

The Elusiveness of Doxastic Compatibilism

Benjamin Bayer

June 3, 2010

ABSTRACT: While moral theorists regularly appeal to compatibilist accounts of freedom of action in order to reconcile the concept of moral responsibility with the prospect of determinism, few epistemologists are as concerned to find a workable compatibilist account of the freedom of belief to underwrite the concept of epistemic responsibility. I suggest that, at least for internalists about justification, epistemic responsibility is crucial and so some version of doxastic compatibilism is necessary for those who take the prospect of determinism seriously. In this paper, I examine Matthias Steup's recent attempt to formulate just such a version of doxastic compatibilism, modeled along the lines of traditional proposals for compatibilism about the freedom of action. Even after strengthening Steup's proposal against objections, I argue that doxastic compatibilism faces the same difficulties as does contemporary compatibilism about freedom of action, perhaps even more acutely. Epistemologists must, therefore, either abandon epistemic responsibility (and internalism) or determinism.

1. Introduction

Ethical theorists who are convinced that moral judgments make substantive prescriptive claims usually believe that this is possible regardless of how controversies in metaphysics or the philosophy of action may be settled. They believe, for instance, that we can make claims about what human beings ought and ought not to do, and praise and blame them for their actions on this basis, while we remain neutral about the debate between determinism and libertarianism. These ethicists are confident that some version of compatibilism can reconcile the prospect of determinism with the existence of a type of human freedom needed to affirm moral responsibility for the bulk of human actions.

While most moral theorists are aware that they need a defensible compatibilism to make their notion of moral responsibility consistent with the prospect of determinism, it is curious how little attention epistemologists seem to pay to a parallel need in their discipline. Just as we assign moral responsibility in order to praise and blame agents for various actions, so it seems we also assign epistemic responsibility in order to praise and blame believers for various beliefs. We criticize as unjustified a person's racist or sexist stereotypes, because we typically regard the person as in some sense responsible for those beliefs. But we do not criticize as logically unjustified the ravings of a paranoid schizophrenic; we regard the schizophrenic as sick and inculpable for these thoughts. The view that *epistemic* or *doxastic* responsibility is presupposed by concepts of epistemic justification is, at the very least, a cherished view of internalist theories of justification (Bonjour 1985, pg. 8), and also in recent theories of virtue epistemology (Zagzebski 2001). And typically, to say that we are responsible for our beliefs is to say that our beliefs are in some sense freely adopted.¹ Therefore, if beliefs are the products or objects of freedom, their free adoption can be criticized as either justified or unjustified.

¹ Sometimes, however, this connection between responsibility and voluntariness is challenged. See, for instance, Engel (2009). Engel believes 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." Engel's definition of responsibility in terms of a capacity to respond to reasons, however, is roughly identical to the compatibilist definition of the freedom or voluntariness of belief in terms of "reasons-responsiveness" given by Steup below, so his challenge to the connection between responsibility and freedom is in effect the same as any compatibilist's challenge to the connection between freedom and the rejection of determinism. In challenging Steup, we'll implicitly challenge Engel.

Of course externalists might dispense entirely with the requirement that knowing subjects exhibit doxastic responsibility. This paper addresses itself mainly to those internalists and virtue epistemologists who do take seriously the requirement of doxastic responsibility (and who take externalism's nonchalance about doxastic responsibility as a mark against it). I want to see if there is a version of compatibilism available to the internalist to ground our attributions of doxastic responsibility, a version that can make sense of any freedom we might have in the formation of our beliefs—our *doxastic freedom*—in a way that is consistent with the prospect of determinism.

A recent attempt to make determinism safe for the epistemologist is found in the work of Matthias Steup (2000, 2001, 2008). Most recently, Steup (2008) has argued provocatively against conventional critics of doxastic freedom, such as Richard Feldman (2001) and William Alston (1989), who suggest that there is an asymmetry between our actions and our beliefs, claiming that while it is highly plausible to regard our actions as free, it is implausible to regard our beliefs as such. Whereas we find ourselves making conscious decisions to move our legs, to speak, to eat, we do not and cannot seem to make conscious decisions to form beliefs. Instead, they seem to be forced on us by our sensory perception or by the testimony of others, and we cannot choose to make ourselves to believe propositions wildly at odds with our evidence.

Against these objections, Steup argues that critics of doxastic freedom have hitherto failed to bring to bear the same compatibilist understanding of freedom that has informed the better-examined question of the freedom of action (which I will here call “practical freedom”). Steup contends that there is no compatibilist understanding of practical freedom that would not, at the same time, apply equally well by analogy to doxastic freedom. Just as compatibilists about practical freedom contend that free action can be defined by specifying the *type* of causation by which paradigmatically free actions are brought about (roughly, causation by internal mental states, rather than external compulsion), so Steup thinks that compatibilists about doxastic freedom can define this freedom by specifying the type of mental causation by which paradigmatically free beliefs are brought about.

In this paper I will survey Steup's compelling proposal to see if it does, indeed, secure a compatibilist case for doxastic freedom.² My contention is that it does not. If anything, I believe that the problems faced by compatibilism are even more acute for the freedom of belief than they are for the freedom of action. Unlike Steup's usual opponents, however, I am not interested in making this point in

² Ryan (2003) and Jäger (2004) also offer brief compatibilist accounts of doxastic freedom. Ryan, however, does not claim to offer anything like a compatibilist *analysis* of doxastic freedom. She simply 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. As I will show in section 5, this purely negative account of freedom is not enough to count as genuinely compatibilist. The compatibilist must also offer an account of the positive causation of freely-formed beliefs. Unlike Ryan, Steup offers a positive account. Jäger's proposal is similar to Ryan's, in that he attempts to develop a view of doxastic responsibility that is not subject to Frankfurt-style objections to the “principle of alternate possibilities” (the view that for one's act to be free, one must have been able to do otherwise than one did). To develop this view, Jäger draws on a distinction between “alternative-possibilities control” (elsewhere called “regulative control”) and “guidance control,” where the latter is possible even if the former is not. “Guidance control” means, in effect, being the cause of one's own actions even if one could not have done otherwise. Unlike Ryan, Jäger does offer a positive account of doxastic freedom, and thinks we do have such “guidance control” over our belief. But Jäger does not say much about what this control consists in. Steup does, in an impressively systematic way, drawing on the version of compatibilism about practical freedom (“reasons-responsive compatibilism”) that is designed specifically to give an account of guidance control that is neutral with respect to the principle of alternative possibilities. This is why we will focus on Steup's presentation.

order to undermine the notion of doxastic freedom. I sympathize with the view that epistemic responsibility requires doxastic freedom, and furthermore the view that our beliefs *are* freely formed. The point of my criticism is to upset the usual complacency among normative epistemologists (and perhaps also ethicists) about metaphysics. I do not believe that the debate between determinism and libertarianism is one that epistemologists can or should be neutral about, as I believe that the prospect of determinism is a serious threat, at the very least, to internalistic epistemologists' conceptions of epistemic responsibility.

