

## **Incompatibilism Refocused: the Case from the Phenomenology of Rational Agency**

Benjamin Bayer

5. August, 2013

*ABSTRACT:* Libertarian incompatibilists are known to argue for their conception of freedom of the will by appealing to introspective awareness of their own agency. However in attempting to articulate how such awareness provides evidence of the ability to do otherwise, these libertarians sometimes suggest that paradigmatically free decisions are “close-call” decisions made as a result of “torn” deliberation. In this paper I argue that libertarians have misidentified the appropriate paradigm cases of free decision, and I recommend refocusing the debate away from our experience of practical decision-making to our experience of choices made in the process of cognitive management—i.e., our introspective awareness of our rational agency. Having sketched this alternative account of the phenomenology of agency, I conclude by examining the compatibilist account of agentic experience offered by Terry Horgan, and suggest it is not the best explanation of all of the relevant evidence about how we experience and make judgments about our freedom.

### **1. Introduction**

Advocates of the “agent causation” approach to libertarian incompatibilism face an uphill rhetorical battle when they attempt to argue for the existence of a form of causality distinct from the efficient causation that is widely agreed to hold between physical events. Agents, they say, can cause actions independent of causation by prior events. Agent causation, if coherent and real, is thought to explain how agents “could have done otherwise,” while still being consistent with robust moral responsibility.

Agent causation libertarians typically appeal to introspective experience as evidence for the existence of incompatibilist freedom. In an early expression of this view, Thomas Reid said “We are conscious of making an exertion, sometimes with difficulty, in order to produce certain effects. And exertion made deliberately and voluntarily, in order to produce an effect, implies a conviction that the effect is in our power” (Reid 1783/1785, 332). More recently, Timothy O’Connor agrees that agent causation is “the way we experience our own activity.” He describes his own experience as follows: “[I]t does not seem to me . . . that I am caused to act by the reasons which favor doing so; it seems to be the case, rather, that I produce my own decisions in view of those reasons, and could have, in an unconditional sense, decided differently” (O’Connor 1995, 196).

But there are serious questions about whether introspective awareness can reveal anything so robust as support for the existence of the controversial form of causation these libertarians are positing. Schopenhauer once suggested that the question of whether any choice could have turned out differently “is a question so remote from self-consciousness that [we] cannot even understand it, much less have a ready answer for it, or even only the undeveloped germ of one” (Schopenhauer 1888/1839, 17). Numerous more recent philosophers, even those who agree that introspection can be a reliable source of information, contend that what experience we have of our own efficacy does not necessarily support the

*incompatibilist* claims that we could have done otherwise or that our that our own efficacy is itself independent of prior determining causes.<sup>1</sup> There are, in other words, ways that compatibilists seek to accept the introspective awareness of our agency in a way that does not support the libertarian’s rejection of determinism.

In this paper I propose a defense of incompatibilist libertarianism by describing our introspective experience of agency in a way that I think avoids common objections to agent causal incompatibilism. I will concede that our introspective awareness of agency does not *per se* include an account of our “ability to do otherwise.” But conceptualization of this experience does provide *prima facie* evidence implying the existence of that ability, especially when we focus on the appropriate aspects of our agency—in particular, on our *rational* agency.<sup>2</sup> In the final two sections of my paper, I will explore an alternate compatibilist account proposed by Terry Horgan (in press) and argue that it fails as the best explanation of the relevant evidence. This leaves the incompatibilist account as the best one standing.

## 2. The phenomenology of rational agency

Before directly presenting my account of our introspective awareness of agency, it will be useful to present a controversy between libertarians and compatibilists about the phenomenology of free will. Nahmias et al. (2004) describe how libertarians and compatibilists differ over which types of decisions to treat as paradigmatic of our freedom. Libertarians who maintain that our freedom includes an ability to do otherwise in a given situation present “close-call” decisions in which we feel torn between two competing motivations as paradigmatically free.<sup>3</sup> Compatibilists who maintain that our freedom consists in some form of internal psychological causation as opposed to external compulsion do not need to focus on “close-call” choices: they instead present straightforwardly confident choices as paradigmatically free choices.<sup>4</sup>

To resolve this dispute, Nahmias et al. suggest that we look to ordinary folk accounts of the introspection of agency, rather than to those of philosophers whose accounts might be subject to confirmation bias. Indeed there are studies which show that subjects claim to feel less control over their actions when their deliberation is fraught with uncertainty (Wells 1927, Perlmutter and Monty, 1979,

---

<sup>1</sup> See, e.g., Grünbaum (1971), p. 306.

<sup>2</sup> I take this evidence to be *prima facie* in a dialectical sense: it implies incompatibilism as long as there is no alternate compatibilist account of the same evidence. I would distinguish the dialectically *prima facie* from another, more robust sense. I do not think, for instance, that ordinary subjects need to refute compatibilists in order to be justified in their conviction that they could do otherwise. For this reason I would not describe the evidence I am about to consider as “*prima facie*” in an unqualified sense. But incompatibilist philosophers do need to refute compatibilist accounts—in order to explain why ordinary subjects are justified in ignoring compatibilist alternatives.

<sup>3</sup> See van Inwagen (1989, 234) and Kane (2002, 228).

<sup>4</sup> See Dennett (1984, 133), Frankfurt (1991, 104) and Wolf (1990, 56).

Wescott, 1989). This isn't good news for incompatibilists who think that freedom is best experienced in the course of uncertain deliberation about "close-call" decisions.

But there is a way to account for ordinary subjects' reactions here without abandoning the incompatibilist conviction that our freedom involves the ability to do otherwise. In the rest of this section I describe the phenomenology of a fundamental alternative that appears to be involved in cases of both uncertain and confident deliberation, the introspection of which I take to be the primary way we come to know of our free agency.

Consider an insight from Reid. Rather than focusing on the uncertainty of deliberation, Reid focuses on the *exertion* involved in deliberate action. Now it is understandable that later incompatibilist libertarians would shy away from focusing on the awareness of effort or exertion, since we can imagine that non-human animals would experience the same, and we do not easily consider non-human animals as possessing the kind of freedom that underpins moral responsibility. But Reid also thought that "moral liberty" was restricted to adult human beings, and criticized compatibilist definitions for implying that the actions of non-human animals, infants, and "madmen" are free. Unsurprisingly, at least some of the exertion Reid highlighted in his introspective account was distinctively human: "The language of all mankind, and their ordinary conduct in life, demonstrate that they have a conviction of some active power in themselves to produce certain motions in their own and in other bodies, *and to regulate and direct their own thoughts*" (my emphasis, 1983/1785, 332).

Discussions of free will usually proceed on the assumption that the paradigmatic control we exercise over our lives is over bodily actions. And clearly there can be cases in which the decision to raise or lower our arm can be as directly intentional as anything else. But most of our bodily movements on a given day—consider just the actions of moving your fingers while typing—occur without conscious thought. Most bodily motions are parts of larger routines of action that are set into motion by decisions we make well in advance. We do have the sense that we could stop them at any moment, courtesy of the effort we experience while acting. And a free choice to move our bodies is easily analyzed by the compatibilist who notes that actions of sane adult human beings are motivated by practical reasons, and so in some important sense *caused* by them.

