidence would also count patently *unfree*, compulsive forms of belief as *free*. There are situations in which compulsive believers could change their beliefs in response to certain kinds of evidence, though these beliefs do not strike us as having the kind of responsiveness to evidence that would make them free or responsible. If, for instance, the compulsive hand-washer were to lose his hands, or be convinced by baseless testimony that all of the germs in the world had been eradicated, he might now believe that he has no need to wash his hands. Perhaps the example of compulsive hand-washing is not even the best example of an unfree action. Perhaps a mild neurosis such as this still counts as free, but strongly irrational. But even paranoid schizophrenics are sensitive to the evidence in some sense. Because John Nash suffered from persecution anxiety manifesting in his adoption of paranoid conspiracy theories, he believed that he was in the presence of one of his persecutors every time he would see a man in a red

tie, but not otherwise.[10] And yet we tend to treat such beliefs as the products of compulsive thinking for which the subject is not responsible, even though they are responsive to evidence in some sense.

Of course, the other conditions on Steup's list might have the effect of protecting his account from the effects of his reliance of an unqualified form of reasons-responsiveness, and thus might guard against the second problem. Perhaps while Nash's belief is responsive to the evidence in some sense, it is not one that he wants to form, and so it would not count as freely formed due to clause (ii). But this does not yet address the first problem. Someone who irrationally insulates his beliefs from refutation by the evidence is intentionally unresponsive to the evidence and so does not fulfill clause (iii): his beliefs would still count as non-responsible by Steup's criterion. In the next section, we will explore a qualified form of reasons-responsiveness that is designed to deal with this problem, among others.

4. Fischer and Ravizza's Moderate Reasons-Responsiveness

Steup never meant to defend reasons-responsiveness in detail, so it is best to turn now to the leading proponents of the view as applied to freedom of action. The problem we have encountered is one that applies to any form of reasons-responsiveness compatibilism. Michael McKenna describes the problem in the form of a dilemma:

> Making the mechanism too responsive to reasons (via strong reasons-responsiveness) sets the bar too high. Those doing moral wrong knowingly would fall short, and hence count as not acting freely *merely by virtue of their wrongful conduct*. But making the mechanism too unresponsive (via weak reasons-responsiveness) allowed a person with only a very limited or insane pattern of sensitivity to reasons to count as satisfying the freedom condition. This set the bar too low. (McKenna 2009)

The typical compatibilist response to this problem has been to attempt to slip between the horns of the dilemma by defining a kind of "moderate" reasons-responsiveness compatibilism, best exemplified in the work of Fischer and Ravizza (1998). Todd and Tognazzini (2008, pp. 685-687) give an apt summary of the two criteria of Fischer and Ravizza's more nuanced definition of moderate reasons-responsiveness:

> *Moderate reasons-responsiveness*
>
> An actually operative kind of mechanism is *moderately reasons-responsive* if and only if
>
> (1) it is at least *regularly receptive to reasons*, some of which are

moral reasons;
(2) it is at least *weakly reactive to reasons* (but not necessarily moral reasons).

An actually operative kind of mechanism is *regularly receptive to reasons* if and only if
(1) There are possible scenarios in which
(i) there is sufficient reason to do otherwise,
(ii) the same kind of mechanism is operative, and
(iii) the agent recognizes the sufficient reason to do otherwise
(2) The possible scenarios described in (1) constitute an *understandable pattern* of reasons-recognition.

An actually operative kind of mechanism is *weakly reactive to reasons* if and only if there is some possible scenario in which
(1) There is sufficient reason to do otherwise;
(2) The same kind of mechanism operates;
(3) The agent recognizes the sufficient reason to do otherwise;
(4) The agent thus chooses and does otherwise for that reason.

Briefly: on this account an act issues from a moderately-reasons responsive mechanism just in case the agent is generally capable of recognizing good reasons for acting otherwise and at least minimally capable of acting on them. (This asymmetry between receptiveness and reactivity is to allow for responsible but unjustified weakness of the will.)

Note that the kind of "doing otherwise" that reasons-responsiveness compatibilism is concerned with here is expressly not an ability to do otherwise *in the exact same situation*. It is crucial to the defensibility of this version of compatibilism to reject the so-called "principle of alternative possibilities" (PAP), the idea that responsibility implies the ability to have done otherwise in the same situation. This principle has been challenged famously by examples from Harry Frankfurt (1969), in which agents seem to be morally responsible for their actions even though their situation is such as to make their acting otherwise impossible. We will explore these examples in more detail later. For now, it is sufficient to note that responsibility defined in terms of moderate reasons-responsiveness is seen by Fischer and Ravizza as defining a form of "guidance control," not "regulative control" (the kind of control presupposed by the PAP).[11]

Applied to freedom of action, moderate reasons-responsiveness helps to avoid the first horn of McKenna's dilemma ("strong reasons-responsiveness") by emphasizing that an agent does not need to act

on an actually good reason to be responsible. She needs merely to be able to act on the basis of a mechanism that is generally capable of recognizing a reason for acting in a contrary way, and which allows her at least occasionally to act on that (good) reason. One can be a responsible but irrational hand-washer, for instance, if one's actions issue from a mechanism that leads one to wash one's hands in some situations when one should not, but that leads one in a range of other circumstances to recognize the detriment of excessive hand-washing, and in at least some of these situations, to refrain from hand-washing for this reason. Thus the criterion of a responsible act is not made coextensive with the actual rationality of the act.

Moderate reasons-responsiveness also helps to avoid the second horn of McKenna's dilemma ("weak reasons-responsiveness") by emphasizing that not just any possibility of responding to a reason to do otherwise makes one responsible: the scenarios in which the agent recognizes these reasons must form an "understandable pattern." So, for example, the fact that there is one situation in which the compulsive hand-washer fails to wash his hands is not enough to make him free or responsible. If, completely arbitrarily, he feels like not washing his hands on Tuesdays, but resumes his compulsiveness in the days that remain, this does not make him any more responsible for his action. There has to be an "understandable pattern" of situations in which he can see a reason to do otherwise.