Of course, Steup himself is not claiming to argue for compatibilism, only for the proposition that "if compatibilism is true, our doxastic attitudes are free" (2008, 390). He acknowledges that compatibilism might be false, and that if it is false, "the reality of doxastic freedom requires that libertarianism be true." Supposing that libertarianism is true, he also claims that there should be no asymmetry between the freedom of our actions and beliefs. Here I agree with him, and for this reason I will close by recommending a form of libertarian doxastic freedom for epistemologists' consideration.

Because I assume that epistemic responsibility requires doxastic freedom, when I argue that we do not have a version of compatibilism that permits the existence of doxastic freedom and the truth of determinism, my conclusion amounts to a simple conditional: If we can be epistemically responsible, then determinism is false. One can, of course, either affirm the antecedent or deny the consequent. A prior commitment to determinism (usually part and parcel of naturalistic epistemologies that find externalism compelling) implies that one should not accept a concept of epistemic responsibility. But if one has a prior commitment to epistemic responsibility (as internalist epistemologists often do), one must reject determinism (and also, perhaps, any form of indeterminism that is equally at odds with doxastic agency).

This internalistic argument from epistemic responsibility to the rejection of determinism is not a new one. Others have previously contended that there is a sense in which determinism is self-refuting on the grounds that if determinism is true, then even one's belief in determinism is determined by antecedent conditions, in which case the determinist cannot assure the objective justification of his own belief in determinism (Jordan 1969).³ Probably a version of this argument goes back at least as far as Kant's *Foundations of the Metaphysics of Morals* (1785/1993, §3, 50).⁴ This argument presupposes that a belief that is determined by antecedent conditions is not a belief that can be epistemically justified (presumably because it is not a belief that can be epistemically responsible). Responding to Jordan's contention that determinism is self-refuting on these grounds, Grünbaum (1971, 309-10) argues that this presupposition may be question-begging, because determinists can hold that it is precisely the causal relationship between one's beliefs and one's awareness of evidence that makes one's belief rational.⁵ Grünbaum's response is essentially the compatibilist's: By explaining rationality (as one might explain freedom) in terms of a specific kind of causality, he appears to be able to make the universe safe for both the truth of determinism and the possibility of a norm of rationality (and in this case, justified belief in determinism). Thus the

³ See also Boyle, Grisez and Tollefson (1972) for a broader survey of arguments that determinism is in some way self-referentially inconsistent.

⁴ See also Paton (1965, 218, 241, 245, 248, 249, 272, 274, 285) for interpretation of Kant's work on this topic.

⁵ Similar replies to the "determinism is self-refuting" charge can be found at least as far back as A.J. Ayer (1963, 266-7).

question that Steup raises about the possibility of compatibilism for doxastic freedom is the same question at issue between these two parties in the more general debate about determinism vs. libertarianism.

So, I believe that assessing the plausibility of doxastic compatibilism has stakes higher than the seemingly modest dispute about the nature and requirements of epistemic responsibility. If we cannot formulate a respectable version of doxastic compatibilism, and if determinism is true, we must either give up the hope for epistemic responsibility and an internalistic, normative epistemology, or decide that we cannot believe rationally in determinism. What's more, if we take it for granted that our beliefs are prominent causes of our actions, then if we cannot make compatibilistic sense of doxastic freedom, we might not be able to make compatibilistic sense of practical freedom, either.

2. Steup's formulation of doxastic compatibilism

Here is the thesis Steup intends to defend:

Compatibilist doxastic freedom

Compatibilism entails that our actions and our doxastic attitudes are mostly free.

To defend this thesis, Steup surveys a variety of formulations of compatibilism for practical freedom, and then looks to see whether our beliefs would count as freely formed through application of analogous applications of 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 Φ ing is free iff (i) S Φ s, 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 this simple formulation of compatibilism is that people sometimes do things which, in spite of their wanting to do them, are nonetheless paradigmatically unfree. Standard examples here include compulsive behavior such as obsessive hand-washing, along with a series of other neuroses and psychoses. This is important to note, not only because more refined versions of compatibilism should seek to exclude such examples from their definition of freedom, but also because it already begins to show the dependence of our account of the freedom of action on a further account of the freedom of belief. The problem with classic compatibilism is that it does not carefully enough specify the particular *sort* of mental cause that issues in free action. Ultimately what the compatibilist about free action is looking for is causation by responsible or freely formed (at minimum, non-neurotic, non-psychotic) belief. The compatibilist about practical freedom *needs* a theory of doxastic compatibilism.

Steup next considers two more refined versions of compatibilism, neither of which I will dwell on since Steup does not incorporate them as prominently into his final proposal as he does a subsequent version. The first version, Strawsonian "reactive attitude compatibilism," claims essentially that an act or

belief is free if it is a fit object for praise or blame. The shortcoming with this position Steup notes is that it 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. The second version, 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 which 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, this does not take us far enough, because higher-order desires might themselves be subject to external influences, such as brainwashing or manipulation, or otherwise subject to neurosis or psychosis.

The final refined version of compatibilism, “reasons-responsiveness compatibilism,” inspired by Fischer and Ravizza (1998), is the view Steup incorporates most prominently into his own proposal for doxastic compatibilism:

Reasons-responsiveness compatibilism

S’s Φ ing is free iff (i) S Φ s; (ii) S wants to Φ ; (iii) S’s Φ ing 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.

To classify compulsive behavior as unfree on the grounds of reasons-responsiveness compatibilism about practical freedom, it’s enough to note that someone who washes his hands because they are genuinely dirty is responsive to a well-defined reason: the practical need to maintain hygiene, for example. But a compulsive hand-washer who washes his hands regardless of the state of his hands—regardless of whether or not he has done anything to seriously contaminate them—is not acting on the basis of a reasons-responsive mechanism. Now consider the corresponding account of doxastic freedom as applied to the same type of example. Presumably what causes the compulsive hand-washer to act in this way is a *belief* that his hands are contaminated. Steup urges that the hand-washer will continue to believe this in a wide variety of situations, even when there is no evidence for its being true—hence he is not acting freely.

Steup notes several preliminary questions about this formulation of compatibilism. First, should it only count responsiveness to *good* reasons, or to reasons of any kind? Steup agrees that a subject must be responding to the “right kind of reasons” (380). He notes that one could, after all, say that the compulsive hand-washer is acting on the basis of a “reason,” broadly construed: his strong desire to wash his hands. Likewise one could say of his belief about contamination that it is the product of a “reason,” broadly construed: a general belief in his own insecurity, for example. Some consideration like this, no doubt, must prompt Steup to stress that to be free, the “right kind of reason” is one that is “responsive to the subject’s evidence” (380).