Some agent causation theorists try to answer the compatibilists by claiming that reasons can explain our actions without causing them. I have deep misgivings about this view.<sup>5</sup> An alternative is to

---

<sup>5</sup> Agent causation theorists like O'Connor (2000) think that while free actions may be explained by such reasons, we need not consider reasons as causes in order to mount this explanation. On O'Connor's views, agents form intentions which reflect the content of their reasons, and so reasons explain by having the same content as intentions to act to satisfy those reasons, but the reasons themselves do not cause our intentions or the actions that result. I find this view of reasons highly implausible, especially since it has the implication that motives and emotions do not in any sense *move* us, even when there is plentiful introspective evidence to suggest that they do, and it is not clear what distinguishes emotions from other propositional attitudes if they do not have this power. So if O'Connor's account of reason explanation does not defang the compatibilist's understanding of bodily

think that reasons cause actions to the extent that agents are the active authors or sources of reasons for action. We can maintain that the agent is the author or source of a reason if we think of both emotions and beliefs as activated and realized in the first place through the intellectual activity of the agent. More obviously, we experience emotions only to the extent that we are aware of their stimuli, and if the attention required for this awareness is itself volitional, then emotions at least have a volitional precondition. Furthermore, while there is controversy about whether there is any sense in which we choose our beliefs as such, there is much less controversy over the idea that our beliefs are at least the products of the control we exercise over acts of inquiry.<sup>6</sup> Finally, though it is more controversial, cognitivist theories of the emotions hold that emotions themselves are the form in which we experience our cognitive value judgments about facts in our environment.<sup>7</sup> The theory receives additional empirical support to the extent that it is presupposed by the highly successful cognitive-behavioral approach to psychotherapy.<sup>8</sup> If this theory has merit, the degree to which we control our emotions reduces to or at least covaries with the degree to which we control our thinking.

One of the earliest critics of compatibilism, Alexander of Aphrodisias, critiqued the Stoic definition of freedom (of “what is up to us”) on the grounds that it failed to include the crucial form of human responsibility, the control we exercise over whether to deliberate or not (Sharples 2007, 56–56). More recently, Binswanger (1991) has proposed an incompatibilist libertarian agency theory according to which *the choice to think or not to think* is the “direct locus of free will” (161).<sup>9</sup> On this view, the distinctively human rational form of awareness, (the ability to form and apply abstract concepts, whether in the form of basic judgments or more advanced inferences) is not automatic but volitional. We face a constant choice to focus or unfocus our minds. Consequently the basis for our conviction about our free agency is our introspective awareness of the difference between the agency we exercise in conceptual cognition vs. our automatic perceptual awareness. It is this volitional nature of conceptual awareness that accounts for the effort we must exert as we “regulate and direct our own thoughts,” as Reid put it. On this

---

movement, how can agent-causal libertarians make sense of the idea that agents and not antecedent events are responsible for our bodily movements? See my suggestion in the main text that follows.

<sup>6</sup> For fairly standard objections to doxastic voluntarism, see Williams (1973) and Alston (1988). For a compatibilist critique of Alston, see Steup (2011). For an incompatibilist libertarian critique of Alston, see Salmieri and Bayer (forthcoming). This latter takes for granted the spirit of Williams’ critique, acknowledging that while we cannot here and now choose in full consciousness to believe any proposition we like, we can engage acts of inquiry which result in a belief in some proposition, which acts themselves partially constitute the mental state of believing. For two views which dispute the claim that we choose our beliefs as such but acknowledge that we control the intellectual activity that results in our beliefs (a point Alston himself concedes) see Heil (1983) and Audi (2001).

<sup>7</sup> The cognitivist view harkens back as far as Aristotle and the Stoics. For contemporary defenders of cognitivism about the emotions, see in particular the psychologists Arnold (1960), Beck (1976), Lazarus (1991) and Martin and Clore (2009), and the philosophers Solomon (1980) and Nussbaum (2001).

<sup>8</sup> See Butler et al. (2006) for a meta-analysis of the literature comparing CBT to rival approaches for treating such disorders as depression, anxiety, phobias, and PTSD.

<sup>9</sup> Binswanger draws on ideas from Rand (1957). For another account of Rand’s view, see Peikoff (1991, pp. 55–72). For a 19<sup>th</sup> century account according to which our basic choice involves an act of attention, see James (1925, chapter 15).

view, it is our choice to raise or lower our level of conceptual awareness that accounts for whether we bring reasons to bear on our actions, and therefore whether we undertake certain actions or not.

Binswanger sketches an example that illustrates how the choice to think or not to think is fundamental to alternatives in bodily action. A student faces the choice to study or to watch TV. In one scenario, when the student thinks of the test, he thinks about how boring he finds Professor Winston and resents having to study. He would rather be watching TV with his friend Maureen. He muses that he can probably get by without studying, so he reaches for the phone to call Maureen. In the second scenario, the same student asks what is at stake in this decision: it will affect his ability to get into graduate school. He asks what he needs to study, and remembers how unclear some of the material was to him and even how boring Professor Winston can be. But he catches himself and notes that the issue Professor Winston's lecture style is irrelevant to the current question. He considers that it would be fun to watch TV, but realizes that he won't really enjoy himself since he'll feel guilty about not studying, and besides he can always find a tape of the show later. He opens his books.

The student in either scenario is going through a kind of deliberation. But on the present theory, it is not the experience of deliberation *per se* that accounts for the student's experience of freedom—it is the experience of the choice to deliberate or not, and if so, to what extent and degree of intensity. We do not (often) experience the choice of whether to deliberate as being torn between two competing motives, not unless we consciously fear the truth. Yet we are continually faced with the choice of exerting cognitive effort or not. Any time we attempt to answer *any* nontrivial question conscientiously, we face the same alternative of whether to believe *p* or not, but this does not mean we face a dilemma. Arguably, we only face dilemmas once we have made the primary choice and are in a position to recognize (fully or partially) a conflict between some of our goals. So, if we reconceive our freedom as centered fundamentally in our cognitive management, we can see how folk reports about free and responsible choices which do not involve dilemmas do not necessarily call into question incompatibilist accounts of freedom.

There is a certain asymmetry here in how reasons cause actions in the case of the active-minded thinker versus that of the passive-minded thinker. Active-minded agents formulate and endorse reasons and go on to act on them, whereas passive-minded agents merely permit or refrain from preventing an uncritically-evaluated reason to operate. This means that there is a sense in which the difference between the active- and passive-minded approaches is not a difference between two kinds of agent-causation, but between agent-causation and the *absence* of agent-causation. Here there is some parallel to the way philosophers in the Kantian tradition think about freedom of the will, as something that has to be achieved or realized and not merely presupposed by any decision. In an important sense, the fundamental choice is

*to be an agent or not.* Active-minded thinkers help make themselves autonomous agents rather than passive pawns of random external influences.

This asymmetry in the experience of being an agent or not points to an additional aspect of the introspective awareness of our free agency *if we are a rational agent*: the fact of having a sense of ownership of one's conclusions and of having managed executive control over one's mental life more generally. Taking ownership in this way or not also has additional affective consequences: active-mindedness is conducive to self-esteem, while passive-mindedness leaves one subject to anxiety. (Cognitive therapists have learned a great deal in the last forty years about how patients' failure to bring to light automatized ruminant thinking patterns can be responsible for a variety of disorders, including especially depression.<sup>10</sup>) But the same asymmetry also gives us an error theory for at least some of those who reject the present theory. Some may have a harder time recognizing the introspective evidence for their own free agency if they choose *not* to make themselves fully rational agents: if they habitually refrain from taking responsibility for their inner mental life, they *will* feel like their choices are determined by forces beyond their control.

