

Injecting the Phenomenology of Agency into the Free Will Debate

Terry Horgan, University of Arizona

The phenomenology of agency is the “what it is like” of experiencing oneself as an agent. The experiential aspect of freedom is an integral part of normal agentic phenomenology. In this paper I will urge the importance of injecting the phenomenology of agency into the free will debate, and I will explore the import, vis-à-vis the major competing positions, of doing so.

1. Some Reliably Introspectible Aspects of Agentic Phenomenology¹

I begin by describing some features of agentic phenomenology which, I submit, are readily ascertainable just on the basis of introspective attention to such phenomenology.

What is behaving like phenomenologically, in cases where you experience your own behavior as action? Suppose that you deliberately do something—say, holding up your right hand and closing your fingers into a fist. What can you ascertain about the phenomenology of this item of behavior, on the basis of introspective attention to this phenomenology? To begin with, there are of course the purely bodily-motion aspects of the phenomenology—the what-it’s-like of being visually and kinesthetically presented with your own right hand rising and its fingers moving into clenched position. But there is more to it than that, of course, because you are experiencing this bodily motion *as your own action*.

In order to help bring into focus this specifically actional phenomenological dimension of the experience, it will be helpful to approach it in a negative/contrastive way, via some observations about what the experience is *not* like. For example, it is certainly not like this: first experiencing an occurrent wish for your right hand to rise and your fingers to move into clenched position, and then passively experiencing your hand and fingers moving in just that way. Such phenomenal character might be called *the phenomenology of fortuitously appropriate bodily motion*. It would be very strange indeed, and very alien.

Nor is the actional phenomenological character of the experience like this: first experiencing an occurrent wish for your right hand to rise and your fingers to move into clenched position, and then passively experiencing a causal process consisting of this wish’s causing your hand to rise and your fingers to move into clenched position. Such phenomenal character might be called *the passive phenomenology of psychological state-causation of bodily motion*.² People often do passively experience

causal processes *as* causal processes, of course: the experience of seeing the collision of a moving billiard ball with a motionless billiard ball is an experience as-of the collision causing the latter ball's subsequent motion; the experience of observing the impact of the leading edge of an avalanche with a tree in its path is an experience as-of the impact causing the tree to become uprooted; and so on. Sometimes people even experience their own bodily motions as state-caused by their own mental states—e.g., when one feels oneself shuddering and experiences this shuddering as caused by of a state of fear. But it seems patently clear that one does not normally experience one's own actions in that way—as passively noticed, or passively introspected, causal processes consisting in the causal generation of bodily motion by occurrent mental states. That too would be a strange and alienating sort of experience.³

How, then, should one characterize the actional phenomenal dimension of the act of raising one's hand and clenching one's fingers, given that it is not the phenomenology of fortuitously appropriate bodily motion and it also is not the passive phenomenology of psychological state-causation of bodily motion? Well, it is the what-it's-like of *self as source* of the motion. You experience your arm, hand, and fingers as being moved *by you yourself*—rather than experiencing their motion either as fortuitously moving just as you want them to move, or passively experiencing them as being caused by your own mental states. You experience the bodily motion as generated by *yourself*.

The language of causation seems apt here too, but differently deployed: you experience your behavior as *caused* by you yourself, rather than experiencing it as caused by *states* of yourself. Metaphysical libertarians about human freedom sometimes speak of “agent causation” (or “immanent causation”), and such terminology seems *phenomenologically* apt regardless of what one thinks about the intelligibility and credibility of metaphysical libertarianism. Chisholm (1964) famously argued that immanent causation (as he called it) is a distinct species of causation from event causation (or “transeunt” causation, as he called it). But he later changed his mind (Chisholm 1995), arguing instead that agent-causal “undertakings” (as he called them) are actually a species of event-causation themselves—albeit a very different species from ordinary, nomically governed, event causation. Phenomenologically speaking, there is indeed something episodic—something temporally located, and thus “event-ish”—about experiences of self-as-source. Thus, the expression ‘state causation’ works better than ‘event causation’ as a way of expressing the way behaviors are *not* presented to oneself in agentic experience. Although agentic experience is indeed “event-ish” in the sense that one experiences oneself as undertaking to

perform actions *at specific moments in time*, one's behavior is not experienced as caused by *states* of oneself.

The phenomenology of doing typically includes another aspect which will be especially important in the context of the present paper: what I will call *core optionality*. (More presently on the reason for the modifier 'core'.) Normally when you do something, you experience yourself as *freely* performing the action, in the sense that it is *up to you* whether or not to perform it. You experience yourself not only as generating the action, and not only as generating it purposively, but also as generating it in such a manner that you *could have done otherwise*. This palpable phenomenology of optionality has not gone unrecognized in the philosophical literature on freedom and determinism, although often in that literature it does not receive as much attention as it deserves. (Sometimes the most explicit attention is given to effort of will, although it takes only a moment's introspection to realize that the phenomenology of effortfully exerting one's will is really only one, quite special, case of the much more pervasive phenomenology of optionality.⁴)

The core-optionality aspect of agentic phenomenology is intimately bound up with the aspect of self-as-source, in such a way that the former is an essential component of normal agentic self-source experience. In experiencing one's behavior as emanating from oneself as its source, one experiences oneself as being able to refrain from so behaving—or at any rate, as being able to refrain from willfully producing such behavior. This is so even when one acts under extreme coercion or duress—e.g., handing over one's wallet or purse to a thief who is pointing a gun in one's face. It also is so even when one acts with an extreme phenomenological "imperativeness"—e.g., a mother's unhesitatingly leaping into the river to save her drowning child, Luther's acting out a sense of moral requirement (as expressed by his declaring "Here I stand, I can do no other"), or the compulsive hand-washer's act of washing hands for the third time in ten minutes. The *core* phenomenology of optionality that is essential to ordinary agentic experience remains present in all such cases, even though there are further, superimposed, phenomenological aspects (duress, moral-obligation experience, intensely strong irrational desires, or the like) whose presence can render appropriate, in context, a judgment that the agent "could not have done otherwise," or "had no other option," or "did not act freely." (Such phenomenological aspects fall under a rubric I will call *superimposed non-optionality*.) Because the phenomenology of core optionality remains present even when the phenomenology of superimposed non-optionality is also present, it can be

contextually appropriate to use ‘could’ and ‘option’ and ‘free’ in a way that reflects this fact (rather than in a way that reflects the presence of one or another kind of superimposed non-optionality phenomenology). For instance, one might say this: “I could have refrained from giving the thief my wallet, and thus I gave it to him freely and with the option of refraining—even though refraining would have been quite stupidly irrational.” Hereafter I will use the expression ‘free-agency phenomenology’, in order to refer to the experience of self-as-source in a way that underscores the aspect of core optionality that is an essential component of normal self-as-source experience.

Agentive phenomenology is more closely akin to perceptual/kinesthetic experience than it is to discursive thought. (Many higher non-human animals, I take it, have some agentive phenomenology, even if they engage in little or no discursive thought.) Of course, we humans also wield *concepts* like agency, voluntariness, and the like (whereas it is questionable whether non-human animals do); but thoughts employing these concepts are not to be conflated with agentive phenomenology itself.

2. Frankfurt Scenarios: What They Show and What They Don’t

Scenarios of the kind first described by Harry Frankfurt (1969) have figured prominently in philosophical discussions of free will ever since Frankfurt’s landmark paper was published. Frankfurt himself took his original scenario to show that an agent can be *morally responsible* for performing a given action even if the agent could not do otherwise. Since then, many philosophers have taken such scenarios to show not only this, but also that an agent can perform an action *freely* even if the agent could not do otherwise. Given the facts about free-agency experience described in the previous section, how should one assess the relevance of Frankfurt scenarios to the free will debate?