So if, for example, our hand washer has neither any good evidence to believe that his hands are still contaminated, nor good evidence to believe in his general insecurity, then on this view his beliefs

would not count as freely formed. He believes in the contamination or in his general insecurity in a wide range of circumstances, regardless of the evidence. But unlike his beliefs, most of our beliefs *are* responsive to evidence, suggesting that our beliefs are, for the most part, freely formed. Here Steup gives the example of our belief that we have hands, and of our disbelief that cats are insects, both of which will change depending upon the evidence we encounter.⁶

We should note, however, that one critical question about “structural compatibilism” raised by Steup concerns whether second-order desires might themselves be influenced by certain causes, such as “systematic conditioning and manipulation,” that make one paradigmatically unfree (379). Critics of compatibilism have recognized that the same questions could be raised about the nature or quality of one’s responsiveness to reasons.⁷ Steup admits that explaining exactly what counts as the right kind of reasons-responsiveness is “not exactly an easy project.” Nonetheless he concludes that “reasons-responsiveness matters,” and proposes responsiveness to evidence as the relevant criterion for reasons-responsiveness. In section 3, I will question whether responsiveness to the evidence is a viable criterion for doxastic compatibilism.

In the second half of his paper, Steup turns to answering a number of arguments against the very possibility of doxastic freedom. For example, he addresses an argument from Feldman (2001) rejecting the possibility of doxastic freedom on the grounds that it would require the explicit intention to adopt beliefs with particular contents, and that we rarely if ever form beliefs this way. In response, Steup gives examples of actions he regards as uncontroversially free but unintentional, such as the steps involved in starting one’s car to drive to work (inserting the key, engaging the clutch, shifting into gear, etc.) If actions can be free but unintentional (especially by the standards of compatibilism), so, presumably, can beliefs. Here I will mention briefly that I agree with Steup that it is a mistake to equate the free formation of beliefs with the explicit choice of belief content. But I should mention that there is more to a belief than its content and what more there is might be that in virtue of which we not only freely form but also *choose* our beliefs, which some accounts of doxastic freedom allow for.⁸

Having rejected Feldman’s argument, Steup formulates a more precise version of reasons-responsiveness doxastic compatibilism 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.

“Weak intentionality,” understood in contrast to the explicit intentionality described by Feldman, is defined as 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

⁶ This point helps Steup address his second main question about reasons-responsiveness compatibilism: Should the reasons concerned be merely prudential or moral, or might they be epistemic as well? Because of his conviction that evidence helps determine the right kind of reason, he argues that both practical and epistemic reasons are relevant here.

⁷ See McKenna (2009).

⁸ See Salmieri and Bayer (unpublished).

comfortable with it). A “weak intention,” then, differs from an explicit intention in that a weak intention is not a propositional attitude. In my mind, this modification to clause (ii) more precisely incorporates the relevant aspects of classical and structural (higher-order desires) versions of compatibilism.

Steup goes on to clarify that in claiming free actions or freely formed beliefs to be weakly intentional, this does not mean that the action or belief must be volitionally caused, i.e. *caused by* the weak intention. Against those who would require volitional causation for freedom, Steup questions whether habitual actions which seem to count as free are really caused by pro-attitudes. Pro-attitudes toward an act might occur only after performing the act. Or even if they are prior to or simultaneous with the act, they might not be the cause (386). This point is largely in keeping with Steup’s rationale for rejecting Feldman’s contention that free action requires causation by explicit intentions. Just as acts that can be intentionally caused can also be caused by a habit, so too can acts that can be caused by a mere pro-attitude. So, Steup thinks, we can conclude that volitional causation is not a necessary condition of freedom, and freedom can exist even where there is no volitional causation.

It is interesting to note, at this point, that Steup has defined a concept of doxastic freedom that has become somewhat distant from the concerns that might otherwise motivate internalist epistemologists to examine the topic. On Steup’s view, we may now be able to define a series of epistemic norms that apply only to epistemically responsible agents, agents whose beliefs are formed freely. But these norms will not be able to *guide* agents in their deliberation about which beliefs to form. If there is a phenomenology associated with weakly intentional, freely formed beliefs, then the agent does not necessarily experience it prior to or simultaneous with the belief formation itself. Any phenomenology associated with weak intentionality, such as any associated with the pro-attitude, is possibly only subsequent to the formation of the belief itself. Norms cannot guide the free acts of agents who do not necessarily know when they are engaging in these free acts. So the epistemic responsibility that Steup’s compatibilism affords is already far removed from the epistemic responsibility of the internalist tradition of Descartes and Locke. In the next section, I will suggest that this version of compatibilism is also at odds with our common sense understanding of doxastic freedom and epistemic responsibility. It may be difficult if not impossible for internalistic epistemologists to acknowledge doxastic freedom without becoming libertarians.

3. Difficulties with Steup’s doxastic compatibilism

Steup, of course, does not see his primary task as defending compatibilism, but as showing that it can be used to make sense of doxastic freedom just as well as it can practical freedom. Whatever shortcomings we may see in his preferred version of compatibilism might be said, then, to be beside the point: his point is to argue against treating practical freedom preferentially over doxastic freedom. But I think that doxastic compatibilism faces special challenges not faced by compatibilism about practical freedom. And doxastic compatibilism has to be treated as a special case, because making sense of the freedom of action *depends* on making sense of the freedom of the beliefs that cause free action. So even if Steup’s main purpose was not to argue for compatibilism, he has done us the service of bringing to light an

issue that has been missing from more general discussions of compatibilism for some time now. Seeing the shortcomings of doxastic compatibilism may bring to light the bigger shortcomings of compatibilism about freedom as such.

First let's consider clause (iii) of reasons-responsiveness compatibilism, the requirement that a belief is free only if S's having taken attitude A toward p is the causal outcome of a reasons-responsive mental process. Steup notes that what characterizes an unfree belief, such as the compulsive belief that one's hands are contaminated, is that one maintains this belief in a wide range of circumstances, regardless of the evidence. What makes our beliefs reasons-responsive, then, is responsiveness to the evidence. But recall Steup's question of whether free action or belief needs to be caused by responsiveness to *good* reasons, vs. reasons of any kind. Steup assumes they must be "the right kinds of reasons." Speaking of action, he says that the compulsive desire to wash one's hands might count as a "reason," broadly construed, but acting on that desire is consistent with acting compulsively and hence being unfree. One could say the same about the belief that one's hands are contaminated. So to count these actions or beliefs as unfree, reasons-responsiveness must involve the *right kind of reasons-responsiveness*, which Steup cashes out in terms of openness to the evidence.