Can this "moderate reasons-responsiveness" be used as a criterion for doxastic freedom? We would have to select the properly analogous "receptivity" and "reactivity" conditions. The "reactivity" we need to characterize here is not, of course, practical action, but the formation of a belief. Steup suggests that we should think about reasons-responsiveness in terms of receptiveness to evidence, and this is how we will flesh it out. Keeping all of these considerations in mind, we can formulate receptivity and reactivity conditions appropriate to doxastic freedom as follows:

Moderate evidence-responsiveness
An actually operative kind of mechanism is *moderately evidence-responsive* iff
(1) it is at least regularly receptive to evidence.[12]
(2) it is at least weakly reactive to evidence.

An actually operative kind of mechanism is *regularly receptive to evidence* iff
(1) There are possible scenarios in which

(i) there is sufficient evidence to believe otherwise,
(ii) the same kind of mechanism is operative, and
(iii) the agent recognizes the sufficient evidence to believe otherwise

(2) The possible scenarios described in (1) constitute an *understandable pattern* of evidence-recognition.

An actually operative kind of mechanism is *weakly reactive to evidence* iff there is some possible scenario in which

(1) There is sufficient evidence to believe otherwise;
(2) The same kind of mechanism operates;
(3) The agent recognizes the sufficient evidence to believe otherwise;
(4) The agent thus chooses and believes otherwise on the grounds of that evidence.

Let's first see if the moderate evidence-responsiveness criterion avoids the first horn of the dilemma, and allows for the possibility of responsible but irrational belief. Consider Sam's irrational belief in the stereotype "All women are bad drivers." Suppose he forms this initially by generalizing hastily from two women he knows who actually are bad drivers. But he does so in a way that ignores evidence that should tell against his generalization, such as the fact that they are ex-girlfriends who spurned him, and the fact that driving is a learned skill, the quality of which he knows depends on attention and practice, not on one's genetic endowment. In this circumstance, Sam the *agent* does not seem to be especially evidence-responsive, but his belief-forming *mechanism* is better-off. Suppose that he encounters new examples of women who are good drivers: for a while, he may be able to retain his original belief by inventing rationalizations that explain away this evidence. But after a certain point, after he has seen enough obvious cases of genuine female driving excellence, he is no longer able to resist, and abandons the stereotype. His mechanism of belief-formation is still responsive in these other situations, and so he can still be held responsible for his irrational belief in the first situation. So at least on a first pass, moderate evidence-responsiveness does seem to allow for the possibility of responsibly formed but irrational beliefs.

Furthermore, moderate evidence-responsiveness also appears to avoid the second horn of the dilemma, the pitfall of "weak reasons-responsiveness," in that not every alternate situation in which one abandons a belief in response to the evidence shows that one has responsibly formed beliefs. If, for instance, Sam abandons his hasty generalization after observing five separate cases of excellent female

driving, but then adopts the generalization again when he sees a sixth case, a seventh case, and so on, his mechanism, while responsive to evidence, is not responsive in a way that exhibits any "understandable pattern." That's in contrast to a mentally healthier Sam, who sees the fifth case, abandons his generalization, and only strengthens his resolve to reject it as he sees more and more cases of excellent female driving. So again, it seems that moderate evidence-responsiveness does not as obviously count compulsive beliefs as free ones simply because they exhibit *some* pattern of responsiveness to the evidence.[13]

So, a criterion of doxastic freedom expressed in terms of moderate evidence-responsiveness does seem to avoid the more obvious counterexamples associated with the strong and weak evidence-responsiveness criteria. In this regard it is vastly superior to a proposal in terms of utterly unqualified evidence-responsiveness. I would, however, like to press more on whether or not the resulting "moderate responsiveness" is moderate in the right way. I want to first raise questions about how well it deals with the second horn of the dilemma, and then return to what I take to be the harder problem, which is the first horn.

The two horns actually have a way of growing together. To avoid the second problem of weak evidence-responsiveness, the current proposal posits that the situations in which one adopts and abandons a belief in light of the evidence must form "an understandable pattern." The most difficult question here is how to understand "understandable pattern." It is tempting to say that it is the pattern we would follow if we were using something like a reliable cognitive mechanism, one that reliably produces true beliefs. But since reliabilist theories are usually thought to define justified or rational beliefs as ones that result from such a mechanism, we would be back to the problem of the first horn, of being stuck with an overly strong criterion of evidence-responsiveness: we would be requiring that responsible beliefs be rational. But, as already noted, there does seem to be good reason to leave room for the possibility of responsibly formed but irrational beliefs.

Perhaps there is some criterion of an "understandable pattern" cashed out in terms of reliability properties that are not sufficient for justification. (Internalists, of course, have long proposed that

reliability is *never* sufficient for justification.) But then the stakes are especially high for a moderate evidence-responsiveness that deals with the first horn of the dilemma. The Fischer-Ravizza-inspired proposal clearly allows some responsible but irrational beliefs, but does it allow enough of them? We are able to portray stereotyping Sam as a responsible believer because even though as an agent he is not responsive to the evidence in a given circumstance, his cognitive mechanism is still responsive as a whole, a fact that becomes manifest when Sam is bombarded with more and more evidence against his hasty generalization. But what if someone *never* actually gives up a belief, no matter what evidence he or she encounters, treating it as arbitrarily unfalsifiable? On the grounds of some deeply entrenched dogma, Cardinal Bellarmine might forever refuse to believe in heliocentrism, even after looking through Galileo's telescope. A moderate evidence-responsiveness view might take this as a sign that his cognitive mechanism is broken and therefore unresponsive to evidence, and thus that he cannot be held responsible for his beliefs. But we might still insist that someone like Bellarmine can be fully responsible for his belief, and yet manifestly irrational precisely because he *freely* and systematically evades the evidence in this way.

The advocate of moderate evidence-responsiveness might claim that even if Bellarmine's *particular* belief is not responsive to the evidence, it does not follow that his *mechanism* of belief formation is not evidence-responsive. Bellarmine may possess numerous other beliefs on other topics that exhibit sensitivity to evidence over the course of his life. However, it is hard to see how the systematic evasion of evidence in relation to one belief could fail to have implications for the rest of Bellarmine's system of beliefs. Suppose that he is refusing to believe heliocentrism because of a *deeply entrenched dogma*, that the Bible is infallible and must be interpreted literally. If that is the dogma, its generality will affect numerous other beliefs of his, for example, about natural history, geology, physics, law, and morality. He would have to evade evidence on any number of further topics if he continued to abide by this dogma.[14] If he did so unflinchingly through the course of his life (and no doubt, there have been such people), it becomes harder to see how moderate evidence-responsiveness could count his cognitive mechanism as evidence-responsive. We may have the reasonable sense that the problem here is not his

cognitive mechanism, but the agent's misuse of a perfectly good cognitive mechanism. But the question then is what could be the basis for this distinction between the agent and its mechanism *in a theory that is supposed to be compatible with determinism*? If determinism is true, what more is there to the agent than some actual cognitive mechanism and its actual performance? Compatibilists may have an answer to this question, but it escapes me.[15,16]

Perhaps moderate reasons-responsiveness compatibilism can be refined still further to explain how beliefs like Bellarmine's could be responsible but irrational. Or perhaps its advocates could bite the bullet and insist that someone exhibiting such stubbornness in the face of evidence really is the victim of some pathology and so not responsible. I see the first possibility as unlikely, and the second as unpersuasive.