Care is needed, with respect to matters of phenomenology, in what one stipulates about a given Frankfurt scenario. A “best case” scenario, I suggest—one that most clearly prompts the desired intuitions (that the agent acts freely, that the agent is morally responsible)—should run along the following lines. The malevolent being—who (unbeknownst to the agent) wants the agent to perform act A, is monitoring the agent’s decisionmaking processes, and is ready to intervene should the agent be about to decide against performing A—will intervene if need be, by causing the agent not only to perform A, but also to undergo an experience that is *phenomenologically indistinguishable* from the experience of freely performing A.

But, since the malevolent being ascertains that the agent is about to freely perform A anyway, the being does not intervene.

What is the significance of the fact that such scenarios seem coherently conceivable? To begin with, it should be noted that in such a situation the agent's free-agency phenomenology would be somewhat nonveridical. For, the intentional content of the phenomenology would include the aspect of core optionality, whereas in fact the agent could not have done otherwise than perform A.^{5,6}

Would this absence of core optionality make it the case that, under the use of 'free agency' most pertinent to questions of moral responsibility, the agent does not exercise free agency in performing act A? Intuitively, the answer seems to be "No." I see no good reason to repudiate this intuitive verdict. But consider now the following two claims:

- (A) It is not the case that in every possible situation S, an agent exercises free agency in S only if the agent can/could choose and act otherwise in S.
- (B) In no possible situation S is it the case that an agent exercises free agency in S only if the agent can/could choose and act otherwise in S.

There are certain Frankfurt scenarios that establish claim (A). So it seems to many philosophers, and so it seems to me too. But it would be a gross non sequitur to infer claim (B) from such Frankfurt scenarios. Indeed, it remains extremely plausible that in *most* possible situations that human agents might encounter—and probably all *actual* situations that humans have ever encountered (or ever will)—an agent exercises free agency in the given situation only if the agent can/could choose and act otherwise in that situation. That is, it remains extremely plausible that core optionality is a *defeasibly* necessary condition for the exercise of free agency.⁷ On this view, which itself should be the default view pending some independent argument in support of claim (B), the real relevance of Frankfurt scenarios is not that they establish claim (B)—they don't even come close to doing that—but rather that they specify potential (though hugely far-fetched) circumstances under which the defeasible requirement of core optionality would actually get defeated—viz., circumstances involving a *pre-empted potential cause*, waiting in the wings and unbeknownst to the agent, that would have caused the agent to have the experience as-of freely performing A had it not been the case that the agent freely performed A by herself/himself.

It is most unfortunate that so many philosophers in the free will debate, and so many compatibilists in particular, have been willing to embrace claim (B) on the basis of Frankfurt scenarios.⁸

This is not to deny that compatibilists who embrace (B) have made important and useful contributions to the case for compatibilism; on the contrary, I think they have, especially in their emphasis on the importance of reasons-responsiveness in decision making (e.g., Fischer and Ravizza 1998). But the fact remains that core optionality is a defeasibly necessary condition for the exercise of free agency. This should have been appreciated by philosophers even apart from considerations of phenomenology, but it becomes all the more palpable once one acknowledges that the aspect of core optionality is inherent to free-agency phenomenology and is highly intertwined with the phenomenological aspect of self-sourcehood.

3. Five Dimensions of the Free Will Debate

The philosophical debate about free will and moral responsibility is complex and multi-dimensional. So far I have been emphasizing the rich, introspectably palpable, phenomenology of free agency. Once this phenomenology is acknowledged, it needs to be taken properly into consideration as one dimension of the free will debate. In the present section I begin to explore its relevance to the debate by discussing its *prima facie* connections to several other dimensions.

To begin with, it is introspectively obvious not only that free-agency phenomenology exists (and is ubiquitous in everyday human experience of being and acting in the world), but also that this phenomenology is richly *intentional*: it experientially *represents* one's choosing and doings as emanating from oneself as their source, and as emanating from oneself in a way that is imbued with core optionality. Being intentional, free-agency phenomenology has *satisfaction conditions*: it imposes requirements on how things must be in the world—how things must be with respect to oneself—in order for one actually to be exercising free agency in the way one experiences oneself as so doing.

Important questions arise, regarding the intentionality of free-agency phenomenology. The following three questions pertain to content, and are labeled 'C' for *content* question:

- C.1. What are the satisfaction conditions of free-agency phenomenology?
- C.2. Does free-agency phenomenology have libertarian satisfaction conditions?^{9,10}
- C.3. Does free-agency phenomenology have compatibilist satisfaction conditions?

The following three corresponding questions pertain to introspectability vis-à-vis questions C.1 – C.3, and are labeled 'I' for *introspectability* question:

- I.1. Can question C.1 be reliably answered directly by means of introspection?
- I.2. Can question C.2 be reliably answered directly by means of introspection?
- I.3. Can question C.3 be reliably answered directly by means of introspection?

These questions will figure centrally in subsequent sections of the paper. Meanwhile they need to be borne in mind, as we consider how the phenomenological dimension of the free will debate is connected to other dimensions.

Four other dimensions each play a significant role: the conceptual, metaphysical, epistemic, and moral dimensions, respectively. I will describe some plausible-looking, *prima facie*, connections between the phenomenological dimension and each of these others, and among the others themselves.

First, the conceptual dimension. It is extremely plausible that satisfaction conditions for the concept of free agency are very largely—perhaps even entirely—inherited from the satisfaction conditions of free-agency phenomenology. Moreover (and this brings in the moral dimension), it is also plausible that context-dependent variations in uses of the concept are directly inherited from distinct aspects of the overall intentional content of agentive phenomenology. Thus, if one correctly remarks, “I could have refrained from giving the thief my wallet, and thus I gave it to him freely—even though refraining would have been stupidly irrational,” the contextually operative satisfaction conditions governing ‘could’ and ‘freely’ are inherited from the core-optionality aspect of one’s agentive phenomenology; whereas, if one instead correctly remarks, “I was not free to refrain from giving the thief my wallet, and I could not have done so because he was pointing a gun in my face,” the contextually operative satisfaction conditions governing ‘could’ and ‘free’ are instead inherited from the superimposed non-optionality aspect of one’s agentive phenomenology. (Normally, the latter contextual setting of the implicit contextual parameters governing the concept of free agency would be the more appropriate in addressing the question of moral responsibility.)

Second, the metaphysical dimension. This mainly concerns (i) the question of what constitutes genuine free agency, and (ii) the question whether humans actually exercise free agency. Question (i) is tightly interconnected with the conceptual dimension, because *prima facie* this question is essentially equivalent to the question of what constitutes the satisfaction conditions of the concept of free agency. And, since there is clearly some degree of context-sensitivity in the workings of the concept of free agency, an important special case of (i) is the question of what constitutes genuine free agency of the kind

typically regarded as a (defeasibly) necessary condition for moral responsibility. This question, which obviously incorporates the moral dimension, also is tightly interconnected to the conceptual dimension, because *prima facie* it is essentially equivalent to the question of what constitutes, *in contexts of moral-responsibility assessment*, the satisfaction conditions of the concept of free agency. So, since the satisfaction conditions of the concept of free agency probably are largely or entirely inherited from the satisfaction conditions of free-agency phenomenology, these latter satisfaction conditions are highly pertinent to question (i) of the metaphysical dimension of the free will debate. (And of course, question (i) is itself highly pertinent to question (ii): whether humans actually exercise free agency depends in large part upon what *constitutes* genuine free agency.)