The trouble is that an attempt to specify the "right kind of reasons" in terms of openness to the evidence runs the risk of conflating the concept of *freely-formed belief* with that of *justified* or *rational belief*. This conflation is made easier by the fact that the term "responsible" can be used equivocally: sometimes to indicate sanity or control, sometimes to indicate the proper exercise of sanity or control. The further effect of this would be to conflate *justified or irrational belief* with *unfree belief*. Consider: Free but irrational beliefs can exhibit the same kind of persistence in the face of evidence as compulsive beliefs. A psychologically healthy person can believe irrationally in the importance of regular hand-washing in a wide variety of circumstances, even when evidence suggests (say) that too much hand-washing weakens our immune system and makes us more likely to get sick. As confirmation holists have widely illustrated, it is possible to maintain any belief in the face of any evidence, simply by adjusting one's auxiliary hypotheses, and this adjustment can be performed consciously and intelligently. It would be a mistake to call it unfree, even if it is irrational. The point of classifying a belief or action as *free*, at least according to those who see compatibilism as a resource for normative theory, is to allow us to evaluate the belief or action, to prescribe it or proscribe it, praise the agent for it or blame the agent for it. *Blaming*, at least in the realm of belief, includes the possibility of criticizing a belief as irrational or unjustified. If compatibilism holds that a belief is free only if it is responsive to good reasons, i.e., to the evidence, then irrational beliefs that are not responsive to the evidence would not count as free.⁹ The first problem with Steup's account, then, is that it does not allow for the possibility of freely formed but irrational beliefs. It is too narrow, and excludes these from being free when it should not.

⁹ Notice that Kant's view of freedom also makes this conflation. His view is that we rule our lives by reason and moral law only when we achieve autonomy of the will, only when we except ourselves from the phenomenal causal order by abandoning the "heteronomous" principles embodied in hypothetical imperatives. The price that Kant paid to avoid the contradiction of determinism (mentioned in section 1) may have been to eliminate the possibility of criticizing immorality.

Perhaps the “reasons-responsiveness” criterion can be broadened so that not every case of a reasons-responsive belief would count as a fully rational, fully justified belief. One way of making this case is to note that many forms of irrational belief resemble rational arguments in a superficial way. In informal logic, for example, the appeal to unjustified authority (e.g., ancient unsourced scriptures or the opinion of influential people) strongly resembles the appeal to reliable testimony (e.g., newspaper reporting or expert witnesses). In formal logic, affirming the consequent looks a lot like *modus ponens*. The examples can be multiplied. Irrational beliefs are not typically beliefs that come from nowhere. They come from forms of rationalization that gain credibility to the extent that they mimic rational belief-forming processes. The irrational is often parasitical on the rational. Perhaps, then, we can say that there is something common to both rational and irrational belief: both involve a common mechanical framework, as it were. Perhaps this common mechanical framework could be isolated, and any beliefs resulting from it could be characterized as free, whether rational or irrational.

Using this concept of a “common mechanical framework” of reasoning in our criterion would imply that a freely-formed belief is caused by *something that either is or resembles* a reason-responsive process. It does not tell us what is common to both disjuncts. *At best* the causes of both rationally and irrationally-formed beliefs taken together form a loosely-defined resemblance class. Keep in mind that this is also *at best*. My attempt to find some “common mechanical framework” of reasons-responsiveness is already quite a charitable stretch. I gave a few examples of irrationally-formed beliefs that parasitize rational belief-forming processes. But it is not obvious that *every* freely-formed but irrational belief falls into such a framework. What about beliefs formed by wishful-thinking, for instance? What about sheer unwillingness to consider new evidence out of mental laziness? What rational processes do such beliefs mimic? It is not obvious to me. Some irrationality is difficult to explain in terms of anything more basic.

But even if we could identify the mechanical framework common to both good and bad reasoning, using this framework to formulate a criterion of reasons-responsiveness could easily end up classifying patently unfree, compulsive forms of belief as free. It is not as if the compulsive hand-washer is incapable of recognizably human inference patterns (fallacious or not). If, for instance, the compulsive hand-washer were to lose his hands, or be convinced that all of the germs in the world had been eradicated, he might validly infer that he has no need to wash his hands. Perhaps the example of the compulsive hand-washer is not even the best example of an unfree action and belief. Perhaps a mild neurosis such as this still counts as free, but strongly irrational. So we might do better to look at even more paradigmatically unfree actions and beliefs, such as those of paranoid schizophrenics. Even then, however, the extent to which schizophrenics maintain recognizably human inference patterns is remarkable. John Nash’s work in mathematics was coherent and often groundbreaking; and even his paranoid conspiracy theories were thoroughly intellectually constructed. His biographer claims that the content of these conspiracies was “organized, in subtle ways, around coherent themes,” with “connections to Nash’s life history and his immediate circumstances” (Nasar 1998, 325). Fallaciously organized, no doubt, but coherent.

Of course, the fact that reasons-responsiveness is not sufficient for epistemic responsibility does not show that it is not necessary. The other necessary conditions on Steup's list might be needed to narrow the concept of doxastic freedom to the appropriate kind of reasons-responsiveness. In section 5 I will argue that these other conditions will not do the job.

Note also that cashing out reasons-responsiveness in terms of *openness to the evidence* does no better to narrow the criterion appropriately. There are actions and beliefs that seem just as compulsive as those of the compulsive hand-washer, in spite of being conditioned by evidence in specific circumstances. Consider first an example of compulsive *action* conditioned by evidence apprehended in specific circumstances. In the movie *Serenity*, the character River Tam has been programmed to become a killing machine when given specific visual cues that carry subliminal messages (in one case, she sees an animated sequence in an advertisement that has been planted by her programmers). Likewise in *The Manchurian Candidate*, Sergeant Raymond Shaw is programmed by the Communist Chinese to obey any order he hears after having seen a Queen of Diamonds playing card. Perhaps because these examples are fictional, there may not be realistic psychological mechanisms allowing for such visual triggering. But surely if these types of behavior *were* real, we *wouldn't* regard them as free. And even if such radical behavioral conditioning is impossible or irrelevant, if there is any such thing as hypnotic suggestion on the basis of perceptual cues, similar, realistic counterexamples could be found. And even if hypnotic suggestion is completely fantastic, there are surely examples of more realistic neurotic or psychotic behavior that might be triggered by perceptual cues. And, of course, a compulsive hand-washer washes his hands constantly because his hands are always in front of him, but a compulsive who avoids stepping on cracks on the pavement will only do so when he is actually in the perceptual presence of cracked pavement.