5. THE ELUSIVENESS OF DOXASTIC FRANKFURT CASES

Up to now, I have discussed the difficulties that arise in the attempt to formulate a compatibilist criterion of doxastic freedom or responsibility that is adequately coextensive with what I take to be our ordinary and plausible ascriptions of responsibility. I would now like to discuss the prospects for motivating doxastic compatibilism by appeal to Frankfurt cases. I have already indicated that Frankfurt cases are crucial for the plausibility of compatibilism. These cases threaten the assumption that responsibility implies the ability to do otherwise, a.k.a. the principle of alternative possibilities (PAP). Compatibilists must deny PAP, for if determinism is true, then no agent could ever have done otherwise. Hence, if the doxastic analog to PAP cannot plausibly be denied, doxastic compatibilism faces a serious difficulty.

A typical Frankfurt case described by Fischer and Ravizza (1998) is as follows. Sam has questionable reasons for wanting to kill the mayor, and informs his friend Jack of his plans. Jack also wants to see the mayor dead, and so implants a device in Sam's brain that will detect any hint that he is wavering from his plan: in that event, it will force Sam to kill the mayor anyway (perhaps the device makes him want to kill the mayor after all, in spite of his relenting). We are then asked to compare two

cases. In the first, Sam acts on his questionable reasons and kills the mayor as planned. In the second case, Sam begins to relent, and Jack's device kicks in, forcing Sam to kill the mayor anyway.

Reasons-responsiveness compatibilism attempts to isolate whatever it is about the first case that makes us say Sam is morally responsible, even though he could not have done otherwise. In the first case, it seems we would hold Sam morally responsible for what he has done because, roughly speaking, he acts on his reasons, even though he could not have done otherwise because of the presence of the brain implant. Fischer and Ravizza propose that it is Sam's possession of "*guidance control*" in the first situation that makes his action morally responsible, in contrast to the second situation in which he lacks this control because he is manipulated into acting against his most recent reasoning (his relenting from wanting to kill the mayor). If there were no difference between guidance control and regulative control (the kind of control one is able to exercise by being able to do otherwise, in line with PAP), there would be much less motivation to define a criterion of responsibility in terms of something like *reasons-responsiveness*, which Fischer and Ravizza go on to single out as the best approximation of whatever separates the first case from the second.

Given these considerations, if there turned out to be no genuine cases in which an agent is responsible but not able to do otherwise, then reasons-responsiveness compatibilism would lose a major source of support. I will contend that it is *exceedingly difficult to see how there could be cases of epistemically responsible belief that could not have been otherwise.* Or, at the very least, it is far more difficult to produce such cases than it is to produce cases of morally responsible action that could not have been otherwise. Consequently, evidence-responsiveness doxastic compatibilism loses much of its support and motivation.

Here it is useful to examine a noteworthy objection to Frankfurt cases for freedom of action and compatibilists' typical response. Frankfurt cases must portray an agent who, in one case, is morally responsible for an act that *that very agent* could not have avoided performing, as illustrated by a second case in which that same agent, this time manipulated, performs *the same act* as in the first situation. Obviously, if the manipulated agent does not perform the same act as the unmanipulated agent, if the

manipulated agent's act has a significantly different description, it seems that the agent can do otherwise and the counter-example has no traction against PAP.

There are numerous ways to register this objection. One way is to argue that while the manipulated and unmanipulated agents perform the same action generic action type, they do not perform the same action if actions or events can be individuated more specifically according to their cause.[17] Another is to claim that their actions are not even of the same essential type: agent-causation theorists, for example, maintain that there is an "intrinsic" difference, not just a relational one, between Sam's killing the mayor on his own and Sam's being made to kill the mayor.[18] Fischer has called all such objections that attempt to show that an alternative remains even for the manipulated agent "flicker of freedom" objections (Fischer 1994, p. 134).

Fischer (1994, p. 147; 2003, pp. 241-242) responds to these flicker theorists by claiming that whatever alternative possibility remains in such cases is not a significant or robust alternative possibility, not one that "grounds" a claim to moral responsibility. Fischer's point is that not just any alternative possibility can account for an agent's responsibility. To use a new example to illustrate his point: suppose that quantum mechanics is indeterministic and Sam is in the position of Schrödinger's cat. Whether or not he kills the mayor depends on whether or not a radioactive atom causes the release of poison. If the poison is not released, Sam kills the mayor; if it is, he doesn't kill the mayor, but only because Sam dies. Surely this kind of alternative possibility doesn't account for why Sam is morally responsible for killing the mayor.

Here, I agree with Fischer, but I doubt his reply shows that there are no alternative possibilities that account for Sam's responsibility for the same action.[19] Even still, I do not wish to pursue the debate about what makes an alternative possibility sufficient to generate moral responsibility. I would rather explore the same style of objection with regard to doxastic Frankfurt cases, an issue that few have yet explored.[20] I will claim that Frankfurt cases that undermine PAP for doxastic freedom are elusive; to that extent, so is support for doxastic compatibilism. In essence: When it comes to doxastic freedom, it is

difficult, if not impossible, to dismiss the flicker of freedom response to Frankfurt cases as highlighting an insignificant alternate possibility that fails to ground epistemic responsibility.

The PAP for *doxastic* freedom would presumably state that in order for an agent to be epistemically responsible for a belief, the agent must have been able to *believe* otherwise.[21] So let's consider a doxastic Frankfurt-style case involving an epistemically blameworthy act that is parallel to a morally blameworthy act.