Third, the epistemic dimension. People at large, and the vast majority of philosophers too, routinely accept as veridical the deliverances of their own free-agency phenomenology: they believe, of themselves, that they really are exercising free agency of the kind they constantly experience themselves as exercising. In addition, people at large, and the vast majority of philosophers too, routinely assume that other persons normally exercise this same kind of free agency. These beliefs are not only deep-seated and pervasive, but they also profoundly affect the nature of human social intercourse: they are presupposed when humans experience self-directed and other-directed reactive attitudes like shame, guilt, resentment, and admiration, they are presupposed in judgments of moral responsibility, and they are presupposed in judgments of “basic desert” for praise, blame, and punishment. But the question arises whether people are *epistemically justified* in regarding their own free-agency phenomenology as veridical, and in thereby forming beliefs whose content is inherited from that phenomenology. And the answer to this question is apt to depend, in part, upon the answers to questions C.1 - C.3 about the content of free-agency phenomenology. The more demanding are the satisfaction conditions of that phenomenology, the greater is the looming worry that such phenomenology is illusory rather than veridical. Once again, the moral dimension is involved too. If the presuppositions behind reactive attitudes, behind judgments of moral responsibility, and behind practices of blame, praise, and punishment are epistemically unjustified, then these attitudes, judgments, and practices may well be largely—or even entirely—unjustified themselves, both epistemically and morally.

Finally, the moral dimension. I have discussed it already above, with respect to the phenomenology dimension vis-à-vis each of the three others.

4. Introspection and the Satisfaction Conditions of Free-Agency Phenomenology

For each of the respective content questions C.1- C.3 posed in section 3, there is a corresponding introspectability question from I.1 - I.3. Now, it is surely fair to say that some who ponder these content and introspectability questions will answer “Yes” both to I.2 and to C.2: they will claim that it is introspectively obvious that the free-agency phenomenology I described in section has libertarian satisfaction conditions. I myself maintain that both “Yes” answers are mistaken. However, I also believe that there is strong prima plausibility in this pair of “Yes” answers, and that this fact imposes a heavy dialectical burden on those, like me, who claim that the answers nevertheless are incorrect. I have addressed that burden in a number of prior papers, and I will return to it in later sections of this paper.

Might there be some philosophers who answer “Yes” both to I.3 and to C.3? Perhaps some compatibilists would be willing to do so (although compatibilists have tended to ignore free-agency phenomenology altogether). Speaking for myself, however, it certainly does not seem *introspectively obvious* that free-agency phenomenology has compatibilist satisfaction conditions; rather, it seems to me that the answer to question I.3 is “No.” (This leaves question C.3 open.)

Many philosophers, among the ranks of both compatibilists and hard incompatibilists, contend that libertarianism is ultimately conceptually incoherent. The claim is that a satisfactory libertarianism would have to make sense of something that is allegedly just nonsensical: viz., the idea of choices and actions that (a) are not nomologically determined by prior states of the world and yet (b) are not, to the extent of such indeterminacy, merely random either. Compatibilists who make this claim might think that it grounds a “No” answer both to question I.2 and to question C.2. But that would be too quick, I submit. For, free-agency phenomenology might have libertarian satisfaction conditions, and this might even be introspectively obvious, even if those satisfaction conditions harbor some form of incoherence and hence are impossible to satisfy.

Hard incompatibilists who claim that libertarianism is incoherent might combine this claim with “Yes” answers both to question I.2 and to question C.2. For, they might think not only that free-agency phenomenology has libertarian satisfaction conditions, but also that this is introspectively obvious—even though these conditions require something impossible, viz., choices and actions that somehow emanate

from oneself in a manner that is neither state-causally determined nor (to the extent they are not thus determined) merely random.

What about questions I.1 and C.1? Libertarians who answer “Yes” to both I.2 and C.2 might perhaps think that the answer to I.1 is “Yes,” and thus that one can ascertain the satisfaction conditions for free agency phenomenology directly by introspection. But they certainly need not think so. For, a libertarian might claim that although it’s introspectively obvious that free-agency phenomenology has libertarian satisfaction conditions, what those satisfaction conditions amount to is far from obvious. The thought is that spelling out the satisfaction conditions would require articulating a satisfying positive metaphysical account of libertarian freedom. And surely it is not obvious how to do that (or even whether it can be done at all).

Hard incompatibilists who answer “Yes” to both I.2 and C.2 are apt to think that the answer to I.1 is “Yes,” and that the right way to spell out the satisfaction conditions for free-agency phenomenology will make it manifest that those conditions are incoherent and impossible to satisfy. Such satisfaction conditions will be *defective* by ordinary standards, harboring inherent conceptual instability or incoherence. (Compare visual renderings of “impossible figures,” which fail to provide a coherent depiction of any figure because they induce partial perceptual construals of the visually presented putative object—partial construals that cannot be combined into any total perceptual construal.)

It might be thought that libertarian satisfaction conditions and compatibilist satisfaction conditions, for free-agency phenomenology and/or for the concept of free agency, would be directly incompatible with one another—and that this would constrain the range of eligible combinations of answers to questions C.1 – C.3 and I.1 – I.3. In principle, however, a philosophical position concerning free agency might somehow invoke both kinds of satisfaction conditions—a possibility I discuss in section 5.3 below.

My own view is this: the correct answer to each of questions I.1 – I.3 is “No,” the correct answer to C.2 is “No,” the correct answer to C.3 is “Yes” (although the case for this must be primarily indirect and abductive, rather than introspective), and addressing question C.1 is an important task for compatibilists to undertake. More on all this below.

I have not canvassed all the conceivable, or potentially credible, answers and answer-combinations that might be given to questions C.1 – C.3 and I.1 – I.3. But enough has been said in this

section, together with what was said in section 3, to provide some initial feel for various respects in which the phenomenology of free agency is potentially relevant to the free will debate. In my view, questions I.2 and C.2 together constitute a dialectically crucial focal point. What the comparative pros and cons are, for the various competing positions in the debate, depends heavily on whether or not “Yes” is the correct answer to both I.2 and C.2. I will call the conjunction of “Yes” answers to these two questions, the thesis of *libertarian introspectability*.

5. Embracing Libertarian Introspectability: Three Package-Deal Positions

It is hard to deny that the thesis of libertarian introspectability has considerable *prima facie* plausibility. After all, since one experiences one’s free actions and free choices as emanating from oneself as their source, and one does not experience them as being state-caused, it is very natural to think that the intentional content of one’s free-agency phenomenology is as-of being a godlike “unmoved mover” of the sort posited by the libertarian conception of free agency—an unmoved *self*-mover. Likewise, since one experiences one’s free actions and free choices as emanating from oneself in such a way that one could have chosen otherwise, it is very natural to think that this core-optionality aspect of one’s phenomenology is as-of one’s choosing and act-initiating taking place outside of the state-causal nexus—again in conformity with the libertarian conception of free agency.

Initially plausible though this seems, I claim that it is mistaken. In my view, introspection alone cannot reliably yield accurate answers to questions about the satisfaction conditions of free-agency phenomenology. Although the *existence* of such phenomenology, with its intimately intertwined aspects of self-sourcehood and core optionality, is introspectively self-evident, the intentional content of this phenomenology is not really self-evident at all. More on this in section 6.

First, however, it is important to consider the consequences, concerning the principal competing positions on the free will issue, of embracing the libertarian introspectability claim. That is the business of the present section. I will consider package-deal versions of libertarianism, hard incompatibilism, and compatibilism, respectively. Each position I describe will be a package deal in this sense: it acknowledges the richly intentional phenomenology of free agency, and it asserts the introspective self-evidence of the claim that free-agency phenomenology has libertarian satisfaction conditions.