I believe but cannot prove now that this consideration may help dissolve another dispute described by Bayne (2008) about whether the phenomenology of agency is primarily "descriptive," as Horgan et al (2003) maintain, or intentional and "directive," as Searle (1983) maintains. The dispute is about direction of fit: Searle thinks that unlike perception, introspection of action is not veridical in virtue of having some object to which it causally responds, but rather it is successful if an action successfully satisfies an agentive attitude. I am not myself sure what to make of the notion of "direction of fit," but on the view that freedom consists of a kind of self-constituting awareness, the awareness of freedom would seem to feature both directions of fit, embodying both the awareness of an independent object but also the satisfaction of a goal. Some philosophers have argued that mental acts can involve both directions (e.g. Millikan 1996), and perhaps they would sympathize with the present account.

If the above is correct, the introspective phenomenology that best reveals our fundamental freedom of agency is the experience of feeling the need to maintain our cognitive effort and of our ability to do so or not. Of course, that we think or feel that we have some ability does not by itself necessarily imply that we have it. In the next section I will argue, however, that the introspective evidence considered above does support the existence of this ability, at least in a *prima facie* way.

---

<sup>10</sup> See Nolen-Hoeksema et al. (2008).

### 3. Rational agential phenomenology and the *prima facie* implication of incompatibilism

Even if the introspective experience described above is the paradigm of free and responsible action, we may still wonder if this experience carries with it any (accurate) information about whether or not any introspected act *could be otherwise*, or about whether a self-caused act is independent of antecedent causes. Introspection would fail to support incompatibilist libertarianism if it carried no information one way or the other.

Terry Horgan (2012) has argued that we cannot resolve the question between compatibilism and incompatibilism by appealing to introspection alone. I agree with Horgan that *simple* introspection of my experience of free agency does not answer this question. If, as I maintain, there is a distinction between raw introspective experience and introspective judgment, raw introspection alone carries no *conceptual* information at all, so certainly nothing about whether it is a fact that we have the ability to do otherwise or not. But even apart from this somewhat controversial distinction, compatibilists may still argue that our conceptualization of the introspective aspect of our rational agency may also be distinct from the conceptualization of our ability to do otherwise or not.

The question then becomes whether through additional reflection, we can go on to judge our introspective experience as providing evidence that we could do otherwise. So I agree with Horgan that resolving the question between compatibilism and incompatibilism requires more sophisticated conceptual apparatus than mere raw introspection, but I think a different category of apparatus is available which supports incompatibilism. The apparatus draws in particular on the distinctive kind of introspective awareness which I described in the previous section, where I described the experience of choosing to focus one's mind or not as central to our experience of ourselves as free *rational* agents.

Horgan points out that it may be true that when the agent reflects on her state of focus, her experience does *not present her behavior as causally determined by prior conditions*. But this does not mean that it *presents her behavior as not causally determined*. This is true as far as it goes. Now it is easy to *deny* that the agential experience we share with animals, that of being the immediate origin of our own action, includes an awareness of being undetermined. But it is much harder to deny that our experience of being a fully *rational* agent include an awareness of being undetermined. For to understand what it is to be have the capacity of rationality, which involves the possibility of irrationality, is to be able to distinguish between these states *normatively*. It is, among other things, to appreciate the difference between a state of mind that one ought to adopt and one which one ought not.

We must now come to grips with the time-honored albeit controversial principle that “‘ought’ implies ‘can.’” If “‘ought’ *does* imply ‘can,’” then it is illegitimate to identify as irrational those subjects who *cannot* choose to focus in a given instance. If determinism is true, they cannot think otherwise and they are neither rational nor irrational. So any agent who thinks of himself that he can think otherwise and

who agrees that “ought” implies “can” will judge his experience of rational agency as fully incompatible with determinism. Of course the “‘ought’ implies ‘can’” principle is controversial, especially among compatibilists. But to make the present, fairly delimited point, I don’t need the principle to be *true*. I only need it to be a principle that agents *believe*. So while it is true that introspection of our rational agency as such does not carry with it the information that such agency is incompatible with determinism, that experience plus some common beliefs (e.g., that “ought” implies “can,” that we ought to be rational and ought not to be irrational) do seem to imply an incompatibilist conclusion. Whether that conclusion is correct would of course depend on the truth of those common beliefs. So no matter what, ordinary subjects who have these common beliefs will be wrong about *something* if determinism is true, even if they are not wrong about their freedom. Determinism is incompatible with *some* ordinary beliefs about normativity and responsibility.

Now some may object that subjects could and maybe ought to be convinced to revise their understanding of normativity in a way that does not imply the ability to do otherwise, for instance. However, the type of introspective experience I have now described also offers independent support for incompatibilism in other ways that will tend to reinforce the connection between normativity and the ability to do otherwise. I see this additional support as coming from two considerations. First, the fact that centering our agency in the control we exercise over our thoughts makes a variety of proposed compatibilist analyses more difficult to defend. Second, evidence of our *rational* agency plays a crucial role in a traditional argument against the rationality of determinism which undercuts compatibilism about *epistemic* responsibility in a significant way. Each of these considerations has received more extensive development elsewhere, so I can only sketch them briefly.

Consider first of all a typical compatibilist’s response to the notion that “ought” implies “can.” “Ought” would clearly not imply “can” if we could reconcile some form of moral responsibility with determinism. In the classical compatibilism of Hume, for example, an agent is “free” and so morally responsible for an act provided that her action has its source in her internal will, as opposed to external forces. Reid and many others criticized this analysis on the grounds that it would count non-human animals, infants, and the insane as morally responsible for their actions. Some additional criterion is needed to distinguish the *type* of internal psychology that could be regarded as producing morally responsible actions. It turns out that it is harder to give an analysis of epistemically responsible *thought* than it is to give one of morally responsible *action*. On the one hand we can’t identify responsible thought with merely any thought that is caused by “reasons,” since there can be pathological reasons. But it also can’t be identified with thought that is caused by rational reasons, since we presume people can be responsible and blameworthy for irrational thinking as well.



More sophisticated forms of compatibilism such as Fischer and Ravizza's (1998) "reasons-responsiveness" semi-compatibilism depend heavily on rejection of the principle of alternative possibilities, the idea that to be morally responsible for an action, we need to be able to do otherwise in the same situation. Usually Frankfurt examples are seen as motivating the rejection of this principle, because they are cases in which we would hold an agent responsible for an action, even though in a counterfactual situation some manipulator would guarantee that the agent undertake the same action even if he would relent from it. So if we can hold an agent morally responsible for an act (say, a blameworthy act such as an assassination) without assuming that this agent could do otherwise, it must be that the concepts of "ought" and "ought not" are applicable even if the agent cannot perform the praiseworthy act (e.g., not assassinating someone).