Suppose that in the first case, Sally sees some weak, circumstantial evidence that implicates her nemesis the mayor as being guilty of a crime, and on this basis *believes* that the mayor is guilty. As in the ordinary Frankfurt cases, we now need to hold Sally responsible for the belief in order to blame her for leaping to this conclusion on the basis of sparse evidence. In both this case and the second, Jack wants Sally to believe that her nemesis is guilty of some crime, so in the event that Sally is about to refrain from believing on the basis of this flimsy evidence, Jack has installed a device that will force her to believe in the mayor's guilt anyway.

The traditional compatibilist evaluation of this example would likely say that because Sally is responsible for her belief in the first case, but unable to believe otherwise due to Jack's device, epistemic responsibility does not require the ability to believe otherwise. However, I will argue that this compatibilist evaluation is implausible. Compared to an ordinary Frankfurt case, it is far more plausible that manipulated Sally's *mental actions are significantly different* from unmanipulated Sally's.

Is Sally's mental action the same in the two cases? In Frankfurt cases about moral responsibility and freedom of action, as in the case of Sam the assassin, it is plausible to say that Sam engages in the same physical action, the act for which he is responsible in the unmanipulated case, in both cases. Why, we might ask, should we think of *killing the mayor on one's own* and *being made to kill the mayor* as intrinsically different, as agent-causation flicker theorists assert? It is plausible to think that there is a physical action type that is common to both cases that we can identify as such independent of its cause. And there are very ordinary ways of understanding "same action" on which what Sam is doing in either case is straightforwardly *the same action.*

But the action we're evaluating in doxastic cases is a mental action. It is much less plausible that the action that counts in both cases is the same mental action. In Sam's case, a mental act of relenting one's commitment to kill the mayor is what triggers the device to make him kill the mayor anyway. But in Sally's case, the *relenting*, the triggering event, is already the alternate possibility that really counts. Yes, the device goes on to somehow stop her from relenting. But she has already done something that makes her mental action significantly different from unmanipulated Sally. We can illustrate this in the following rather crude way:

	Unmanipulated Sally	*Manipulated Sally*
t_1	Sally considers scant evidence of the mayor's guilt.	Sally considers scant evidence of the mayor's guilt.
t_2	Sally is incensed by this evidence and allows her critical faculties to be overwhelmed.	Sally is incensed by this evidence but still manages to engage her critical faculties.
t_3	Sally believes that the mayor is guilty.	Jack's device makes Sally believe that the mayor is guilty anyway.

Sally's engaging her critical faculties at t_2 is what triggers Jack's device and makes Sally believe something unjustifiable at t_3. But the difference between the two Sallys' mental states at t_2 is already quite significant. There are philosophers who think it doesn't make sense to say that Sally can choose her beliefs anyway. As far as they're concerned, the only thing epistemology can praise or blame is mental actions like those at t_2.[22] As a mere mental difference, this may seem like a mere flicker of a difference, especially if Sally goes on to kill the mayor in both cases because of her belief. But *in the realm of epistemological evaluation, this flicker of freedom is probably all there is of epistemic significance to evaluate.*[23]

Part of the reason I described the chart above as crude is that there is a serious question to be asked about whether the states described in t_2 and t_3 are being artificially separated. Since in t_1, Sally has already considered the evidence, what Sally is doing in t_2 is recognizing that evidence as relevant or not. But then it is hard to see how she is not already therein believing or not that the mayor is guilty. Some philosophers have noted how asking the question of whether I believe that p is "transparent" to asking whether p (Shah and Velleman 2005). And it is not obvious that we can make sense of the way we see

beliefs as providing reasons for other beliefs on the model of how one event causally explains another in a temporal sequence. Some philosophers now think that belief is sufficiently unique that it may be misleading to describe it as simply an end-state product that is realized instantaneously at the terminus of a series of mental processes. For instance Boyle (2011) suggests that we exercise agency in believing by maintaining a state that persists through time, and Gregory Salmieri and I (2014) have argued that believing as such is an active state, one that is maintained through a series of constitutive cognitive actions (such as the act of mental management). If we are right, it is misleading to parse the epistemically responsible cognitive action from the state of believing itself. Then not only is it the case that there is an epistemically significant flicker of freedom, but that flicker is already the alternative between believing and not believing, and there is no instantaneous sign that can trigger a Frankfurt device to make one act differently--any sign of acting differently will already partially constitute acting differently.[24]

But even if this account of what makes belief unique as a mental state is not true, the fact that it is even tempting to adopt the account already shows that Frankfurt cases supportive of doxastic compatibilism are a much harder sell. Besides, there is already good reason to think that the t_2 alternatives taken by themselves are epistemically significant.

A compatibilist could respond here by insisting that the difference between manipulated Sally's and unmanipulated Sally's thinking at t_2 is not significant enough. Of course, the alternative between the two conditions at t_2 does not look like a mere accident. It's not like the difference generated by a radioactive isotope in the Schrödinger's cat case. But compatibilists may want to impose a higher standard for what makes an alternative robustly relevant to generating responsibility. Fischer (1994, p. 142), for example, has claimed that the kind of alternative that matters to PAP is the alternative over which an agent deliberates. Pereboom (2009) offers a similar criterion, claiming that an agent is responsible only when she knows that she could avoid responsibility by choosing the opposite alternative. For instance, if Sam deliberates between acting to kill the mayor and not doing so, and is aware that he's praiseworthy for the first and blameworthy for the second, then if he goes ahead and kills the mayor anyway, we regard him as responsible for what he does and blame him accordingly.

We can answer the compatibilist objection by rejecting these criteria of robustness, and I think we have good independent reasons for doing this. It is understandable why some might think the alternatives we deliberate over are the kind that PAP presupposes. Libertarians often claim that deliberation is a paradigmatic example of free will. However, we should challenge these libertarians too. To claim that we can be responsible only for an act we could deliberate about is to claim that there is no such thing as epistemic responsibility for a whole category of mental action we ordinarily think is subject to epistemic evaluation. I've already noted that some philosophers think we can't choose a belief by deliberating what to believe, and yet many think we can still epistemically evaluate our beliefs at least in an indirect way in virtue of a more direct evaluation of the cognitive processes used to form them.[25] But none of these cognitive processes are the sorts that typically are deliberated over first, and arguably some couldn't be deliberated over first on pain of regress. And surely we can epistemically evaluate some deliberation as rational and some as irrational. We praise or blame deliberators according to how well they are engaging their critical faculties in so doing. But it can't be that to do that, we have to think that these deliberators either did or could have deliberated about how to deliberate in order to be held responsible in this way. Indeed, we will even evaluate agents for whether they deliberate or not. Failing to deliberate when faced with certain questions is an epistemic vice. One early libertarian, Alexander of Aphrodisias, proposed that it was the choice to deliberate or not that was the fundamental alternative in which our free will consists, and others have proposed similar accounts in terms of an even broader choice to focus one's mind or not.[26]

Perhaps Fischer and Pereboom would retort that they do not mean that a robust alternative needs to be one that we can overtly and consciously deliberate about. Perhaps all that is needed is lower-level sensitivity to the stakes of the alternative. But then I think this kind of awareness surely does exist in the Sally case. People know on some level the important difference between engaging their faculties and not doing so.