Each of the three package-deal positions will turn out to have important theoretical liabilities in comparison to one or both of the other two. The liabilities, in combination with the reasons I describe (in section 6) for *doubting* libertarian introspectability, will provide strong abductive grounds for the claim that libertarian introspectability is *false*—and also for the claim that introspection alone cannot reliably answer any of questions C.1 – C.3. And thereafter the liabilities, in combination with this claim about the limitations of introspection, will contribute substantially to the overall abductive case for a package-deal version of compatibilism.

5.1. *Package-Deal Libertarianism*

Package-deal libertarianism (henceforth, *PD* libertarianism) makes the following claims. Free-agency phenomenology has libertarian satisfaction conditions, and the concept of free agency has libertarian satisfaction conditions that are inherited from those of free-agency phenomenology. Genuine free agency requires one to be a godlike self-source of one’s choices and actions, which entails that these choices and actions are neither state-causally determined nor (to the extent that they are not thus determined) merely random either. Normally, genuine free agency also requires the ability to do otherwise, in a sense that is incompatible with state-causal determinism—although this requirement is defeasible, because of the conceivability of Frankfurt scenarios.¹¹ Moral responsibility for one’s actions requires that they be freely performed, and so does the aptness of reactive attitudes, of attributions of “basic desert,” and of basic-desert based praise, blame, and punishment. People believe they exercise free agency primarily because they experience themselves as doing so; moreover, their free-agency experience provides strong epistemic justification for this belief. And indeed, people often do exercise free agency.

Given the assumption of libertarian introspectability, *PD* libertarianism fares well in one way with respect to the conceptual dimension of the free will debate: it smoothly integrates the content of the concept of free agency with the content of free-agency phenomenology, and it plausibly treats the former as derived from the latter. But it fares badly in another way, because it is notoriously difficult to give any satisfying positive account of what libertarian self-sourcehood amounts to, over and above the mere absence of state-causal determination; such putative self-sourcehood is *at least* deeply mysterious, and perhaps is downright conceptually incoherent.

PD libertarianism fares well in one way with respect to the metaphysical dimension: it honors the strong belief, held by almost everyone (including most philosophers), that free agency is a genuine phenomenon. But in other ways it fares badly. First, it conflicts with a very credible scientific hypothesis about human cognition—viz., that the cognitive processes leading to choices and actions are implemented by brain processes that either are deterministic at the neurochemical level of description or are at least near-deterministic—with any quantum-level indeterminacies, if such there be, being “drowned out” rather than affecting the behavior of neurons.¹² Second, libertarian self-sourcehood is very probably conceptually incoherent—in which case its actually occurring is impossible.

PD libertarianism fares well in one way with respect to the epistemic dimension: it acknowledges the fact that people believe they exercise free agency largely because of the experiential character of their own free-agency phenomenology, and it honors the thought that this phenomenology renders the belief in free agency epistemically well justified. But in another way it fares badly: because of the very real scientific possibility that brain processes are deterministic or near-deterministic at the neurochemical level of description, and the very real possibility the very notion of libertarian self-sourcehood is ultimately incoherent, the hypothesis that humans exercise libertarian free agency lacks strong epistemic justification *on balance*—even granting that people experience themselves as exercising libertarian free agency. Indeed, the hypothesis that humans *do not* exercise libertarian free agency *has* strong epistemic justification on balance, notwithstanding the character of free-agency phenomenology. If free-agency phenomenology really has libertarian satisfaction conditions, then the possibility that such experience is non-veridical and illusory looms very large indeed—so large that it undermines the epistemic justification of belief in the reality of free agency.

PD libertarianism fares well in one way with respect to the moral dimension: it acknowledges the fact that an act’s being the product of free agency is a prerequisite for holding the agent morally responsible for the act, for having reactive attitudes directed at the agent because of the act, and for regarding the agent as having basic desert that warrants blame and punishment. But in another way it fares badly: since belief in libertarian free agency is epistemically unjustified on balance, so likewise are the reactive attitudes, and the associated social practices, that presuppose that people actually exercise free agency.

So although PD libertarianism is certainly theoretically attractive in some respects, especially so given the libertarian introspection introspectability claim, its theoretical costs are very high indeed.

5.ii. *Package-Deal Hard Incompatibilism*

Package-deal hard incompatibilism (henceforth, *PDH* incompatibilism) makes the following claims. Free-agency phenomenology has libertarian satisfaction conditions, and the concept of free agency has libertarian satisfaction conditions that are inherited from those of free-agency phenomenology. Genuine free agency requires one to be a godlike self-source of one's choices and actions, which entails that these choices and actions are neither state-causally determined nor (to the extent that they are not thus determined) merely random either. Normally, genuine free agency also requires the ability to do otherwise, in a sense that is incompatible with state-causal determinism—although this requirement is defeasible, because of the conceivability of Frankfurt scenarios. Moral responsibility for one's actions requires that they be freely performed, and so does the aptness of reactive attitudes, of praise and blame, of attributions of “basic desert,” and of basic-desert based punishment. People believe they exercise free agency primarily because they experience themselves as doing so; moreover, they regard their free-agency experience as providing strong epistemic justification for this belief. But that belief is both false and, on balance, epistemically unjustified. It is false because it conflicts with the scientifically well warranted hypothesis that the brain processes that implement human cognition are deterministic or near-deterministic at the neurochemical level of description, and also because a choice or act, to the extent that it is not state-causally determined to occur, is to that extent merely random rather than being genuinely free.

To a large extent, the various theoretical strengths and weaknesses of *PDH* incompatibilism are a mirror image of those of PD libertarianism. Given the assumption of libertarian introspectability, PD libertarianism fares well in one way with respect to the conceptual dimension of the free will debate: it smoothly integrates the content of the concept of free agency with the content of free-agency phenomenology, and it plausibly treats the former as derived from the latter. (In this regard *PDH* incompatibilism fares the same as PD libertarianism.) But in another respect, *PDH* incompatibilism does better: it avoids the need to render intelligible the *prima facie* incoherent notion of libertarian free agency.

PDH incompatibilism fares well in one way with respect to the metaphysical dimension: it comports with the very credible scientific hypothesis that the cognitive processes leading to choices and actions are implemented by brain processes that are deterministic or near-deterministic at the neurochemical level of description. But in another respect it fares badly: it fails to honor the strong belief, held by almost everyone (including most philosophers), that free agency is a genuine phenomenon.

PDH incompatibilism fares well in one way with respect to the epistemic dimension: it honors the fact that on balance, the hypothesis that humans exercise libertarian free agency lacks strong epistemic justification, and the further fact that on balance, the hypothesis that humans lack libertarian free agency has strong epistemic justification. But in another way it fares badly: it is committed to the contention that among those who believe in the reality of free agency (*viz.*, almost all people), many are manifesting a serious lapse in epistemic competence by maintaining this belief. They should know better, for two reasons. First, they are informed enough about science to realize, concerning the scientific hypothesis that the cognitive processes leading to choices and actions are implemented by brain processes that are deterministic or near-deterministic at the neurochemical level of description, that this hypothesis is *at least* a live, open, and viable scientific possibility (and perhaps—as I think myself—is very likely true, given current scientific knowledge). Second, they should appreciate that it is *at least* seriously doubtful whether a conceptually coherent account of libertarian free agency can be given (and perhaps—as I think myself—is impossible).

PDH incompatibilism fares well in one way with respect to the moral dimension: it honors the very plausible contention that in order for someone to be genuinely morally responsible for an act, and in order for reactive attitudes toward the person to be morally justified, and in order for basic-desert based praise, blame and punishment to be morally justified, the person must have exercised the kind of moral agency that we humans experience ourselves to exercise. But in another respect it fares badly: it is committed to the contention that nobody is ever genuinely morally responsible in the way that confers moral justification upon reactive attitudes, attributions of basic desert, or basic-desert based praise, blame, and punishment.