However, a popular response to semi-compatibilism is that it does not seem that Frankfurt examples really provide examples in which *the same agent* cannot help but perform *the same act* in either case. A would-be assassin who relents but who is forced into performing the assassination after all is not performing the same act as one who performs it of his own will, at least not if we individuate actions by causal history (Van Inwagen 1983). And furthermore, the manipulated assassin is not even the agent of action, the manipulator is (O'Connor 2000, Rowe 2003). This point is even more pressing when we consider *doxastic* versions of Frankfurt cases, in which the agent's relenting looks to be actually constitutive of the relevant alternate possibility. Bayer (unpublished--b) develops this line of criticism of doxastic Frankfurt examples, urging that there is little sense to the idea that the doxastic agent's relenting to believe something is a mere "flicker of freedom" bereft of epistemic responsibility, given that this relenting would seem to constitute a doxastic decision in and of itself. So it is not clear that Frankfurt cases easily challenge "ought" implies "can" for the kinds of norms—epistemic norms—that matter most in the phenomenology of rational agency.<sup>11</sup>

There is a second way in which introspection of our rational agency lends further support to incompatibilism. This time, however, it comes in the form of a traditional argument against determinism itself, one that is especially prominent in the Kantian tradition (Kant 1783/1999; 1785/1993) and which was given contemporary expression by Jordan (1969).<sup>12</sup> According to this argument, a belief in determinism is self-defeating. The argument runs roughly as follows:

1. If determinism is true, our assessments of the relevance of our reasons are determined by causes that needn't be sensitive to objective criteria of relevance.
2. If our assessments of the relevance of our reasons are determined by causes that needn't be sensitive to objective criteria of relevance, we cannot rationally believe anything on the basis of our reasons.

---

<sup>11</sup> For an additional and recent challenge to the compatibilist's take on doxastic Frankfurt example, see Peels (forthcoming).

<sup>12</sup> For more on the general Kantian doctrine, see Allison (1989). For a critique of Jordan's argument with recommendation for how to improve it, see Boyle et. al (1972).

3. If we cannot rationally believe anything on the basis of our reasons, we cannot rationally believe determinism is true.
4. Therefore, if determinism is true, we cannot rationally believe determinism is true.

From this, of course, it follows that if we believe that determinism is true, we must believe that we cannot rationally believe that determinism is true. This admittedly controversial argument is itself only an argument against the rationality of believing in determinism, not an argument against determinism as such. Still, one important aspect of the argument contributes to the case for incompatibilism. The crucial premise is (2), which alleges that the truth of determinism implies a kind of skepticism. If this is correct, and nothing we believe as a product of inference is rational, then the introspective awareness we seem to have an ability to be rational or not is once again called into question. So premise (2) implies a kind of incompatibilism: it urges that determinism is incompatible with the kind of freedom we associate with *epistemic* responsibility: it suggests that if determinism is true, no one is epistemically responsible (whether rationally or irrationally).

Compatibilists will of course object to premise (2). A.J. Ayer (1963, 266–67), for instance, once alleged that the argument falsely assumes (or begs the question in assuming) that reasons cannot themselves be sufficient causal conditions. But Jordan contends that it assumes no such thing. Even if we assume that reasons can be a type of cause, for a subject to qualify as a rational agent, she needs to be able to do more than respond to reasons. The reasons also need to be *good* (rational) reasons, and the subject needs to have some *awareness* that they are good reasons, i.e. that they are relevant to the conclusion.

What Jordan describes is essentially an epistemic regress problem for the determinist, one which he suggests we can solve only by rejecting determinism. Here premise (1) plays an important role: because the determinist must hold that our assessment of the relevance of reasons will, like everything else, be determined, there is then a question of how the determinist subject's relevance assessment can be objectively superior to another's. At some point, to terminate a regress, as in Lewis Carroll's "What the Tortoise Said to Achilles," a subject must simply rely on some logical rule or standard without adding it as a premise. The reliance on this rule will constitute the subject's assessment of the relevance of his reasons, but at pain of regress, a determined subject's assessment is not dependent on any further articulated reasons. Instead this assessment is simply caused, not by the consideration of some additional premise, but by external factors not necessarily connected to the objective logic of the argument. Some subjects could be necessitated to find affirming the consequent primitively compelling, while others might not. Still others could find a case of modus ponens compelling but for the wrong reasons, perhaps because they fail to distinguish it from affirming the consequent.<sup>13</sup> So if a subject's reasoning must ultimately rely

---

<sup>13</sup> The skeptical problem here is analogous to the skeptical puzzle attributed to Wittgenstein by Kripke about how a subject can know he is relying on the "plus" rather than the "quus" function in performing an arithmetical operation.

on an assessment of relevance whose cause is inscrutably related to whatever makes principles of logic non-arbitrary, then whether the subject is relying on a non-arbitrary criterion here seems completely fortuitous, and the subject is not in a position to rationally believe the results.

Only if we reject determinism, on this view, does the subject know that she has applied the right rule of inference. She knows this through a special form of self-awareness, awareness of her own volitional acts: she knows she is being rational because, she knows that *by an act of will she made it true that she is being rational*.<sup>14</sup> Suppose we consider the most general of logical norms (rather than specific rules like *modus ponens*). For instance, how does the subject know that she is accepting some conclusion *because it conforms to the evidence*, rather than to some prejudice? The subject can know it *because she knows she directly opened her mind to the evidence through an act of will, to the exclusion of prejudicial factors*.<sup>15,16,17</sup>

Because this argument assumes that we need to *know* if our assessment of the relevance of a reason is rational, the question of whether we should accept premise (2) then turns on a resolution to the controversy between internalism and externalism in epistemology. If we accept internalism and suppose that we must have conscious access to all the relevant justifiers of our beliefs, we accept that a rational agent is not one whose beliefs *simply* causally track the truth or which are the products of a reliable

---

<sup>14</sup> This is a point not found in Jordan's account which I believe is necessary to make his argument complete and answer typical objections to it, such as those found in Boyle, et al, (1972). Indeed something like this idea may well be substantively the same as Boyle et al.'s own attempt to defend the argument that determinism is self-defeating. I'm indebted to Darryl Wright for coming to see the idea that an act of will is needed to *make* a foundational justifier true here. I suspect that the same point could also play a role in solving Wittgenstein's (or Kripke's) puzzle about rule-following.

<sup>15</sup> On this view, this ability to directly apply a logical rule would also express the subject's choice of a basic doxastic motivation, something like a will to believe the truth. Understanding this is also important for resolving the debate about doxastic voluntarism, the question of whether there is any sense in which we choose our beliefs as such. Pamela Hieronymi (1996) has alleged that the voluntary control we exercise over our cognition does not amount to choosing our beliefs at will, because choosing at will presupposes choosing in response to a practical reason. As Bayer (unpublished-a) has argued, however, this will to believe the truth could be understood as a practical reason relevant to all epistemically responsible belief choice. In that case, there could be an important sense in which we do choose our beliefs as such.

<sup>16</sup> In connection with the idea that we know our rationality in virtue of our self-constituted rational activity, it is probably worth considering the fact that agent-causal theories have long maintained that there is an important sense in which we *also* only come to grasp the idea of cause and effect in general first by grasping our powers as agents to affect the world. This idea appears in Reid (1983/1788, p. 332) and Chisholm (1982, pp. 31–32), who each claim that our grasp of agent causation is prior to and the basis for our grasp of event-event causation. An interesting possibility that might account for *how* knowledge of the first kind of causation might account for knowledge of the second kind arises when considering the so-called manipulability theory of causation (von Wright 1971, Hausman 1986, Price 1991, Menzies and Price 1993). According to this theory, we understand a cause as that which we could manipulate in order to bring about a desired effect. Think here of a controlled experiment in which we discover a cause by holding everything constant between two cases and varying only one factor between the two, such that the effect obtains in one case but not in the other. Here it is the agent's own knowledge that *the agent* varied one factor and kept the others the same that guarantees the knowledge that only one factor has changed and could be the cause. So perhaps self-constituted rational activity not only provides a foundation for all epistemic norms, but also specific norms of inductive reasoning about cause and effect.