6. CONCLUSION

Even if manipulated Sally believes the same as she would when unmanipulated, I have argued that she still exhibits an epistemically significant flicker of freedom, a genuine alternate possibility that helps account for her epistemic responsibility in the unmanipulated cases. It is much easier to make this case for epistemic responsibility than it is for moral responsibility, because there is such an intimate connection between the act for which she is responsible and the object of epistemic evaluation.

I have not offered a knock-down argument against doxastic compatibilism. I have, however, attempted to show that one of the most plausible versions of the view (the moderate reasons-responsiveness theory) is difficult to formulate in a way that respects our ordinary ascriptions of epistemic praise and blame. And I have tried to show that one of the most compelling arguments for compatibilism—the appeal to Frankfurt cases—cannot successfully be deployed. Perhaps a form of compatibilism could be proposed that dispenses with the need for respecting our ordinary ascriptions of epistemic responsibility, and which obviates appeal to Frankfurt examples. But until I see such a proposal, to me, at least, a viable doxastic compatibilism remains elusive.

Loyola University New Orleans

WORKS CITED

Alston, W. (1988). The Deontological Conception of Epistemic Justification. *Philosophical Perspectives, 2*, 257-299.

Arnold, M. (1960). *Emotion and Personality*. New York : Columbia University Press.

Audi, R. (2001). Doxastic Voluntarism and the Ethics of Belief. In M. Steup (Ed.), *Knowledge, Truth and Duty* (pp. 93-114). New York: Oxford University Press.

Ayer, A. (1963). *The Concept of a Person.* London: St. Martin's Press.

Ballin, M. (unpublished). Moral Luck and the Scope of Volition.

Bayer, B. (2012). Internalism Empowered: How to Bolster a Theory of Justification with a Direct Realist Theory of Awareness. *Acta Analytica, 27*(4), 383-408.

Bayer, B. (2013). An Evidentialist Account of Epistemic Possibility. Unpublished Manuscript. Retrieved from http://www.benbayer.com/epistemic-possibility.pdf

Beck, A. (1976). *Cognitive Therapy and the Emotional Disorders.* New York: Meridian.

Binswanger, H. (1992). Volition as Cognitive Self-Regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 165-178.

Bonjour, L. (1985). *The structure of empirical knowledge.* Cambridge, MA: Harvard University Press.

Boyle, M. (2011, December). 'Making Up Your Mind' and the Activity of REason. *Philosophers' Imprint, 11*(17), 1-24.

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The Empirical Status of Cognitive-Behavioral Theraphy: A Review of Meta-analyses. *Clinical Psychology Review, 26 (1)*, 17-31.

Cain, J. (2014). A Frankfurt Example to End All Frankfurt Examples. *Philosophia, 42 (1)*, 83-93.

Chisholm, R. (1982). Human Freedom and the Self. In G. Watson (Ed.), *Free Will* (pp. 24-35). New York: Oxford University Press.
Chrisman, M. (2012). The Normative Evaluation of Belief and the Aspectual Classification of Belief and Knowledge Attributions. *Journal of Philosophy, 109*(10), 588-612.
Clifford, W. (1877, May). The Ethics of Belief. *Contemporary Review, 29*, 289-309.
Engel, P. (2009). Epistemic responsibility without epistemic agency. *Philosophical Explorations, 12*(2), 205–219.
Feldman, R. (2001). Voluntary Belief and Epistemic Evaluation. In M. Steup (Ed.), *Knowledge, Truth and Duty* (pp. 77-92). New York: Oxford University Press.
Fischer, J. M. (1994). *The Metaphysics of Free Will.* Oxford, UK: Blackwell.
Fischer, J. M. (2002). Frankfurt-type Examples and Semi-Compatibilism. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 281-308). New York: Oxford University Press.
Fischer, J. M. (2003). Responsibility and Agent-causation. In D. Widerker, & M. McKenna (Eds.), *Moral Responsibility and Alternative Possibilities* (pp. 235-250). Burlington, VT: Ashgate Publishing Limited.
Fischer, J. M., & Ravizza, M. (1998). Responsibility and Control. Cambridge, UK: Cambridge University Press.
Frankfurt, H. (1969, December). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy, 66 (23)*, 828-839.
Heil, J. (1983). Doxastic Agency. *Philosophical Studies, 43 (3)*, 355-364.
Jäger, C. (2004). Epistemic deontology, doxastic voluntarism, and the principle of alternate possibilities. In W. Löffler, & P. Weingartner (Eds.), *Knowledge and belief. Wissen und glauben* (pp. 65–75). Vienna: Öbvahpt.
Lazarus, R. (1991). *Emotion and Adaptation.* New York: Oxford University Press.
Martin, L., & Clore, G. (2009). *Theories of Mood and Cognition: A User's Guidebook.* Mawah, NJ: Lawrence Erlbaum Associates, Inc.
McKenna, M. (2009). Compatibilism: State of the Art. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter edition). Retrieved June 4, 2010, from http://plato.stanford.edu/entries/compatibilism/supplement.html
Nagel, T. (1979). Moral Luck. In *Mortal Questions* (pp. 24-38). New York: Cambridge University Press.
Nasar, S. (1998). *A Beautiful Mind: The Life of Mathematical Genius and Nobel Laureate John Nash.* New York: Simon and Schuster.
Nussbaum, M. (2001). *Upheavals of Thought: The Intelligence of Emotions.* Cambridge, UK: Cambridge University Press.
O'Connor, T. (2000). *Persons and Causes: The Metaphysics of Free Will.* New York: Oxford University Press.
Peels, R. (2013). Does Doxastic Responsibility Entail the Ability to Believe Otherwise? *Synthese, 190*(17), 3651-3669.
Peels, R. (forthcoming). Against Doxastic Compatibilism. *Philosophy and Phenomenological Research*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/phpr.12040/abstract
Pereboom, D. (2009). Further Thoughts about a Frankfurt-Style Argument. *Philosophical Explorations, 12*(2), 109-118.
Rand, Ayn. 1964. "The Objectivist Ethics," in *The Virtue of Selfishness: A New Concept of Egoism* (New York: New American Library), pp. 13-35.
Rowe, W. L. (2003). Alternate Possibilities and Reid's Theory of Agent-Causation. In D. Widerker, & M. McKenna (Eds.), *Moral Responsibility and Alternative Possibilties* (pp. 219-234). Burlington, VT: Ashgate Publishing Limited.
Ryan, S. (2003). Doxastic Compatibilism and the Ethics of Belief. *Philosophical Studies, 114*, 47-79.
Salmieri, G., & Bayer, B. (2014). How We Choose Our Beliefs. *Philosophia, 42*, 41-53.
Shah, N., & Velleman, D. (2005, October). Doxastic Deliberation. *Philosophical Review, 114*(4).
Sharples (Trans.), R. 2007. *Alexander of Aphrodisias on Fate*. London: Duckworth Publishers.
Solomon, R. (1980). Emotions and Choice. In A. Rorty (Ed.), *Explaining Emotions* (pp. 251-81). Los Angeles: University of California Press.
Steup, M. (2000). Doxastic Voluntarism and Epistemic Deontology. *Acta Analytica, 15 (1)*, 25–56.
Steup, M. (2001). Introduction. In M. Steup (Ed.), *Knowledge, Truth and Duty* (pp. 3-20). New York: Oxford University Press.
Steup, M. (2008). Doxastic Freedom. *Synthese, 161 (3)*, 375–392.
Steup, M. (2011). Belief, Voluntariness and Intentionality. *Dialectica, 65*(4), 537-559.
Todd, P., & Tognazinni, N. (2008). A problem for guidance control. *Philosophical Quarterly, 58*(233), 685–692.
van Inwagen, P. (1983). *An Essay on Free Will.* Oxford: Clarendon Press.