So although PDH incompatibilism is certainly theoretically attractive in some respects, especially so given the libertarian introspectability claim, its theoretical costs are very high indeed.

5.iii. *Package-Deal Libertarian Compatibilism*

Package-deal libertarian compatibilism, as I will call the view about to be described (henceforth, *PDL* compatibilism), is modeled on the dual-content approach to color-presenting phenomenology and color-ascribing judgments. On one version of this approach, advocated for instance by David Chalmers (2006), both the phenomenology and the corresponding judgments have two kinds of color-content. The first kind has satisfaction conditions requiring the instantiation, on the surfaces of external objects, of sensuous, non-dispositional, color-properties whose essential nature is given by how they are experienced. (Chalmers calls this “edenic” content, and I will adopt that usage here; the idea is that, whereas such properties would be instantiated in an ideal Eden in which things around us really are just as they seem experientially, this is not so in our actual world with respect to color.) The second kind has satisfaction conditions requiring something more modest—e.g., the instantiation, by external objects, of Lockean dispositions to produce, in normal perceivers under suitably good viewing conditions, certain color-presenting phenomenology. (I will call this “Lockean” content.) On another version of the approach, which I myself have advocated in a recent paper (Horgan, in press), visual color-presenting phenomenology has edenic content but not Lockean content, whereas color-ascribing judgments, as normally deployed, have Lockean content but not edenic content.¹³ Both versions claim that edenic color-presenting content is non-veridical—a view which in effect goes back to the dawn of modern science, with Galileo and Locke. Both versions also claim that judgments with Lockean content are quite often veridical.

PDL compatibilism is modeled on this dual-content approach to color experience and color-ascribing judgments. For simplicity of exposition, I will describe a version of the view that is modeled on the second version of the dual-content approach to color.¹⁴ *PDL* compatibilism makes the following claims. Free-agency phenomenology has libertarian satisfaction conditions, but these are never actually met when humans form intentions and perform actions. Free-agency judgments, on the other hand, have compatibilist satisfaction conditions, and these are frequently met when humans form intentions and perform actions. The relation between the two kinds of satisfaction conditions is analogous in pertinent respects to the relation that obtains, according to the corresponding version of the dual-content approach to color experience and color judgments, between Lockean color-content and edenic color-content. Roughly, at least, compatibilist satisfaction conditions advert to *what typically goes on*, both

psychologically and in the neurochemical implementation of psychological processes, when someone forms an intention or performs an action in a way that would be regarded as a clear, paradigm, instance of free agency. Although such compatibilist satisfaction conditions for free-agency ascribing judgments do require the *presence* of free-agency phenomenology in an agent's decision or action, in order for that decision or action to be correctly judged as a paradigmatic case of free agency, they do not require the phenomenology to be veridical.¹⁵

Given the assumption of libertarian introspectability, PDL compatibilism fares well in one way with respect to the conceptual dimension of the free will debate: it acknowledges and honors the contention that free-agency experience has libertarian satisfaction conditions. (This is in common with both PD libertarianism and PDH incompatibilism.) It fares well in a second way too: it avoids the need to render intelligible the *prima facie* incoherent notion of libertarian free agency. (This is in common with PDH incompatibilism.) It does incur a non-negligible theoretical cost, by introducing two distinct kinds of free-agency content plus a somewhat indirect relation between the content of free-agency phenomenology and the content of free-agency judgments. But in the case of color, the analogous cost is arguably well worth paying; and an advocate of PDL compatibilism would urge that the same is true here.

PDL compatibilism fares well in one way with respect to the metaphysical dimension: it honors the strong belief, held by almost everyone (including most philosophers), that free agency is a genuine phenomenon. And it fares well in another way too: it comports with the very credible scientific hypothesis that the cognitive processes leading to choices and actions are implemented by brain processes that are deterministic or near-deterministic at the neurochemical level of description.

PDL compatibilism fares well in one way with respect to the epistemic dimension: it honors the fact that on balance, the hypothesis that humans exercise libertarian free agency lacks strong epistemic justification, and the further fact that on balance, the hypothesis that humans lack libertarian free agency has strong epistemic justification. And it fares well in another way too: it acknowledges the fact that people believe they exercise free agency largely because of the experiential character of their own free-agency phenomenology, and it honors the thought that this phenomenology renders the belief in free agency epistemically well justified.

In short, with respect to the phenomenological, conceptual, metaphysical, and epistemic dimensions, PDL compatibilism inherits the major, offsetting, theoretical advantages of both PD

libertarianism and PDH incompatibilism, while also avoiding the major, again offsetting, theoretical liabilities of both positions. So far, so theoretically attractive! But there remains the moral dimension to take into account.

Unfortunately, PDL compatibilism fares quite badly with respect to the moral dimension. The following contention is extremely plausible. (I label it “RMR” for *Requirement for Moral Responsibility*):

RMR. In order for someone to be genuinely morally responsible for an act, and in order for reactive attitudes toward the person to be morally justified, and in order for basic-desert based praise, blame and punishment to be morally justified, the person must have exercised the kind of moral agency that we humans experience ourselves to exercise.

Given RMR, PDL compatibilism is committed both to the claim (a) that moral attitudes and associated social practices that presuppose that people actually exercise free agency are epistemically unjustified, and also to the claim (b) that nobody is ever genuinely morally responsible in the way that confers moral justification upon reactive attitudes, attributions of basic desert, or basic-desert based praise, blame, and punishment. (This is a conjunction of how PD libertarianism fares badly on the moral dimension with how PDH incompatibilism does so.) Fans of PDL compatibilism might hope to fend off this very serious-looking disadvantage by trying to make plausible the following contention (which I label “HI” for *Harmless Illusion*):

HI. In order for someone to be genuinely morally responsible for an act, and in order for reactive attitudes toward the person to be morally justified, and in order for basic-desert based praise, blame and punishment to be morally justified, the person **need not** have exercised the kind of libertarian free agency that we humans experience ourselves to exercise. (The systematically non-veridical phenomenology of libertarian free agency is a harmless illusion.) Rather, it is enough that (i) the person experienced free-agency phenomenology, (ii) the person’s cognitive processes leading up to intention-formation and/or action were neurochemically implemented in the way that normally accompanies this illusory phenomenology, and (iii) the person was not subject to coercion, restraint, or the like.

Prima facie, the claim HI is extraordinarily implausible. Best of luck to the fans of PDL compatibilism in seeking to defend it anyway.¹⁶

It bears emphasis how different PDL compatibilism is, with regard to its apparent consequences regarding matters of real-life concern, from the dual-content approach to color experience and color judgment. Nothing of moral or practical importance seems to be lost by embracing the view that color phenomenology is systematically non-veridical. On the contrary, color experience is no less useful as an aid to perceptual discrimination, is no less aesthetically pleasing, and so forth even if the only real property-instantiations that it tracks are of Lockean dispositional properties rather than instantiations of the vivid, sensuous, non-dispositional, apparent color-properties that visual experience presents. The putative non-veridicality of free-agency phenomenology *matters*, in a way that the putative non-veridicality of color phenomenology does not.

So although PDL compatibilism is certainly theoretically attractive in some respects, especially so given the libertarian introspectability claim, its theoretical costs are very high indeed.

6. Against Libertarian Introspectability

In several prior papers (Horgan 2007a, 2011, 2012, Horgan and Timmons 2011) I have argued that human powers of introspection are not up to the task of reliably ascertaining, by themselves, the satisfaction conditions of free-agency phenomenology—and more specifically, that introspection alone cannot reliably ascertain whether or not free-agency phenomenology has libertarian satisfaction conditions. Here I will briefly sketch my arguments, which are further elaborated in these other papers.