<sup>17</sup> Understanding the role of the will here in helping to *constitute* one's justification helps to explain why Sellarsian arguments against "the myth of the given" fail. Sellars and McDowell suggest that foundationalists who see appeal to the given as an ultimate justifier claim that the resulting foundationalism is naturalist-externalist, assimilating justification with explanation or exculpation. They insist that rational agents are not mere reliable responders. But if the present account is correct, it is precisely a robust rejection of naturalistic externalism, in the form affirming a libertarian account of rational agency, that permits a foundationalist account of justification.

belief-forming process. If, however, we accept externalism and deny that a rational agent has conscious access to all of one's justifiers, premise (2) fails.

We cannot hope to resolve in this space the controversy between internalism and externalism, though we should confess our independent conviction that the internalists are fundamentally right. It is not surprising that a leading criticism of the externalist view of justification was that it implied a kind of skepticism not too different from the sort being ascribed here to determinism (Stroud 1981, 1989). These allegations are convincing enough to lead critics of internalism like Quine (1981) to concede that "naturalized" approaches to epistemology are not even attempting to answer the skeptic in the traditional way, for such is to concede too much to the foundationalist.<sup>18</sup> Critics of internalism, in that case, had better embrace a pragmatist approach to knowledge and simply make free use of science, without concern for question-begging, to predict and explain the behavior of subjects. We do not need to embrace Quine's radical naturalism to see that many externalists are content to abandon the goal of traditional normative epistemology: the provision of guidance to the individual knower. We cannot expand here in full on why we think it would be a mistake to abandon the normative project (and why it need not be abandoned<sup>19</sup>). But we can say that to the extent that an internalist conception of rationality underpins the argument that determinism is self-refuting, it follows that the whole repertoire of evidence (introspective or otherwise) that is normally offered in defense of internalism thereby becomes part of the evidence supporting incompatibilism.<sup>20</sup>

However, what I have shown at this point is at most that the introspective awareness of and conceptualization of rational agency *prima facie* implies incompatibilism. Whether the implication avoids defeaters is subject to scrutiny by compatibilist interpretations of our phenomenology of agency. In the next section I will present one such compatibilist interpretation. .

#### **4. A compatibilist account of the phenomenology of agency**

Terry Horgan (in press) has argued for a form of compatibilism, according to which our introspective experience of our own free agency can be veridical and our judgments about the existence of such agency can be true even if determinism is true.

Horgan describes his compatibilism as a complex of different theses, many involving complexities I do not need to summarize here. For instance Horgan claims that although one's introspective experience of free agency is compatible with the truth of determinism, one cannot

---

<sup>18</sup> For more on why Quine's response to Stroud amounts to a concession to skepticism in favor of a kind of untraditional pragmatism, see Bayer (2010).

<sup>19</sup> For a defense of internalism from a perspective we find favorable, see Bayer (2012).

<sup>20</sup> Here I have in mind, in particular, the array of counterexamples to reliabilist forms of externalism proposed by Bonjour (1980) which suggest that the reliabilist conception of justification would license patently irrational inferences (e.g. those based on clairvoyance).

experience a bodily event as also being caused by a mental or non-mental state. That is, the introspective experience of agency is *experienced* as being mutually exclusive with being causally determined. But Horgan thinks this phenomenological incompatibility is not itself evidence for the incompatibility of our experience and the possibility that our action might *in fact* be causally determined. Furthermore, Horgan maintains that our *judgment* that we are free agents is at least very often compatible with our actions being determined. In other cases he claims we can correctly judge that we are *not* free, e.g. in cases when we explicitly and philosophically consider whether our agency is compatible with state-causation, but he considers these as cases governed by “implicit contextual parameters” of our judgments about freedom. As long as our judgments that we are free are often true in many cases when we experience ourselves as free agents, even when our actions are determined, he sees this as sufficient for his version of compatibilism.

Horgan’s argument for this form of compatibilism is an inference to the best explanation from what he regards as data including facts about our intuitive judgments about freedom and about our phenomenological experience of agency. He regards the truth of compatibilism as the best explanation for the existence of this set of facts. Because there is an additional collection of facts that seem to support incompatibilism, he also attempts to show how his account would explain them away. Before I explain why I think the truth of compatibilism is not the best explanation for these facts, I will first summarize the facts in question. I begin by listing the facts Horgan cites about our *judgments* of freedom, and then facts about the phenomenology of our experience of free agency. While his thesis and this paper are primarily about introspective phenomenology rather than judgment, he presents the facts about judgment first, and I will defer to this order of presentation.

#### *4.1 Evidence from judgments about freedom*

*Type 1:* One has strong intuitive judgments that in some situations one is free, while in others one is not.

Horgan argues that the truth of compatibilism better explains the existence of our practice of making these judgments because according to compatibilism, these judgments are (often) *true*. In other words, we often make such judgments *because* they are (often) true. Incompatibilism, by contrast, “holds the truth of freedom-attributions hostage to a very demanding additional condition, over and above the conditions that are clearly satisfied in ordinary paradigmatic cases of free decision and free action—viz., the requirement of causal indeterminism in the processes generating the decisions and actions.” In other words, incompatibilism cannot as easily say that we make these judgments because they are true, for on incompatibilism they may very well not be true. So incompatibilism would need to offer some other

explanation for why we make potentially systematically false judgments, and this explanation would not be as simple as saying that we make them because they are true.

*Type 2:* Even though one judges that one's acts are free in paradigmatic situations, one's judgment can be pulled in the opposite direction when one considers the relation between freedom and determinism explicitly and philosophically.

Horgan leaves this data aside until after he has considered how compatibilism best explains the phenomenology of our experience of free agency. I will delay consideration of it until then as well.

*Type 3:* One's evidential standards for making the judgments above are commonly taken to yield judgments that are epistemically warranted independently of consideration of evidence for or against the truth of determinism.

Horgan's argument here is difficult to distinguish from his argument from type-1 data. Just as he thinks that our practice of making the judgments about freedom that we do is best explained by the *truth* of these judgments, he also thinks that our practice of regarding these judgments as warranted is best explained by the fact that these practices actually *are* warranted. This last is precisely what compatibilism maintains, in contrast to incompatibilism, which holds out the serious possibility that the evidential standards of our practices of judgment are too lax and hence our practices may be unwarranted.

*Type 4:* One's practices of making judgments about freedom play a role in moral and social practices which are seen as important whether our actions are determined or not.

Central to Horgan's argument here is that the concepts we use to make our judgments about freedom serve certain practical purposes and that failure of their "satisfaction conditions" can thwart these purposes. For example, we call some acts free because doing so allows us to hold agents morally responsible, which in turn justifies rewarding them or punishing them. I take it that Horgan means here that if it turns out that incompatibilists are correct then the truth of determinism would threaten our freedom *and* our ability to praise and blame, and hence our ability to reward and punish. Presumably we need to be able to reward and punish whether or not determinism is true, and this is the very purpose of the concept of "freedom." So the best explanation of the fact that we engage in the practice of reward and punishment at all is supposedly the fact that our judgments about freedom are themselves true independently of whether determinism is true or false.