If anything, perceptually-triggered compulsive *beliefs* are probably easier to find and more realistic than are perceptually-triggered actions. A man suffering from a fear of heights will think anxious thoughts precisely under the conditions when he sees himself standing in elevated positions. A paranoid schizophrenic suffering from persecution anxiety may believe that he is in the presence of a conspirator every time he sees a man in a red tie (as was, apparently, the case for John Nash). And so on. There can, in fact, be method in one's madness. Compulsive beliefs can be triggered methodically, in the presence of specific evidence, and this does not seem to make them any less compulsive.

4. Refining doxastic compatibilism with moderate reasons-responsiveness, and its difficulties

The problem of finding a criterion of reasons-responsiveness that is neither too exclusive nor too inclusive is also a problem encountered by reasons-responsiveness compatibilism about practical freedom. Michael McKenna explains it this way:

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. Fischer and Ravizza (1998) formulate just such a version of reasons-responsiveness as part of a case for compatibilism about freedom of action. Here I borrow from Todd and Tognazzini’s (2008, 685-7) apt summary of the two criteria of Fischer and Ravizza’s 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.

More briefly: 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. Note that the kind of “doing otherwise” that reasons-responsiveness compatibilism is concerned with here is not an ability to do otherwise *in the same situation*. This version of compatibilism is consciously attempting to side-step a definition of freedom in terms of the so called “principle of alternative possibilities” (that freedom means the ability to have done otherwise), a principle that has been called into question by Frankfurt examples. Reasons-responsiveness compatibilism is concerned with defining a form of “guidance control,” not “regulative control.”

Is moderate reasons-responsiveness, formulated thusly, adequate to overcome the dilemma we posed earlier? Applied to freedom of action, it is supposed to avoid the problem of the criterion’s being too strong or exclusive by emphasizing that an agent does not need to act on an actually recognized good reason to be responsible, she needs merely to act on the basis of a mechanism that is generally *capable* of permitting her to recognize a reason for acting otherwise, and be capable of at least occasionally acting on it. One can be an irrational but responsible hand-washer, for instance, if one relies on a mechanism that recognizes a reason to wash one’s hands a bit too often, but which would lead one in a range of other circumstances to recognize that excessive hand-washing interferes with one’s quality of life, and in at least

some of these situations, one would then refrain from hand-washing because of this reason.¹⁰ Thus the criterion of responsibility is not conflated with actual rationality in a given circumstance.

Moderate reasons-responsiveness also appears to avoid the problem of the criterion's being too weak or inclusive by emphasizing that not just any possibility of recognizing a reason to do otherwise will do; the scenarios in which the agent recognizes these reasons form an "understandable pattern." So, for example, the fact that there is one situation in which the compulsive hand-washer fails to see a reason to wash his hands is not enough to make him free or responsible. If, completely arbitrarily, he feels like not washing his hands one day of the week, but resumes his compulsiveness every day, 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, like the recognition about quality of life in the example above.

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, physical action, but the very "act" of forming a belief. What kind of "reasons" can one then be receptive to which are not already in the form of a belief? The answer has to be some kind of sensory evidence, which (depending on your philosophy of perception) either is already in a conceptual form capable of being affirmed or denied in judgment, or is in a non-conceptual form such that one forms a judgment only by applying concepts to it in the first place. 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* if and only if

- (1) it is at least regularly receptive to perceptual evidence.¹¹
- (2) it is at least weakly reactive to perceptual evidence.

An actually operative kind of mechanism is *regularly receptive to perceptual evidence* if and only if

- (1) There are possible scenarios in which
 - (i) there is sufficient perceptual evidence to believe otherwise,
 - (ii) the same kind of mechanism is operative, and
 - (iii) the agent recognizes the sufficient perceptual evidence to believe 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 perceptual evidence* if and only if there is some possible scenario in which

- (1) There is sufficient perceptual evidence to believe otherwise;
- (2) The same kind of mechanism operates;
- (3) The agent recognizes the sufficient perceptual evidence to believe otherwise;

¹⁰ Here we'll consider the quality of life consideration to be at least a *good* reason, leaving aside questions about whether or not every good reason need be a "moral reason."

¹¹ Depending on one's philosophy of perception, 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." But the version of the philosophy of perception I happen to endorse (a form of direct realism) does not recognize the possibility of non-veridical perceptual evidence, and so I'll leave that aside for the moment.

- (4) The agent thus chooses and believes otherwise on the grounds of that perceptual evidence.

Let's first see if these modified criteria will account for the possibility of a responsible but irrational belief that dangerous contaminants are present. If we think this agent's responsibility for an irrational belief is accounted for by its issuing from a moderately reasons-responsive mechanism, then we suppose that even if he does not recognize the appropriate perceptual evidence in a given situation, there is some other situation in which he will recognize perceptual evidence that would ground a belief that no dangerous contaminants are present. So, suppose that his belief in dangerous contamination issues from a "mechanism" which treats as relevant perceptual evidence offered by health experts who warned of the danger of various surface-borne viruses and bacteria. Suppose that once in the past, the hand-washer saw a microscopic photograph of the surface of a standard American kitchen counter, and was shocked by how many contaminants were present. But suppose that he has not carefully or correctly interpreted that evidence in the present circumstance, and has even refused to consider additional perceptual evidence that would counter the belief that dangerous contaminants are present in his kitchen. Suppose that the same experts who presented him with the original photograph mention that it was of a cleaned counter, and want to show him an additional picture of an uncleaned counter. Their point is to show that some contamination is inevitable, and that the "poison" is in the dose. This new perceptual evidence might be enough to convince the hand-washer that he needn't be so finicky about hand-washing and that he is interpreting the original photograph out of context. If in a different circumstance, the agent would consider this additional perceptual evidence (and there is an "understandable pattern" of such circumstances), and if occasionally he would "act" on this evidence by forming the belief that there is no dangerous level of contamination, then we can say that he is responsible for his current (admittedly) irrational belief, because the more general mechanism it issues from is both regularly receptive to and weakly reactive to perceptual evidence. If we say he is *not* responsible, it is because *there is no understandable pattern* in which he would recognize and react to perceptual evidence to believe otherwise.