First: Normal human conceptual competence is often mainly a matter of being able to correctly apply a given concept *to a concrete case*—or more precisely, do so correctly *modulo one's available evidence*. Conceptual competence alone is apt to be fairly limited as a basis for answering abstract general questions about the nature of satisfaction conditions. What typically happens is that (i) some general claim about the semantics of a given concept or class of concepts is put forward as a *hypothesis*, and (ii) various data are adduced as evidence for or against the hypothesis; in effect, the argument is that the data would be well explained (or would fail to be well explained) by the hypothesis. (Thus, in effect the argumentation is a matter of inference to the best explanation.) One especially salient form of data is spontaneous intuitive judgments about whether or not to apply a given concept to a given concrete scenario—e.g., whether or not to apply the concept *knowledge* to a given Gettier case, whether or not to apply the concept *water* to the substance XYZ on Twin Earth, and so forth.

In light of these remarks about ascertaining satisfaction conditions of concepts, consider what is apt to happen when one introspectively attends to one's free-agency phenomenology with the goal of ascertaining whether or not it has libertarian satisfaction conditions. The claim that it does is a *general hypothesis* about satisfaction conditions, notwithstanding the fact that in this case, it is a hypothesis about the intentional content of one's self-presenting phenomenology. The envisioned task is the *direct, intuitive*, formation of a judgment about this general hypothesis—just by introspectively attending to one's phenomenology and then deploying one's conceptual competence to form a judgment about it. It is unreasonable and unwarranted to expect one's capacity for concept-wielding to be that splendid when it is directed at general hypotheses concerning the intentional content of agentic phenomenology, just as it is unreasonable to expect it to be that splendid when it is directed at general hypotheses concerning the satisfaction conditions for concepts themselves. When it comes to reliability and accuracy of direct, intuitive, judgments, the forte of conceptual competence is concrete judgments about specific cases. General hypotheses about satisfaction conditions are a matter for abductive inference—even when these hypotheses concern facts about the intentional content of self-presenting phenomenal character, facts that are fully fixed by that phenomenal character itself.

Second: The idea that one can tell introspectively that free-agency phenomenology has libertarian satisfaction conditions reflects a fairly subtle form of introspective confabulation. It is one thing for agentic experience to have this feature:

(A) *Not* presenting one's behavior *as* causally determined by prior conditions.

It is quite another thing for agentic experience to have this distinct feature:

(B) Presenting one's behavior *as not* causally determined by prior conditions.

Feature (A) does seem directly accessible introspectively. (One introspectively ascertains that one's free-agency experience lacks the causal-determination aspect that is sometimes present in other experiences—e.g., an experiences of a billiard-ball collision and its aftermath.) But to think that introspection directly reveals feature (B) is to be guilty of conflating the introspectable feature (A) with the distinct feature (B).

Third: Elsewhere I have argued (i) that the concept of free agency is governed by implicit, contextually variable, semantic parameters, (ii) that everyday default settings on these parameters are inherited from the intentional content of free-agency phenomenology, (iii) that there is strong abductive evidence that free-agency phenomenology has compatibilist satisfaction conditions, and hence (iv) that

these default settings render free-agency attributions compatible with determinism. However, I have also argued (v) that explicitly posing the problem of freedom and determinism tends to drive the implicit parameters away from their default setting and toward a setting under which freedom-attributions become incompatible with determinism (Horgan 1979, 2011, Graham and Horgan 1994). Suppose that this contextualist, compatibilism-friendly, orientation is right with respect to the *concept* of free agency. What, then, should be said about attempts to introspectively ascertain whether or not free-agency *phenomenology* has libertarian satisfaction conditions? Well, when one introspectively attends to the freedom aspect of one's agentic phenomenology, while simultaneously seeking to form a spontaneous judgment concerning this matter, the implicit contextual parameters governing the free-agency *concept* are apt to be driven away from their default setting and toward an extreme, determinism-incompatible, setting. Introspective inquiry is therefore prone to the following kind of subtle error. When one undertakes to answer the question on the basis of introspective attention to the free-agency aspect of one's agentic experience, one's judgmental deployment of the concept of free agency is apt to be unwittingly infected from the very start—by the unnoticed changes in the contextual settings of the implicit parameters governing this concept. That is, one is apt to mistakenly think that it is introspectively self-evident that the intentional content of one's free-agency phenomenology is incompatible with state-causal determinism—when one's tendency to think so has actually arisen because the *judgmental* aspect of one's introspective endeavor has unwittingly been prodded into a non-default parameter-setting by the posing of the very question one is seeking to answer. One's introspective judgment thus gets perverted by an unnoticed context-effect.

7. **Package-Deal Compatibilism**

Package-deal compatibilism (henceforth, PD compatibilism) makes the following claims. Free-agency phenomenology has compatibilist satisfaction conditions, and does not have libertarian satisfaction conditions; likewise for the concept of free agency, which has satisfaction conditions that are largely inherited from those of free-agency phenomenology.¹⁷ Normally, genuine free agency also requires the ability to do otherwise—although this requirement is defeasible, because of the conceivability of Frankfurt scenarios.¹⁸ Moral responsibility for one's actions requires that they be freely performed, and so does the aptness of reactive attitudes, of attributions of “basic desert,” and of basic-desert based praise,

blame, and punishment. People believe they exercise free agency primarily because they experience themselves as doing so; moreover, their free-agency experience provides strong epistemic justification for this belief. And indeed, people often do exercise free agency.¹⁹

I myself advocate PD compatibilism. As mentioned above, I favor a version of it that also incorporates a form of contextualism about the concept of free agency, according to which this concept is governed by implicit contextual parameters that can take on unusual, limit-case, settings under which attributions of free agency take on libertarian satisfaction conditions. One source of philosophical puzzlement about whether free agency is compatible with determinism, I claim, is that the very posing of the question tends to induce the limit-case, incompatibilist, setting of the implicit parameters, without one noticing that this has happened. In most ordinary contexts, including moral-responsibility ascribing contexts, the satisfaction conditions are compatibilist rather than libertarian—and are inherited from the compatibilist satisfaction conditions of free-agency phenomenology.²⁰

How does PD compatibilism fare with respect to the five dimensions of the free will problem—viz., phenomenological, conceptual, metaphysical, epistemological, and moral dimensions? That depends very heavily on whether or not the thesis of libertarian introspectability is true. If it is, then PD compatibilism fares quite badly indeed. If it is false, however, then PD compatibilism has all the theoretical advantages of each of the three package-deal positions I canvassed in section 5, and it has none of their theoretical disadvantages. (With respect to the conceptual dimension in particular, it shares with PD libertarianism and PDH incompatibilism the advantage of smoothly integrating its treatment of the satisfaction conditions for free-agency with those for the concept of free agency—although of course its treatment of both is compatibilist rather than libertarian.)