Williams, B. (1973). Deciding to Believe. In B. Williams, *Problems of the Self* (pp. 136-151). Cambridge: Cambridge University Press.
Zagzebski, L. (2001). Must Knowers Be Agents? In *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility* (pp. 142–157). New York: Oxford University Press.

NOTES

I would like to thank a number of people who have offered their encouragement and feedback on this paper in the last few years. The Colorado Springs Philosophy Discussion Group read an early version of my paper and offered kind and useful feedback. (One attendee was my current colleague at Loyola University New Orleans, Leonard Kahn.) Both Matthias Steup and John Martin Fischer graciously encouraged me to publish the paper. A number of anonymous reviewers at a series of journals were particularly influential in encouraging me to reconfigure the overall argument and structure of the paper. Anne Briard's moral support in the last year of writing and editing was irreplaceable (she also helped with final proofreading).

[1] Many following the work of Williams (1973) and Alston (1988) reject entirely the idea that we choose our beliefs but maintain that we can still be epistemically responsible. See, for instance, the work of Heil (1983) and Audi (2001) who claim that we are indirectly responsible for our beliefs in virtue of the control we exercise over various intellectual processes, if not over the beliefs themselves. See Steup (2011) for the suggestion that compatibilists should not see an asymmetry between freedom of action and doxastic freedom. But see also Salmieri and Bayer (2014) for an argument in favor of the claim that there is a robust sense in which we choose our beliefs in a manner relevant to epistemic assessment.

Semicompatibilists who follow Fischer and Ravizza (1998) claim that one can formulate a criterion of responsibility without calling it a criterion of freedom. Engel (2009) follows this line of thought and proposes that we can be epistemically responsible, in the sense of being "able to answer the specific kind of reasons which govern the theoretical domain," but that this kind of responsibility does not "involve [a] voluntary act of the will"(206). Engel's proposal is roughly identical to semicompatibilist "reasons-responsiveness" view proposed by Steup below. In challenging Steup (and later, Fischer and Ravizza), I will implicitly challenge Engel.

In this paper, I will often treat the concepts of responsibility and freedom interchangeably. Because I'll try to show that the semicompatibilist account of responsibility fails, I think it is safe to treat the concepts as roughly interchangeable for my purposes, at least in the build-up to their position.

[2] I have defended a robustly internalist theory of justification in "Internalism Empowered."

[3] In addition to Steup, both Ryan (2003) and Jäger (2004) have recently offered brief compatibilist accounts of doxastic freedom. Ryan, however, does not claim to offer anything like a compatibilist *analysis* of doxastic freedom. She only claims that paradigmatic cases of the lack of doxastic freedom involve compulsion (i.e., neurosis or psychosis), and that our beliefs are not normally compulsive. A virtue of Steup's approach is to offer such a positive account. Jäger's proposal is closer to the moderate reasons-responsiveness compatibilism I sketch in section 3 in light of Fischer and Ravizza's views. For a more recent critical discussion of doxastic compatibilism, especially of the contention that it is motivated by Frankfurt examples, see Peels (2013 and forthcoming).

[4] Steup does this even though Fischer and Ravizza, as semicompatibilists, do not themselves offer reasons-responsiveness as a criterion of *freedom*, but only of *responsibility*. As I mention in footnote 1 above, I am following his practice through much of this paper.

[5] One can argue that we control our actions in virtue of control over emotions, which we exercise in virtue of control over our beliefs. Binswanger (1992) argues for such a chain of control (though he thinks that even control over our beliefs depends on a more fundamental form of cognitive management—see also footnote 26). We might understand the mechanics of such a chain of control through various "cognitivist" theories of emotion, which hold that emotions either are a form of cognition or are the form in which we experience our cognitive appraisals or value judgments about facts in the world. The view dates back to Aristotle and the Stoics, but has been defended recently by philosophers such as Solomon (1980) and Nussbaum (2001). A great deal of empirical support can also be found in cognitive psychology (Arnold (1960), Beck (1976), Lazarus (1991), and Martin and Clore (2009)), and in the extraordinary success of cognitive-behavioral therapy, which operates on a cognitivist premise (Butler et. al (2006)).