But is a criterion like this *really* adequate to the task of distinguishing between recognized cases of responsibility and irresponsibility? Think again about the example of our irrational but responsible hand-washer's belief in the presence of dangerous contamination. Is it realistic to say that for a belief to be responsible but irrational, there must be an "understandable pattern" of cases in which the believer would recognize and react to perceptual evidence? The claim is that an *agent* is responsible if his belief issues from a reasons-responsive mechanism, but the implication seems to be that it is really the *mechanism* that is responsible for rationality or irrationality, not the agent. The idea is that if a responsible mechanism is bombarded with the right kind of perceptual evidence, usually of a greater intensity, then rational beliefs will occasionally be issued, showing that the mechanism in general, abstracted from the agent, is still receptive and reactive to perceptual evidence. If the agent has such a mechanism, he is responsible. But this mechanism-centric account of responsibility is at odds with our ordinary understanding of responsible irrationality, which often admits of no mechanical "understandable pattern." We do not necessarily

consider an agent to be insane or otherwise irresponsible simply because of the absence of an “understandable pattern” of evidential receptivity. An irrational person might recognize or react to perceptual evidence in a haphazard way, or might *never* recognize or react to perceptual evidence that would undermine some cherished belief. Cardinal Bellarmine might refuse to look through Galileo’s telescope under any circumstance, and might excuse what he sees there as hallucination, even if he is forced to look. Indeed we sometimes regard as one hallmark of irrationality the fact that a given belief is held in an “unfalsifiable” manner. The fault here is not a mechanism of reasoning that is insufficiently sensitive to perceptual evidence, but of the *agent* who refuses to be sensitive or appreciative of this evidence. Insofar as moderate reasons-responsiveness, as stated, requires an understandable pattern of recognizing and reacting to perceptual evidence for believing otherwise as a criterion of responsibility, and insofar as there may be no such understandable pattern for many irrational believers, this version of compatibilism would still classify many obviously irrational believers as non-responsible—failing to avoid one horn of the dilemma.

I see two lines of response open to the compatibilist here. They might respond that the receptivity and reactivity criteria do not specify that the agent receiving or reacting to evidence in other situations need not be the agent whose responsibility we are assessing: an agent would still count as responsible as long as there is *another* agent who, using *the same mechanism*, would recognize evidence for an alternate belief and form that belief on its basis. But then we would run into serious difficulties about how to individuate the “same mechanism” in ways that do not presuppose the principle of alternative possibilities that an account of “guidance control” is supposed to obviate. Compatibilists might also respond by biting the bullet and suggesting that we should simply give up looking for a way to praise or blame agents rather than mechanisms. In this case, notice that an epistemology that accepts moderate reasons-responsiveness as its criterion of doxastic responsibility is likely to be very distant from traditional internalism. Epistemic norms would no longer be relevant to evaluating the *agent’s* guidance of *his* beliefs, but relevant only to the evaluation of the agent’s mechanism. This sounds more like a form of reliabilist externalism than internalism.

Notice further this problem of the absence of an “understandable pattern” is a *special* problem facing doxastic compatibilism. A non-compatibilist about doxastic freedom could still accept that an agent who *acts* irrationally or immorally requires some kind of counterfactual “understandable pattern” of reasons-receptivity to be practically free. To characterize an agent as receptive to moral or practical reasons is already to hold as fixed the agent’s set of background beliefs. An irrational agent who fails to recognize a good reason for acting rationally in a given circumstance is one who already believes in the relevance of a bad reason, or who holds other beliefs that would be conducive to believing in the relevance of the bad reason. An agent’s ability to consider a reason for *action* is already a sophisticated intellectual capacity depending on a variety of background beliefs (is the agent a moralist or an amoralist?, an egoist or an altruist?, religious or non-religious?, etc.). Because holding an agent morally responsible presupposes the

attribution of this fixed set of background beliefs, it suggests that there is an “understandable pattern” of reasons-receptivity.

But note: we will only hold the agent responsible for actions issuing from these background beliefs and their resulting conduciveness to his capacity to recognize reasons for actions if we hold the agent to be responsible *for these beliefs*. We do not hold morally responsible the psychotic agent incapable of recognizing reasons against killing innocent people; we put him in an insane asylum.¹² So the attribution of freedom of action already presupposes a base level of doxastic responsibility, and the attribution of doxastic responsibility does not necessarily presuppose any further fixed set of beliefs, or that they have been formed responsibly. As discussed above, a responsible but irrational believer has no fixed mechanism that determines the perceptual conditions that will issue in rational beliefs. So, insofar as an account of practical freedom depends on an account of doxastic freedom, any unique problems for doxastic compatibilism brought to light here will shed light on problems for compatibilism in general.

5. Last attempts to save doxastic compatibilism

Doxastic compatibilists might respond at this point by jettisoning the reasons-responsiveness requirement and focusing on enriching the other criteria in Steup’s definition. They might focus on criterion (ii) in Steup’s original scheme, the one requiring that S’s attitude A be weakly intentional and thus “non-accidental.” This, they might say, already rules out “cognitive dysfunctions” associated with a lack of doxastic responsibility. So responsible belief would simply be the kind that is *not* the result of a cognitive dysfunction.

Note that this is expressly *not* the way Steup presented his case. He interpreted reasons-responsiveness as openness to the evidence precisely in order to exclude such examples as the compulsive hand-washer. In any case, I do think that, in the absence of an independent definition of “non-accidental,” the stipulation that beliefs be non-accidentally formed is extremely *ad hoc*. To say that causation is non-accidental is to say that it is non-accidental *with respect to* some other positive factor, i.e., that a normal causal factor is unpredictably interrupted by another. To characterize an effect as non-accidental, we need to know what “accidental” is, and to know that, we need to know something about the normal cause. What is it? We have already dismissed the epistemologically-oriented “reasons-responsiveness” account as an adequate answer. Are there fields besides epistemology in which we should look for the answer?

Let’s look at some of the science behind paradigmatic compulsion. Consider this description of the causes of schizophrenia:

Schizophrenia appears to be due to a complex interaction between environmental factors and inherited susceptibility genes. For example, the risk of *schizophrenia* is increased in the offspring of pregnant women who experienced viral infections or malnutrition, or complications during delivery. . . . A genetic basis for *schizophrenia* has been demonstrated from multiple findings. . . . [T]win studies suggest that about 80% of the risk for *schizophrenia* is due to inheritance. . . .

¹² See Todd and Tognazzini 2008: 688.

To understand the neurobiological effects of the interactions between environmental factors and susceptibility genes in *schizophrenia*, much research has focused on the working memory functions of the prefrontal cortex. Working memory is the ability to keep in mind and manipulate a limited amount of information to guide thought or behaviour, and is often impaired in subjects with *schizophrenia*. . . . [N]ot only do individuals with *schizophrenia* exhibit impaired performance on [location memory] tasks, but they also fail to show the normal activation of the prefrontal cortex when attempting such tasks, suggesting that the prefrontal cortex may be an important area of brain dysfunction in *schizophrenia*. . . .

Postmortem studies have also shown that in the prefrontal cortex from subjects with *schizophrenia* the number of neurons does not appear to be changed, but there is evidence for significant disruptions in the connectivity between neurons. . . . Since the brain relies on intricate connections between neurons for proper function, disruptions in such connections can have devastating functional consequences and might play critical roles in the clinical features of *schizophrenia* (Konopaske and Lewis, 2007).