Dialectically, the fact that PD compatibilism would fare so well if the libertarian introspectability thesis were false should figure heavily in the collective body of considerations which feed into the hopper of wide reflective equilibrium whereby one assesses—abductively—the comparative epistemic credentials of a philosophical position vis-à-vis those of the competitor-positions. In section 6 I sketched reasons for at least *doubting* the libertarian introspectability thesis. One might or might not regard those reasons as strong enough by themselves to warrant *denying* the thesis. But be that as it may, I submit that when those reasons are fed into the hopper *along side* all the ways that PD compatibilism would fare well if the libertarian introspectability thesis were false, the net import of all the pertinent considerations

together is this: the libertarian introspectability thesis is very likely false, and PD compatibilism is considerably more likely to be true than either PD libertarianism or PDH incompatibilism or PDL compatibilism.²¹

References

- Balaguer, M. (2010). *Free Will as an Open Scientific Problem*. (Cambridge, MA: MIT Press.)
- Chalmers, D. (2006). Perception and the Fall from Eden. In T. Gendler, T. and Hawthorne, J. (eds.), *Perceptual Experience*. (Oxford: Oxford University Press). 49-125.
- Chisholm, R. (1964). Human Freedom and the Self. The Langley lecture (University of Kansas).
(Reprinted in Feinberg, F. and R. Shafer-Landau, R. (eds.) (2002), *Reason and Responsibility: Readings in some basic problems of philosophy, 11th Edition*. (Wadsworth.) 492-499.
- Deery, O. (2014). Is Agentive Experience Compatible with Determinism? *Philosophical Explorations*.
COI:10.1080/1386975.2013.874495
- Deery, O. (In press). The Fall from Eden: Why Libertarianism Isn't Justified by Experience. *Australasian Journal of Philosophy*.
- Ekstrom, L. (2000). *Free Will: A Philosophical Study*. (Boulder, CO: Westview Press.)
- Fischer, J. M. and Ravizza, Mark. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. (Cambridge: Cambridge University Press.)
- Frankfurt, F. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66, 829-39.
- Franklin, C. Farewell to the Luck (and Mind) Argument. *Philosophical Studies* 156 (2), 199-230.
- Ginet, C. (1990). *On Action*. (Cambridge, UK: Cambridge University Press).
- Ginet, C. (2007). An Action Can Be Both Uncaused and Up to the Agent. In C. Lumer and S. Nannini (eds.), *Intentionality, Deliberation and Autonomy*. (Farnham, UK: Ashgate Press). 243-255.
- Graham, G. and Horgan, T. (1994). Southern Fundamentalism and the End of Philosophy. *Philosophical Issues* 5, 219-47.
- Hobart, R. E. (1934). Free Will as Involving Determination and Inconceivable Without It. *Mind* 43, 1-27.
- Horgan, T. (1979). 'Could', Possible Worlds, and Moral Responsibility. *Southern Journal of Philosophy* 17, 345-58.
- Horgan, T. (2007a.) Agentive Phenomenology and the Limits of Introspection. *Psyche* 13/2.
- Horgan, T. (2007b). Mental Causation and the Agent-Exclusion problem. *Erkenntnis* 67, 183-200.

- Horgan, T. (2011). The Phenomenology of Agency and Freedom: Lessons from Introspection and Lessons from Its Limits. *Humana Mente* 15 (Jan. 27, 2011), 77-97. Issue: *Agency: From Embodied Cognition to Free Will*.
- Horgan, T. (2012). Introspection about Phenomenal Consciousness: Running the Gamut from Infallability to Impotence. In Smithies, D. and Stoljar, D. (eds.), *Introspection and Consciousness*. (Oxford: Oxford University Press.) 405-422.
- Horgan, T. (In press). Phenomenal Intentionality and Secondary Qualities: The Quixotic Case of Color. In Brogaard, B. (ed.), *Does Perception Have Content?* (Oxford: Oxford University Press.)
- Horgan, T., Tienson, J., and Graham, G. (2003). The Phenomenology of First-Person Agency. In Walter, S. and Heckmann, H. D. (eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. (Exeter: Imprint Academic.) 323-340.
- Horgan, T. and Timmons, M. (2011). Introspection and the Phenomenology of Free Will: Problems and Prospects. *Journal of Consciousness Studies* 18, 1, 180-205.
- Kane, R. (1996). *The Significance of Free Will*. (Oxford: Oxford University Press.)
- Markosian N. (1999). A Compatibilist Version of the Theory of Agent Causation. *Philosophical Quarterly* 80, 257-277.
- Markosian, N. (2012). Agent Causation as the Solution to All the Compatibilist's Problems. *Philosophical Studies* 157, 383-398.
- McCann (1998). *The Works of Agency: On Human Action, Will, and Freedom*. (Ithaca: Cornell University Press.)
- Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. (Oxford: Oxford University Press.)
- Stephens, G. L. and Graham, G. (2000). *When Self-Consciousness breaks: Alien Voices and Inserted Thoughts*. (Cambridge MA: MIT Press.)
- Whittle, Ann. (Forthcoming.) Ceteris Paribus, I Could Have Done Otherwise. *Philosophy and Phenomenological Research*, forthcoming.

¹ This section is adapted, with some modifications, deletions, and additions, from similar sections in Horgan et al (2003) and in Horgan (2007b, 2011).

² Here and throughout I speak of ‘state-causation’ rather than ‘event-causation’. More below on my reasons for this choice of terminology. States can be short-lived, and often when they do they also fall naturally under the rubric ‘event.’

³ For discussion of a range of psychopathological disorders involving similar sorts of dissociative experience, see Stephens and Graham (2000).

⁴ This is not to deny, of course, that there is indeed a distinctive phenomenology of effort of will that *sometimes* is present in the phenomenology of doing. The point is just that this aspect is not always present.

⁵ At any rate, the agent could not have performed some alternative action, and could not have prevented her/his body from moving in the A-conforming way. Some might argue that intervention by the malevolent being would prevent that A-conforming motion from being a full-fledged *action of the agent*.

⁶ What about a Frankfurt scenario in which one knows about the malevolent being and one therefore firmly believes that one is in a situation in which core optionality is absent? Perhaps here one’s agentic phenomenology would be as-of *non*-full-fledged self-source-hood in which the core-optionality aspect is lacking. But that would be extremely unlike ordinary agentic phenomenology. Alternatively—as I myself suspect would be the case—perhaps even here the core-optionality aspect still would be present in one’s agentic *phenomenology* despite one’s *belief* that core optionality itself is absent. Compare experiences of the Muller-Lyer illusion, in which one horizontal line still looks longer than the other even when one firmly believes the two lines are the same length.

⁷ For an excellent defense of this take on the real import of Frankfurt scenarios, see Ann Whittle (forthcoming).

⁸ Perhaps they do so because they think that claim (B) provides the simplest explanation of the existence of cases where agents exercise free agency and yet lack core optionality. But a major problem with that this idea is that it fails to honor free-agency phenomenology, with its deeply intertwined aspects of self-sourcehood and core optionality. A better explanation, I submit, goes as follows. First, *paradigmatic* cases of free agency are ones that fully accord with free-agency phenomenology. (In contexts of moral-responsibility attribution, this includes not only self-sourcehood and core optionality, but also the absence of superimposed non-optionality.) Second, an exercise of agency that lacks core optionality can count as a *non-paradigmatic* case of free agency only if it is sufficiently relevantly similar, in all contextually pertinent respects, to a paradigmatic case. Third, an exercise of agency in a Frankfurt scenario would exhibit, in a context of moral-responsibility attribution, such contextually pertinent relevant similarity to a paradigmatic case of free agency; this is because the presence of the malevolent being, unbeknownst to the agent, would play no actual role whatever in the person’s exercise of agency, and instead would only constitute an (unknown) pre-empted potential cause. Hence fourth, an exercise of agency in a Frankfurt scenario would count, in a context of moral-responsibility attribution, as a non-paradigmatic case of free agency.

This explanation of Frankfurt cases not only honors free-agency phenomenology, but also undergirds in the following way my contention that core optionality is a defeasibly necessary condition for free agency: unusual cases aside, an action can count as an exercise of free agency only if it exhibits all the features of *paradigmatic* free actions.