*Type 5:* There is evidence that our decisions result from physical brain processes and no evidence that these processes are indeterministic.

Finally, Horgan claims the best explanation of this evidence and the absence of evidence of indeterminism is that there is simply not "the dramatic kind of causal indeterminacy of thought and action

that incompatibilism requires.” This means, with reference to type-4 data, that there is all the more reason to think that the purpose of our concepts of freedom would be thwarted if incompatibilism were true.

#### *4.2 Evidence from phenomenology of the experience of free agency*

To emphasize that the phenomenology he is concerned with explaining is the kind that involves “presentational” rather than “judgmental” content—roughly, the kind of content that is non-conceptual, non-linguistic and directly experienced—Horgan suggests that it is likely the sort that would also be experienced by non-human as well as human animals. If we think of this phenomenology as something we share in common with non-human animals, it then becomes a datum clearly subject to evolutionary-biological explanation (unlike practices of judgment which are subject to sociolinguistic explanations). What then is the best explanation of the fact that we, presumably along with other animals, sometimes experience ourselves as free agents?

According to Horgan, whether or not determinism is true, there are still “real distinctions among phenomena” that would be tracked by this phenomenology, the tracking of which would have adaptive value for survival and reproduction. Organisms that are able to experience the difference between those aspects of their environment that can be influenced by their bodily motion and those that are not will be likely to survive and flourish; those that are not, not so much. Horgan gives the example of a bear who is aware that its bodily motions will carry it towards a source of food: it is important that it be able to distinguish this self-caused motion from state-caused motions that it cannot control, e.g. the motion of a boulder rolling towards it. The data to be explained here is mainly the fact that it is adaptively useful to be able to distinguish these types of experience: Horgan suggests that this data would not be better explained if we assumed that the “satisfaction conditions” of our experience tracked more than the distinction between the self-caused and the state-caused. This suggests that phenomenology, even if non-conceptual, is still “representational” in some sense, and so either veridical or non-veridical. Hence we can understand his point by analogy to his point about the best explanation of our practices of judgment above: the best explanation for why we make those judgments—and for why we have these experiences—is that the judgments are true and that the experiences are veridical. Incompatibilism would hold out the possibility that our phenomenology is not veridical, and would require a more complicated explanation for why systemically non-veridical experience nonetheless yields useful results.

Horgan, of course, does not want us to infer from the fact that we experience a contrast between the self-caused and the state-caused that we can derive a case for *incompatibilism* from mere introspection, as one might at first think. Horgan insists that the introspective awareness of our own free agency by itself does not answer the question of whether free agency is compatible or not with determinism. His reason is that answers to such questions are simply not “manifest” to us in the way that

other facts are, a point we have discussed above. This then makes the experience of a distinction between the self-caused and the state-caused consistent with compatibilism.

The problem, of course, is the type-2 data from facts about our judgments of freedom: we can be tempted to judge that we are unfree when, for instance, we are asked explicitly and philosophically if the phenomenology of our agentic experience can be veridical if determinism is actually true. To account for this problem, Horgan would note that he has readily observed that the *experience* of self-causation and state-causation are mutually exclusive: we cannot experience our behavior as simultaneously self-caused and state caused. But the fact that we cannot experience an event in this way does not mean that an event cannot *be* both self-caused and state-caused, especially since not experiencing it as state-caused does not imply an experience of its *not* being state-caused. And as Horgan has been insisting, it makes evolutionary sense for us to experience the two kinds of causation as distinct.

Further, Horgan contends that there are “implicit contextual parameters” in our judgments about freedom, and that asking questions in explicit philosophical contexts drives these parameters to a “maximally strict setting.” So, just as contextualists would say that thinking about knowledge in the philosophy seminar drives up the standards for knowledge attribution to the point where we think we might have no knowledge, even when we would insist that we do in ordinary contexts, so thinking about our freedom in the philosophy seminar also raises our standards to the point of a similar kind of skepticism about freedom. And just as epistemic contextualists think that the philosophy seminar actually destroys our knowledge in philosophic settings, Horgan thinks that it is *true* in philosophic contexts that determinism is incompatible with freedom, because it is correct for philosophical subjects to say that we are not free, according to the stricter standards for what counts as “free.” This, however, does not mean that our judgments about the existence of freedom are false in ordinary contexts, and this is all the compatibilism he thinks he needs.

## **5. Why the compatibilism is not the best explanation of the phenomenology of agency**

It is hard to deny much of the data that Horgan musters for his abductive argument for compatibilism. The main question is whether the truth of compatibilism is in fact the best explanation of this data. Here I will survey each of the pieces of data presented and, in light of my positive case at the beginning of this paper, suggest that it is not.

Let us start with the type-1 data, that we often have strong judgments that we are free, while in other situations we judge that we are not. Horgan says that the best explanation of the fact that we make these judgments is that these judgments are true. I agree with Horgan that the best explanation of these judgments would be their truth. The question is why the chances of their being true are better on compatibilism than they are on incompatibilism.



To make this point, Horgan has to assume that it is more likely than not that determinism is true, and that incompatibilism would then imply hard determinism and falsify all judgments about our freedom. This outcome would certainly complicate things for the incompatibilist and demand a complicated error theory for our judgments of freedom. And Horgan has offered some reason to think it is more likely than not that determinism is true, namely his type-5 evidence, the scientific findings that show that human decisions and actions result from neurophysiological processes, together with his assumption that every physical event is causally determined (physical state-causal closure).

This inference assumes, however, that the case for determinism is settled on purely scientific grounds independent of the present philosophical debate. However there is at least some reason to think that this is not so. As I urged in my earlier section, part of the argument that determinism is self-refuting helps support incompatibilism. I can concede that many arguments for incompatibilism are independent of the case against determinism, but this one is not. If incompatibilism is true, it likely true on grounds that also make it more likely that libertarian incompatibilism is true than that hard determinist incompatibilism is. To suppose, then, that determinism is more likely true than not in order to defend compatibilism is precisely to beg the question in favor of compatibilism. This is underscored by the point that in order to answer the charge that determinism is self-refuting, a critic would need to provide a compatibilist analysis of the rationality of the reliance on various criteria of relevance. This kind of analysis, as I have suggested, is far more difficult for compatibilism about doxastic freedom than it is about practical freedom.