The passage above suggests that a combination of environmental and genetic factors converge to physically damage the schizophrenic's prefrontal cortex. The irrational but responsible believer, who has a healthy brain, is *physically* capable of revising his beliefs in a rational way (even in there is no situation in which he truly *would* choose to do so). No amount of sensory evidence may ever prompt Cardinal Bellarmine to revise his geocentrism, but if we perform an autopsy on his brain, we may find nothing wrong with it. An autopsy of the schizophrenic's brain will reveal something else: he is physically incapable of revising his beliefs.

The compatibilist might welcome the neurophysiological basis of cognitive dysfunctions like schizophrenia as the key to an improved definition of compatibilistic freedom. They could claim that a freely-formed belief is one that is non-accidentally formed, with respect to the normal functioning of a healthy brain, where "healthy" is unpacked in terms of naturalistic considerations about the flourishing of the organism. This would serve to exclude patently unfree beliefs. Just as classic compatibilism about the freedom of action is plausible because freedom is understood as *freedom from external compulsion*, doxastic compatibilism's position could be plausible as long as doxastic freedom is understood as *freedom from internal damage*.

In classic compatibilism, however, one is free *from* external compulsion, and therefore free *to* act on one's desires. *Acting on one's desires* is the positive element of the definition. What is the positive element of the present "freedom from damage" definition of doxastic freedom? On what basis are freely-adopted beliefs formed, rather than on the basis of internal "compulsion"? Here it would be helpful to be able to refer to a reasons-responsive mechanism as our positive element, but this would not tell us enough, as we have already seen how irrational but responsible beliefs do not always appear to issue from a reasons-responsive mechanism, even if the agent *has* such a mechanism available.

A zealous compatibilist might insist that there be some proximate cause of freely-formed beliefs, some factor common to both rational and irrational responsible beliefs. Call it the "neurophysiological X-factor." But this strategy misses the usual point of compatibilism. One adopts compatibilism because one is convinced that one's actions or beliefs are free *in some ordinary sense*, even if they are determined by antecedent conditions. A compatibilist wants to be able to reconcile an ordinary concept of "freedom" with

determinism, in order to make room for the possibility of patently ordinary moral or epistemic judgments. The project loses its point if it loses contact with these ordinary concepts. Suppose that determinism is true and there really is some neurophysiological X-factor that antecedes all healthy, freely-formed beliefs. Even if this is true, it is highly implausible to insist that an understanding of this neurophysiological X-factor is contained in the ordinary concept of doxastic freedom as it is implausible that folk psychological concepts consciously refer to or even track neurophysiological factors. Determinists are of course more likely than not to be skeptical of the scientific acceptability of folk psychological concepts, but eliminativism about folk psychology is not the usual strategy of the compatibilist. (Compatibilism about practical freedom relies heavily on belief/desire psychology, for instance.) I know of no folk psychological resources for making sense of the positive component of doxastic freedom, even though there are folk psychological resources for doing the same for the freedom of action. The concept of “reasons-responsiveness” might have been explicable in ordinary folk psychological terms, but we have so far failed to see how this criterion delivers an adequate account of freedom.

Even if the compatibilist could offer a convincing account of the positive element of doxastic freedom, there is an additional problem with any compatibilistic reliance on the negative element, the component that refers to how responsible beliefs are not caused, that they are not caused by cognitive damage. Obviously there are many ways for a brain to be damaged, not all of which involve neurosis or psychoses that generate a lack of freedom. There is damage that might inhibit memory or processing speed or domain-specific cognitive functions, without inhibiting one’s ability to think freely, i.e. to be rational or irrational. The negative component of doxastic compatibilism would need to specify which kind of brain health is relevant to doxastic freedom. I’ve suggested above that brain health could be defined in terms of a description of the healthy functioning of the organism as a whole, but this was, perhaps, too quick. Normally we would suppose that sanity—the ability to form beliefs rationally and dispense with irrationality—is the relevant aspect of an organism’s healthy functioning. But this the concept of this ability either presupposes or just is identical to the very concept of doxastic freedom that we are trying to define. So it is hard to see how any definition of doxastic freedom in terms of the relevant form of health could avoid circularity. Even if the relevant form of health or brain region were described in purely physiological terms this would not involve overt circularity, but it would once again run up against the problem that this is not folk psychological knowledge, and not suitable for a compatibilist explication. It would also run up against the problem that one could only identify these portions of the brain by first identifying schizophrenic symptoms that correlate with them—and understanding them as symptoms again presupposes a prior understanding of mental health, leading to the circularity problem again.

Some might respond that it is a mistake to expect that compatibilism include *any* positive causal element in its account of freedom. Compatibilism is only concerned with showing that freedom is *consistent* with determinism. It need not show the specific form of causal determination of our “responsible” beliefs. A determinist *qua* determinist need not explain how any given event is determined by antecedent events. But a compatibilist determinist does, if we are convinced that the ordinary concept of

“freedom” has a positive element. Especially when we think of it as “responsibility,” it seems unavoidable that it does have a positive element. *Something* is responsible for a belief when we call it responsible. Doxastic freedom defined merely as the absence of a particular kind of cause would also count as “free” any beliefs caused by other out-of-the-ordinary means, such as hypothetical Frankfurt-style manipulation—or beliefs caused by nothing at all. To show that determinism is consistent with responsibility or freedom, we have to say at least enough about what the responsible thing is, to show that it is not the libertarian self that could always do otherwise according to agency theory.

6. Conclusion

Part of the reason that we lack folk psychological resources for making doxastic compatibilism plausible is that there is, in fact, an important asymmetry between practical freedom and doxastic freedom, both in terms of our ordinary understanding of these concepts, and in terms of their situation with respect to each other.

Classic compatibilism begins by defining practical freedom essentially as freedom *from* external compulsion and freedom *to* act on one’s desires. (This is only the starting point for compatibilism about freedom of action, of course—we have already seen the refinements that have to follow it.) What makes this plausible is that virtually everybody is already willing to call this a form of freedom. The only serious dispute is whether the compatibilist errs in conflating metaphysical and social freedom. And even those who think it is *not* acceptable to conflate the two think that social freedom is still *a* precondition for rendering moral judgment against someone. A slave or hostage is not normally held morally culpable for his actions, since he is acting under compulsion. Critics of compatibilism will say that there are *additional* preconditions for the applicability of moral judgment, such as a metaphysical freedom that is distinct from social freedom, but they still agree that social freedom is *one* of the preconditions for this applicability. When we talk about doxastic freedom, however, there is not already some agreed-upon *positive* definition of doxastic freedom in additional folk psychological terms waiting to meet our need for a precondition for epistemic judgment. There is at least the freedom from internal compulsion, the negative element of a possible definition, but there is no agreed-upon *positive* element of the definition, nothing that would be obviously compatible with determinism.