[6] Steup goes on to consider two more refined versions of compatibilism inspired by Strawson and Frankfurt, but I will pass these by since Steup does not incorporate them prominently into his final proposal, and I agree with his reasons for finding them to be just as implausible as classic compatibilism. Strawsonian "reactive attitude compatibilism," claims essentially that an act or belief is free if it is a fit object for praise or blame. Steup notes that this position tells us nothing about what it is about the act or belief that makes it fit for these attitudes, and so does not *explain* freedom. Frankfurtian "structural compatibilism," is essentially the same as classic compatibilism, except that it adds a third clause: "S's wanting to Φ is in harmony with S's higher-order desires." This does have the effect of ruling out some neurotic and psychotic behaviors or beliefs that we would otherwise not regard as free, in cases where the subject does not *want to want* to engage in these behaviors. But as Steup notes, higher-order desires

might themselves be subject to external influences, such as brainwashing or manipulation, or otherwise subject to neurosis or psychosis.

[7] One might raise objections to clause (ii) of this formulation that I will not pursue here. There might be examples of beliefs one forms against one's wishes that might seem to be formed in a perfectly epistemically responsible manner, for example, learning plot spoilers from reliable testifiers. But I will not pursue this issue further, since my main point of contention will be with the reasons-responsiveness clause (iii). The problems I raise for this clause will apply even in the absence of (ii). Steup, in fact, does not retain (ii) in his ultimate formulation, which is expressed in terms of "weak intentionality," not desire (see footnote 11).

[8] Here I say "seems" because I'm not convinced that "compulsive" hand-washing rises to the level of legitimate compulsion. I will give more plausible examples later, but since this is Steup's example, I'll run with it for the time being for the purpose of illustration for the time being.

[9] His ultimate formulation is as follows: "*Weakly intentional reasons-responsiveness doxastic freedom*: S's attitude A toward p is free iff (i) S has attitude A toward p, and (ii) S's attitude A is weakly intentional; (iii) S's having taken attitude A toward p is the causal outcome of a reason-responsive mental process" (Steup 2008, p. 386). "Weak intentionality," understood in contrast to explicit intentionality, is defined in terms of *non-accidental* causation (i.e., not being caused by cognitive dysfunction, for example) and involving a *pro-attitude* towards the belief (i.e., one endorses one's resulting belief, or is at least comfortable with it).

[10] Nash's biographer claims that the content of his conspiracy theories was "organized, in subtle ways, around coherent themes," with "connections to Nash's life history and his immediate circumstances" (Nasar 1998, p. 325).

[11] For the sake of full disclosure, while it is true that the criteria above define Fischer and Ravizza's account of reasons-responsiveness, there is more to their account of responsibility than reasons-responsiveness. Like Steup, they add at least one other major necessary condition. Free actions also need to be ones for which agents "*take* responsibility," which involve the agent's viewing himself as an agent, as an "apt target for the reactive attitudes" (such as resentment and guilt), which views are based on the agent's own evidence (Fischer and Ravizza, 1998, pp. 238). As we shall see, however, these additional necessary conditions will not overcome the broader problems of compatibilism defined in terms of moderate reasons-responsiveness.

[12] Depending on one's view of evidence, one might need to insert a qualifier here analogous to "some of which are moral reasons," probably like "some of which is veridical perceptual evidence."

[13] Curiously, nothing in the example of Sam's hasty generalization points to an asymmetry between receptivity and reactivity to reasons. It is not yet clear what it would mean to be receptive to evidence, that is, what it would mean to recognize it as a good reason to form a belief, without actually reacting and forming the belief.

[14] This is one of the holistically-themed lessons of Clifford's classic essay, "The Ethics of Belief": "If a belief is not realized immediately in open deeds, it is stored up for the guidance of the future. It goes to make a part of that aggregate of beliefs which is the link between sensation and action at every moment of all our lives, and which is so organized and compacted together that no part of it can be isolated from the rest, but every new addition modifies the structure of the whole. No real belief, however trifling and fragmentary it may seem, is ever truly insignificant; it prepares us to receive more of its like, confirms those which resembled it before, and weakens others; and so gradually it lays a stealthy train in our inmost thoughts, which may someday explode into overt action, and leave its stamp upon our character for ever." (Clifford 1877, pp. 291–292)

[15] Alternately, the advocate of moderate reasons-responsiveness might stress that Bellarmine *would* abandon geocentrism in an alternate situation, provided that he revised some of his auxiliary hypotheses, such as those about the authority of the Bible or the necessity of interpreting it literally. This, however, comes far too close to a classical compatibilist conditional analysis of "S could have done otherwise" as "S would have done otherwise if S had desired otherwise," a proposal that does not have a happy history (McKenna (2009, §3.3) and Chisholm (1982)).

[16] It will not do to respond by citing other possible worlds in which Bellarmine does change his mind. It is not obvious why pointing to nomologically impossible but logically possible alternate realities would help show that Bellarmine's *actual* cognitive mechanism was really reasons-responsive after all. When a moderate reasons-responsiveness says there must be "possible scenarios" in which the subject reacts to the evidence, the possibilities in mind must be nomological possibilities, not logical possibilities. We are trying to assess a believer's responsibility in the actual world, by reference to the actual mechanism of which his beliefs are products, and the point is to be able to do this in an actual world where determinism is true and so every event in his life is nomologically necessitated. Statements about nearby possible worlds may support counterfactuals about Bellarmine, but the evidence-responsiveness of his cognitive mechanism is not a counterfactual property, nor obviously reducible to any counterfactuals—and this is quite by design. Remember that in clause (1) of the receptivity condition, when it speaks of the agent's believing otherwise, it is expressly not about the ability to have believed otherwise in the same situation. That is the very counterfactual that PAP expresses, which compatibilists must reject.

[17] See also Peter van Inwagen's (1983) distinction between *event-particulars* and *event-universals*. The same event-universal, killing the mayor, may be brought about by several different causes. *Killing-the-mayor-because-Sam-has-always-wanted-to* is a different event-particular from *killing-the-mayor-because-a-brain-probe-makes-him-want-to*.

[18] See for instance O'Connor (2000, pp. 82-83) who offers this interpretation in contrast to the claim that between the manipulated and unmanipulated case there is the same action but caused in two different ways. Rowe (2003, pp. 226-227) explains an important consequence: while it may not have been in Sam's power to avoid willing to kill the mayor in the alternate manipulation case, it may have been in Sam's power not to cause the will to kill the mayor.