Furthermore, there is very likely an even more direct way in which Horgan's argument here begs the question in favor of compatibilism. I said that I agreed with Horgan that, *ceteris paribus*, the best explanation for the fact that we judge that we are free is that our judgments are true. However it is very often true that "*ceteris* is not *paribus*." There are certainly whole categories of discourse that can be in need of error-theoretic explanation, e.g. discourse about witches. To suppose that our discourse about freedom is not on par with our discourse about witches is already to suppose that the meaning of the folk concept of freedom is analyzable in a way that allows for the truth of compatibilism. Horgan cannot suppose this without offering his own compatibilist analysis, especially given that centuries of debate suggest that it is a challenging task. Yes, if we assume that "freedom" simply means causation by internal psychological states, judgments involving such a concept of freedom would be compatible with determinism. But countless counterexamples show that such a concept of freedom, which might be best described as *physical freedom* or *voluntariness*, is far from equivalent to *moral freedom*, the kind we see as relevant to making the very judgments of moral responsibility which Horgan sees as necessary for our practices of praise and blame. As Thomas Reid argued, while the concept of physical freedom or voluntariness is certainly one common sense concept, the distinct concept of moral freedom seems to

presuppose not only the ability to act according to one's will but the power to control one's will itself. I have already mentioned the standard counterexamples of infants, animals and madmen, whose actions are caused by internal psychological states but not often held to be morally responsible, because they have no such control. This is why even ordinary people outside of the philosophy seminar will attempt to excuse themselves of moral responsibility by claiming (justifiably or not) that they are not responsible for things they do when their action resulted from a will that was beyond their control (e.g. "the devil made me do it" or "the drink made me do it"). And yet to the extent that he speaks of it, Horgan seems to presuppose little more than the classical Humean analysis of freedom as physical freedom (this is evident when he claims that a phenomenological incompatibility between agency and state-causation would have adaptive value because animals need to know when an action originates in their body.) There are of course more refined compatibilist analyses that avoid these counterexamples (such as the reasons-responsiveness compatibilism discussed above), but these are contentious as well. This means that the reliability of folk judgments about freedom is still up for grabs if compatibilists cannot offer an acceptable analysis that saves them from error.

In light of the above, Horgan's explanation of the uncomfortable type-2 data on behalf of compatibilism seems especially *ad hoc*, and not the best explanation. To say that our concept of freedom in effect changes in different contexts from a compatibilist to an incompatibilist analysis is far from a "simple" explanation. Horgan offers no independent evidence of any psychological mechanism that would account for or realize the "implicit contextual parameters" he ascribes to our concepts of or judgments about freedom, nor does he offer any explanation for what the evolutionary advantage of such parameters would be. It seems he posits them only to explain away the conflict between the data and his compatibilism. If we assume that determinism is more likely true than false, as Horgan does, then in the absence of specific evidence of the psychological mechanisms in question, it would seem that the simpler explanation is that people's denial of freedom in philosophical contexts is *false* by the standard of the allegedly compatibilist folk concept. Alternately, still simpler than Horgan's *ad hoc* explanation is that people's affirmation of freedom in *ordinary* contexts is false by the standard of a superior philosophical concept (superior, perhaps, on grounds of pragmatic explication). But then Horgan is harder pressed to argue for the superiority of compatibilism on the grounds that the truth of our judgments about freedom is the best explanation for why we make them. True, compatibilism's simplicity would work in its favor to explain non-philosophical attribution of freedom, but it too would require an error theory for philosophical cases. And in a modern culture where more and more people are influenced by philosophy to deny the existence of freedom, this error theory could become very complex indeed.

We can now deal with the type-3 and type-4 data quickly. Type-3 data says that we presume our evidential standards for attributing freedom are warranted. As in my analysis of type-1 data above,

compatibilism is the best explanation for this data only if we assume that determinism is more likely to be rational to believe than not, which I have suggested is a question that is itself not independent of the outcome of the debate over compatibilism. Likewise to suppose that compatibilism would make these standards warranted would also assume the much contested compatibilist analysis of freedom, and beg the question still more directly. Type-4 data notes that our judgments of freedom serve practical purposes, e.g. the assignment of moral responsibility and the justification of reward and punishment. But to the extent that Horgan seems to presuppose a relatively simple compatibilist analysis of freedom that does not fully explain human practices of praise and blame. His is more akin to the *voluntariness* or freedom of action that is common to adult human beings, infants, madmen and animals, not the *moral freedom* that concerns our ability to control the will itself.

Finally, while type-5 data does seem to support the case for determinism, it is far from conclusive. If the argument that determinism is self-defeating is successful, we would need to find a way to make this data consistent with the failure of determinism. One possibility is that while human mental states causally depend on the physical, they do not reduce to the physical and so are not subject to deterministic laws of physics. A proposal along these lines has been advanced recently by E.J. Lowe (2008). Because Lowe rejects epiphenomenalism, he argues that free mental states still exercise downward causation on the physical. This implies a rejection of the causal closure of the physical, but Lowe argues that this principle, when interpreted most plausibly, does not rule out non-reductive mental causation, and when it does rule it out, it is not plausible on independent grounds. While this conclusion will be discomfiting to many naturalistically-inclined philosophers, I would suggest that its impact can be softened by showing why the failure of causal closure does not imply the violation of various physical laws, e.g. conservation laws, as both Lowe and, separately, Barbara Montero (2006) have argued.

Finally, Horgan's inference to the best explanation from the data of our introspective phenomenology fares little better in light of the above. To begin with, as I suggest above, the kind of phenomenology he considers is really that of our experience of *voluntariness*—which admittedly we do share with non-human animals. But on the pain of begging the question in favor of the most simplistic forms of compatibilism, this tells us nothing about the phenomenology of moral freedom, which presupposes the experience of doxastic freedom I have described in my earlier section. And as I have argued, it is precisely this latter kind of freedom which even the most refined versions of compatibilism have the most difficulty accounting for.

Though Horgan's argument is offered as a compatibilist account of the phenomenology of agency, it is much less focused on the phenomenology than it is on our judgments about freedom. When it focuses on these judgments, it fails to offer the *best* explanation for why we make them as we do. When it focuses on the phenomenology, it focuses on the wrong phenomenology. And when it relies implicitly on

the assumption that incompatibilism unduly risks hard determinism, it does not consider how the kind of incompatibilism supported by the phenomenology of *rational* agency lends special support to libertarian incompatibilism.

## 6. Conclusion

No doubt there is more compatibilists could say to answer the case I have outlined from the phenomenology of rational agency. Most crucially, they could attempt to offer compatibilist accounts of the norms of rationality, in particular by doubling down on externalist accounts of justification. Compatibilism about freedom and externalism are usually allied in spirit, owing to their common allegiance to a form of scientism. I remain skeptical that externalists can offer a theoretically satisfying epistemology for reasons I cannot expand on here. But I do have the sense that internalism is somewhat more popular as a theory of justification than libertarian incompatibilism is as a theory of freedom. So I hope at least to have convinced philosophers favorable to internalism to see that an implication of their concept of rationality is the ability to make a stronger case for libertarian free will. If nothing else, this should help show that a belief in robust libertarian freedom is not exclusively the province of supernaturalistic philosophy. To the extent that the internalist conception of justification represents the evidentialist tradition of Enlightenment philosophy, philosophers who respect the strictest standards of rationality can still be comfortable embracing a concept of free will so intimately bound up with this understanding of rationality. Say what you like about the prospects for determinism, but the concept of freedom defended here is embraced for reasons far removed from mystical irrationalism.

## Bibliography

- Allison, H. (1989). Kant's Refutation of Materialism. *The Monist*, 72(2), 190-208.
- 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). Fatalism. In A. Ayer, *The Concept of a Person and Other Essays* (pp. 235-69). London: St. Martin's Press.
- Bayer, B. (2010). Quine's Pragmatic Solution to Sceptical Doubts. *International Journal of Philosophical Studies*, 18(2), 177-204.