Part of the reason there is no agreed-upon positive element is that unlike freedom of action, for which one can easily point to internal mental processes (beliefs and desires) as the obvious positive causes of free external behavior, once the subject is the freedom of these very internal processes—the freedom of belief—there is no obviously additional realm to point to as the cause of free internal acts. It’s not as if there’s an inner-internal realm that everyone grasps as being the cause of the outer-internal. And there is no agreed-upon subcategory of beliefs and desires which is understood as fundamental to the rest, in virtue of which the others count as being freely caused. To overcome this problem, doxastic compatibilists don’t bother trying to point to an additional third realm. Instead they in effect point *outside* the mental—to the external causes that impinge upon the senses. Of course there is a sense in which this is pointing to a subset

of internal mental processes, since, on this view, it is states of sensory awareness that are the *proximate* causes of our freely-formed beliefs. But since, on their view, beliefs formed on the basis of the senses are passive with respect to their external objects, the upshot is to point to external causes. We have already seen how a criterion of freedom in terms of causation by reason-responsive mechanisms has the consequence of conflating irrational free belief with unfree compulsive belief. But when we think about what motivated compatibilists in the first place to highlight causation by the internal as a criterion for freedom of action, it becomes especially strange that doxastic compatibilists would now point to causation by the *external* as a criterion for doxastic freedom. It's true that external perceptual causation is not the same kind of external force that compatibilists originally sought to distinguish from internal mental causation when identifying a source of practical freedom. Still, it is somehow unsettling and contrary to whatever ordinary understanding of doxastic freedom we have to say that freedom consists in the external world forcing various beliefs on us, something unsettling for reasons similar to the contention that free actions might be forced on us from the outside.

In the above, I have not offered anything like a knock-down argument against the possibility of compatibilism about doxastic freedom. What I have done is to offer counterexamples to Steup's best formulation of it, in order to show that it is not as easy to formulate as his article suggests. I have offered improved formulations which, to me, remain implausible. I have also expressed general skepticism about the availability of the kind of resources that more plausible versions of compatibilism about the freedom of action typically draw upon. Possibly Steup or another philosopher can offer a more persuasive, refined formulation of compatibilism for the freedom of belief-formation. Until such time, I remain skeptical and suspect that we must choose between a libertarian conception of epistemic responsibility, and a deterministic rejection of epistemic responsibility.

One libertarian account of doxastic freedom, Binswanger's (1992), identifies the essence of human freedom in our fundamental ability to raise or lower our level of cognitive awareness. Salmieri and Bayer (unpublished) have shown how this proposal can be made consistent with a sense in which we actively choose our beliefs. The proposal would help to show why it is so difficult to articulate a difference between irrational non-responsiveness to reasons and compulsive non-responsiveness. The only difference is that in the first case, it is the *agent* that is positively responsible for ignoring or evading evidence, and not in the second. There is no further space to explore the plausibility of such a theory. What I hope to have done here is to recommend that internalist epistemologists concerned with epistemic responsibility explore it; it may be their only option. In the current philosophical climate, it is understandable that philosophers should seek to isolate debates in epistemology from debates in the philosophy of action. But at this point, the tool needed to accomplish this isolation, doxastic compatibilism, is elusive.

References

Alston, W. (1989). *Epistemic justification: essays in the theory of knowledge*. Ithaca: Cornell University

- Press.
- Ayer, A.J. (1963). *The concept of a person*. London: St. Martin's Press.
- Binswanger, H. (1992a). 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, J.M., Grisez, G., & Tollefsen, O. (1972). Determinism, freedom, and self-referential arguments. *Review of Metaphysics*, 26 (1), 3–37.
- Engel, P. (2009). Epistemic responsibility without epistemic agency. *Philosophical Explorations*, 12 (2), 205–219.
- Feldman, R. (2001). Voluntary belief and epistemic evaluation. In Steup, M. (Ed.), *Knowledge, truth and duty* (pp. 77–92). New York: Oxford University Press.
- Fischer, J.M. and Ravizza, M. (1998). *Responsibility and Control*. Cambridge: Cambridge University Press.
- Grünbaum, A. (1971). Free will and the laws of human behavior. *American Philosophical Quarterly*, 8 (4), 299–317.
- Jäger, C. (2004). Epistemic deontology, doxastic voluntarism, and the principle of alternate possibilities. In Löffler, W & Weingartner, P. (Eds.), *Knowledge and belief. Wissen und glauben* (pp. 65–75). Wien: Öbvahpt.
- Jordan, J.N. (1969). Determinism's dilemma. *The Review of Metaphysics*, 23 (1), 48–66.
- Kant, I. (1785/1993). *Grounding for the metaphysics of morals*. James W. Ellington, trans. Indianapolis: Hackett Publishing.
- Konopaske, G. and Lewis, D. (2007). Schizophrenia. *Encyclopedia of life sciences*. Online edition, accessed June 1, 2009. Resource document. Wiley Interscience. <http://0-mrw.interscience.wiley.com/emrw/9780470015902/els/article/a0000062/current/html?hd=All,schizophrenia>. Accessed 4 June 2009.
- McKenna, M. (2009). Compatibilism: State of the art. *The In Edward N. Zalta (Ed.). Stanford encyclopedia of philosophy (Winter2009 Edition)*. Resource document. <http://plato.stanford.edu/entries/compatibilism/supplement.html/>. Accessed 4 June 2010.
- Nasar, S. (1998). *A Beautiful mind: The life of mathematical genius and Nobel Laureate John Nash*. New York: Simon and Schuster.
- Paton, H.J. (1965). *The Categorical imperative*. London: Hutchinson.
- Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies*, 114, 47–79.
- Salmieri and Bayer (unpublished). How we choose our beliefs. Resource document. <http://www.benbayer.com/doxastic-voluntarism.pdf>. Accessed 4 June 2010.
- Steup, M. (2000). Doxastic voluntarism and epistemic deontology. *Acta Analytica*, 15, 25–56.
- Steup, M. (2001). Introduction. In Steup, M. (2001), *Knowledge, truth and duty* (pp. 3–20). New York: Oxford University Press.

Steup, M. (2008). Doxastic freedom. *Synthese*, 161, 375–392.

Todd, P and Tognazinni, N. A problem for guidance control. *The Philosophical Quarterly*, 58 (233), 685–92.

Zagzebski, L. (2001). Must knowers be agents? In Fairweather, A. and Zagzebski, L. (Eds.), *Virtue epistemology: essays on epistemic virtue and responsibility* (pp. 142–157). New York: Oxford University Press.