[19] The case I've just described partially resembles Fischer's modified Frankfurt case in which, rather than installing a device that will force Sam to kill the mayor, Jack arranges so that if it becomes clear that if Sam will relent from his plans, Sam will be

instantaneously killed. Fischer thinks it is clear that Sam's alternate path "does not contain any voluntary behavior by the agent or anything with sufficient 'oomph' to ground moral responsibility" (2003, p. 242).

I use an example in the main text that does not involve a triggering device because it is a much clearer case of the kind of alternative that does not establish moral responsibility, as it is not triggered by relenting, which I would argue is enough to make the alternative robust enough to illustrate one's responsibility. If Sam's relenting from the assassination means that he dies and is therefore unavailable to kill the mayor, it's true that that Sam doesn't intend to die in doing so, but he does intend to avoid killing the mayor, and that's what happens.

Perhaps the worry here is that the accidental relationship between his intention and the result is not enough to give Sam full credit for the result. Indeed Fischer's case resembles a standard case of moral luck of the kind popularized by Nagel (1979). Nonetheless, we may not need to give Sam credit for not killing the mayor to give him credit for something here. Ballin (unpublished manuscript) has proposed a response to Nagel's puzzle that is remarkably similar to my response to Frankfurt's: if we regard an agent's control over his thinking as fundamental to an agent's freedom of action, agents subject to luck still get to be praised or blamed for significant choices that antecede the outcomes of their action.

James Cain (2014) has written a clever paper showing that we can create Frankfurt examples that help demonstrate why the alternative between doing something on one's own and not doing it on one's own can be significant and generate moral responsibility. He gives the example of someone who knows that the Frankfurt device is present and is asked by investigators not to do something he wants on his own, so they can learn how the device works. But he does what he wants anyway, and as a result, the device does not function. Not doing it on his own would have earned their praise, while doing it on his own earns their blame. Now one might respond that even if the difference matters in this case, it doesn't always matter. But I would argue that the kind of alternative I will cite is always significant.

[20] Peels (2013) does apply van Inwagen's distinction to argue that doxastic Frankfurt cases do not refute the PAP, but he does not consider Fischer's response about the significance of responsibility conferred by flickers of freedom.

[21] One may wonder if this implies treating *believing* as an action. Surely there are some differences between beliefs and the sorts of actions to which we normally attribute moral as opposed to epistemic responsibility. The point here is to apply PAP to beliefs *mutatis mutandis*. For present purposes, I think it is sufficient to treat *believing* it as a condition, where "condition" is a term we'll suppose is neutral between "action" and "state." But I think it is too quick to assume that believing is not an action. I have argued with Gregory Salmieri that we should reject the state-action dichotomy in application to belief (2014, pp. 48–50).

Some philosophers who reject the idea that believing is an action and who doubt that doxastic responsibility is analogous to moral responsibility include Boyle (2011) and Chrisman (2012). Boyle argues that we cannot understand the seemingly timeless form of reason explanation we use for beliefs on the model of actions that occur in time, and Chrisman argues that linguistic data supports the idea that "believe" is a stative rather than active verb.

I think that Chrisman's linguistic data is inconclusive for reasons I cannot elaborate on now. I suspect that Boyle's point turns on an equivocation between first-person and third-person uses of "belief." Clearly one cannot give a reason-explanation for what one currently believes by thinking of one's belief as an event that needs causal explanation. But a third-person use of "belief"—which we can apply to ourselves retrospectively—has no such implications. I suspect that the distinction here is related to Moore's paradox. It is paradoxical to say "it is raining but I don't believe it" even though we know that there can be false beliefs. The Moore-paradoxical usage is first-person and conversationally implies that one is making a knowledge claim, but saying that "I once believed that it was raining, but I was wrong" does not involve this first-person usage or this implicature. I would suggest that the kind of "belief" at stake in the assertion that we could have believed otherwise is clearly the third-person usage applied to oneself as well, not the transparent first-person usage that implies knowledge claims or timeless truth conditions.

[22] See Heil (1983) and Audi (2001).

[23] I do not think that the point here is an accident of fact that the first scenario involves a blameworthy belief while the second involves relenting from it. We could just as easily sketch a scenario in which Sally in the first case believes something difficult but rational (perhaps she comes to see evidence of spousal infidelity), and in the alternate scenario she irrationally relents from believing it. In the alternate scenario she does not need to form a fully contradictory belief (that her spouse is innocent of infidelity) in order to be blameworthy. It is enough that she relents from believing what the evidence makes it rational for her to believe.

[24] A separate compatibilist response that I will not develop further is to say that some mental event prior to t_2 is what triggers the device, so that manipulated and unmanipulated Sally no longer face the epistemically significant alternative of engaging or not engaging their critical faculties. But the further back we push the triggering event, say, to some unconscious neural events that are thought to be responsible for the conscious experience of engaging those faculties or not, the closer we come to begging the question in favor of determinism. Frankfurt himself recognizes the need to avoid the assumption that determinism is true in the course of making his argument (1969, p. 835, n3). His thought experiment is supposed to be neutral between determinism and non-determinism, testing only our concept of moral responsibility. But it looks like the thought experiment challenges PAP only on the assumption that a prior sign can be causally responsible for a choice to engage one's critical faculties. There is, of course, an extended debate about whether or not Frankfurt examples used to argue for semi-compatibilism beg the question in this way. (See Fischer 2002, for a good overview of different replies by compatibilists.) Very notably, many of the replies attempt to find Frankfurt examples that make no assumption of determinism. I think the examples that make no assumption leave plenty of room in which agents can still face important epistemic alternatives, and the others are not as free of the assumption as their advocates believe. I can't elaborate why in this space, and this debate is clearly beyond the scope of the present article. For his part,

Fischer's main critique of the "flicker of freedom" problem is that the alternatives evident in the flicker are not significant, and I think I have now shown how this is not the case.

[25] Gregory Salmieri and I think that it is in the very act of initiating acts of inquiry that we choose to form beliefs (Salmieri and Bayer 2014).

[26] For the Alexander reference, see Sharples (2007, pp. 55-56. See also Rand (1964, pp. 20-21); Binswanger (1992; 2014, pp. 321-328); and Salmieri and Bayer (2014).