- 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. (Unpublished--a). Believing at Will and the Will to Believe the Truth. Retrieved from <http://www.benbayer.com/will-to-believe-the-truth.pdf>
- Bayer, B. (Unpublished--b). The Elusiveness of Doxastic Compatibilism. Unpublished manuscript. Retrieved from <http://www.benbayer.com/doxastic-compatibilism.pdf>
- Bayne, T. (2008). The Phenomenology of Agency. *Philosophy Compass*, 3(1), 182-202.
- Beck, A. (1976). *Cognitive Therapy and the Emotional Disorders*. New York: Meridian.
- Binswanger, H. (1991). Volition as Cognitive Self-Regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 165-178.
- BonJour, L. (1980). Externalist Theories of Empirical Knowledge. *Midwest Studies in Philosophy*, 5, 53-74.
- Boyle, J., Grisez, G., & Tollefsen, O. (1972). Determinism, freedom, and self-referential arguments. *Review of Metaphysics*, 26(1), 3-37.
- Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The Empirical Status of Cognitive-Behavioral Therapy: A Review of Meta-analyses. *Clinical Psychology Review*, 26, 17-31.
- Chisholm, R. (1982). Human Freedom and the Self. In G. Watson (Ed.), *Free Will* (pp. 24-35). New York: Oxford University Press.
- Dennett, D. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.
- Frankfurt, H. (1991). *Necessity, Volition and Love*. New York: Cambridge University Press.
- Grunbaum, A. (1971). Free will and the laws of human behavior. *American Philosophical Quarterly*, 8(4), 299-317.
- Hausman, D. M. (1986). Causation and Experimentation. *American Philosophical Quarterly*, 143-54.
- Heil, J. (1983). Doxastic Agency. *Philosophical Studies*, 43, 355-364.
- Hieronymi, P. (2006). Controlling Attitudes. *Pacific Philosophical Quarterly*, 87, 45-74.
- Horgan, T. (2012). Introspection about Phenomenal Consciousness: Running the Gamut from Infallibility to Impotence. In D. Smithies, & D. Stoljar (Eds.), *Introspection and Consciousness* (pp. 405-22). New York: Oxford University Press.
- Horgan, T. (in press). Causal Compatibilism about Agentive Phenomenology. In T. Horgan, M. Sabates, & D. Sosa (Eds.), *Supervenience in Mind*. Cambridge, MA: MIT Press.
- Horgan, T., Tienson, J., & Graham, G. (2003). The Phenomenology of First-Person Agency. In S. Walter, & H.-D. Heckmann (Eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action* (pp. 323-40). Exeter, UK: Imprint Academic.
- James, W. (1925). *Talks to Teachers on Psychology: And to Students on Some of Life's Ideals*. New York: Henry Holt and Company.

- Kane, R. (2002). *Free Will: New Directions for an Ancient Problem*. In R. Kane (Ed.), *Free Will*. Malden, MA: Blackwell.
- Kant, I. (1783/1996). Review of Schultz's Attempt at An Introduction to a Doctrine of Morals For All Human Beings Regardless of Different Religions. In P. Guyer, & A. Wood (Eds.), *Practical Philosophy* (M. Gregor, Trans., pp. 7-10). Cambridge: Cambridge University Press.
- Kant, I. (1785/1993). *Grounding for the metaphysics of morals*. (J. W. Ellington, Trans.) Indianapolis: Hackett Publishing.
- Lazarus, R. (1991). *Emotion and Adaptation*. New York: Oxford University Press.
- Lowe, E. (2008). *Personal Agency: The Metaphysics of Mind and Action*. New York: Oxford University Press.
- Martin, L., & Clore, G. (2009). *Theories of Mood and Cognition: A User's Guidebook*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Menzies, P., & Price, H. (1993). Causation as a Secondary Quality. *British Journal for the Philosophy of Science*, 44, 187-203.
- Millikan, R. G. (1996). Pushmi-pullyu representations. (J. Tomberlin, Ed.) *Philosophical Perspectives IX*, 185-200.
- Montero, B. (2006). What Does the Conservation of Energy Have to Do with Physicalism? *Dialectica*, 60(4), 383-96.
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2004). The Phenomenology of Free Will. *Journal of Consciousness Studies*, 11(7-8), 162-79.
- Nolen-Hoeksema, S., Wisco, B., & Lyubomirsky, S. (2008, September). Rethinking Rumination. *Perspectives in Psychological Science*, 3(5).
- Nussbaum, M. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- O'Connor, T. (1995). Agent Causation. In T. O'Connor (Ed.), *Agents, Causes and Events* (pp. 173-200). New York: Oxford University Press.
- O'Connor, T. (2000). *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.
- Peels, R. (forthcoming). Does doxastic responsibility entail the ability to believe. *Synthese*. Retrieved from <http://link.springer.com/article/10.1007%2Fs11229-012-0217-5#page-1>
- Peikoff, L. (1991). *Objectivism: The Philosophy of Ayn Rand*. New York: Dutton.
- Perlmutter, L., & Monty, R. (Eds.). (1979). *Choice and Perceived Control*. Hillsdale, NJ: Lawrence Erlbaum.
- Price, H. (1991). Agency and Probabilistic Causality. *British Journal of the Philosophy of Science*, 42, 157-76.
- Quine, W. (1981). Reply to Stroud. *Midwest Studies in Philosophy*, 6, 473-5.

- Rand, A. (1957). *Atlas Shrugged*. New York: Random House.
- Reid, T. (1983/1788). *Inquiry and Essays*. (R. a. Beanblossom, Ed.) Indianapolis: Hackett Publishing Company.
- Rowe, W. L. (2003). Alternate Possibilities and Reid's Theory of Agent-Causation. In D. Widerker, & M. McKenna (Eds.), *Moraal Responsibility and Alternative Possibilities* (pp. 219-234). Burlington, VT: Ashgate Publishing Limited.
- Salmieri, G., & Bayer, B. (forthcoming). How We Choose Our Beliefs. *Philosophia*. Retrieved August 5, 2013, from <http://link.springer.com/article/10.1007%2Fs11406-013-9462-1>
- Schopenhauer, A. (1999/1839). *Prize Essay on the Freedom of the Will*. (G. Zoller, Ed., & E. Payne, Trans.) Cambridge: Cambridge University Press.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- 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. (2011). Belief, Voluntariness and Intentionality. *Dialectica*, 65(4), 537-559.
- Stroud, B. (1981). The Significance of Naturalized Epistemology. *Midwest Studies in Philosophy*, 6, 455-71.
- Stroud, B. (1989). Understanding Human Knowledge in General. In M. Clay, & K. Lehrer (Eds.), *Knowledge and Skepticism* (pp. 31-50). Boulder: Westview Press Inc.
- van Inwagen, P. (1989). When is the Will Free? *Philosophical Perspectives*, 3, 399-422.
- von Wright, G. (1971). *Explanation and Understanding*. New York: Cornell University Press.
- Wells, H. (1927). The Phenomenology of Acts of Choice: An Analysis of Volitional Consciousness. *British Journal of Psychology: Monograph Supplements*, 4(11).
- Wescott, M. (1988). *The Psychology of Human Freedom: A Human Science Perspective and Critique*. New York: Springer-Verlag.
- Williams, B. (1973). Deciding to Believe. In B. Williams, *Problems of the Self* (pp. 136-151). Cambridge: Cambridge University Press.
- Wolf, S. (1990). *Freedom within Reason*. New York: Oxford University Press.