⁹ Here and throughout, by ‘libertarian’ and ‘libertarianism’ I am referring to the position that nowadays is often called ‘agent-causal libertarianism’—and not to so-called ‘event-causal libertarianism’ (versions of which are advocated, for example, by Kane 1996, Ekstrom 2000, Balaguer 2010, and Franklin 2011), or to so-called ‘non-causal libertarianism’ (versions of which are advocated, for example, by Ginet 1990, 2007, and McCann 1998). Although these other kinds of libertarianism are sociologically important in recent and contemporary philosophy, nonetheless to my mind, insofar as they differ substantively (and not just verbally) from agent-causal libertarianism, they simply are not viable contenders to provide an account of the kind of free agency that is required for full-fledged moral responsibility. Derk Pereboom (2014) argues in support of this claim—decisively, in my opinion—in a chapter entitled “Problems for Event-Causal and Non-Causal Libertarians.” To quote Pereboom:

Critics of libertarianism have argued that if actions are undetermined, agents cannot be morally responsible for them. A classical presentation of this concern is found in Hume’s *Treatise of Human Nature*.... I think the objection in this family that reveals the deepest problem for event-causal libertarianism is what I call the *disappearing agent objection*.... [B]ecause no occurrence of certain antecedent events settles whether the decision will occur, and only antecedent events are causally relevant, *nothing* settles whether the decision will occur.... The concern raised is that because event-causal libertarian agents will not have the power to settle whether the decision will occur, they cannot have the role in action that secures the control that this sort of moral responsibility demands.... But at this point one might contend, as Ginet does, that the non-causal position fares just as well [as agent-causal libertarianism]. In his view, an agent’s agent-causing simple mental acts would have no advantage over her simply *performing* such acts Suppose that the non-causalist claimed that if a “making happen” relation is not law-governed, it does not qualify as causal.... Suppose we abandon language involving the term ‘cause’ and restrict ourselves to talk of making happen and difference-making. We can then ask the indeterminist non-causalist whether she agrees that in the case of a free action, the agent makes the action happen or makes the difference whether the decision happens. If the response is positive, then the next question to ask is whether agents making actions happen or making the difference whether actions happen reduces to agent-involving events making actions happen or making the difference whether actions happen. If so, then the problem of the disappearing agent arises again. If not, then it appears we are left with the view that agents-as-substances make actions happen or

make the difference whether actions happen. But then there would be no substantive difference between the agent causal and “non-causal” positions. (pp. 31-32, 39-43)

¹⁰ Having approvingly quoted in the preceding footnote Pereboom’s critical remarks about event-causal and non-causal libertarianism, I should acknowledge that he himself maintains that in order for a person to possess the kind of control that is presupposed and required by full-fledged moral responsibility—the kind of responsibility that includes “basic desert” for praise, blame, and punishment—the person must have, and must exercise, the kind of free agency that is posited by agent-causal libertarianism. I also acknowledge that Pereboom denies that people possess and exercise that kind of freedom, and for this reason he denies that people are ever morally responsible in this full-fledged way. However, I think that his critique of event-causal and non-causal libertarianism can be happily embraced by a compatibilist, because I think that (with one important caveat, to be described presently) an adequate compatibilist account of the agentive control required by full-fledged moral responsibility should appeal to reasons-responsive cognitive decision-making processes that operate either deterministically or near-deterministically. For a classic and powerful articulation of this line of thought, see Hobart (1934)—although I think that Hobart errs in advocating a simple, conditional, analysis of ‘can/could do otherwise’, and I also think that his dismissive treatment of alternative positions on the free will issue reflects an overly crude and flat-footed form of ordinary-language meta-philosophy, is irritatingly arrogant, and fails to recognize the true depth of the problem of free will and determinism. Now for the caveat: sometimes one freely performs a given *course-grained* action in such a way that one is totally indifferent as to which specific *fine-grained* version of that action, among several that are available, gets implemented. For instance, one grabs a specific can of peas from the shelf in the grocery store, when several other exactly similar cans are equally within reach. There need not be any *psychological* reason why one chooses that specific can, rather than any of those others. (There could still be a deterministic, sub-psychological, *cause* of one’s choosing that very can, however—a cause that is embodied in the specific way that one’s overall psychological state is neuro-physically realized on this occasion. Psychological indeterminism can co-exist with sub-psychological determinism, via the multiple neuro-physical realizability of total psychological states.) And if there isn’t any psychological reason why one chooses that specific can of peas rather than one of the others, then one’s freedom would not be undermined if one’s reaching for that particular can is purely a matter of indeterministic *chance*.

¹¹ In order to construct a Frankfurt scenario that even a libertarian could construe as one in which the agent exercises free agency, one must stipulate that the malevolent being has the capacity for foreknowledge of what the agent is going to freely choose to do, even though that choice will not be state-causally determined.

¹² I acknowledge that it is somewhat tendentious to claim, as I do, that this hypothesis is very credible. My own take on the matter is that (i) the absence of strong reasons to think that there are subatomic-physics level indeterminacies that get amplified up to the level of neuronal behavior constitutes (ii) a strong reason to think that the neurochemistry of brain processes is either deterministic or near-deterministic. For a conflicting view, see, for instance, Balaguer (2010).

¹³ People who are competent in the use of color concepts need not realize that two kinds of content are involved. Nonetheless, Lockean content comes into play under roughly the following circumstances: (a) color-properties of the kind presented in visual phenomenology are never actually instantiated in our world, and yet (b) certain instantiations of Lockean properties play the right kind of role vis-à-vis color ascribing judgments to qualify as truth-makers of those judgments.

¹⁴ Both for color and for free agency, the second version seems to me more natural and more plausible than the first. But nothing of substance will turn on that here. The main advocate of PDL compatibilism I am aware of is Oisín Deery (2014, in press), who models his account on Chalmers' version of the dual-content approach to color.

¹⁵ Similarly for the corresponding version of the dual-content approach to color: although the Lockean satisfaction conditions for color-ascribing judgments do require the presence of certain color phenomenology in the experience of a suitably well positioned perceiving agent, they do not require the phenomenology to be veridical.

¹⁶ Alternatively, they could side with advocates of PDH incompatibilism in embracing the claim that free-agency presupposing reactive attitudes, attributions of free-agency based moral responsibility, judgments of basic desert, and basic-desert presupposing blame, praise, and punishment are always unjustified both epistemically and morally. But that goes wildly contrary to the beliefs, practices, and ways of being-in-the-world of almost everyone, including most philosophers. Out of the frying pan, into the fire!

¹⁷ Two important questions for PD compatibilism to address are these. (1) Do the satisfaction conditions of free-agency phenomenology include the requirement that an agent must undergo this phenomenology itself in order to be choosing and acting freely? (2) If so, why? And if not, why not? An idea that currently appeals to me is this: They do include that requirement, and the reason why is that (a) in order to exercise free agency one must be a creature "in the space of reasons" (in Wilfrid Sellars' famous phrase), and (b) free-agency phenomenology is partly constitutive of being an agent in the space of (practical) reasons.

¹⁸ PD compatibilism is clearly committed to explaining why and how it can be true that an agent *could have chosen/acted* otherwise even if state-causal determinism is true. In my view there is significant work yet to be done here. Dispositional accounts and so-called "conditional analyses" have long seemed hopeless to me. I am unhappy with possible-worlds satisfaction conditions according to which the possible worlds that are "accessible" to a freely

choosing/acting agent include worlds in which a “divergence miracle” occurs shortly before the agent chooses/acts otherwise than how the agent chooses/acts in the actual world. I am even more unhappy with satisfaction conditions according to which some “accessible” possible worlds are allowed to differ somewhat from the actual world at all moments in time prior to the agent’s non-actual choice/act. (These dissatisfactions arise regarding possible-worlds satisfaction conditions for *any* modal claims, under the assumption of determinism—not just modals about agency. Likewise for counterfactuals.) An idea that currently appeals to me is this: do the semantics of modals and counterfactuals in terms of “scenario-specifications” that (a) are *epistemically* possible (relative to some contextually pertinent body of background information), and (b) need not be metaphysically possible. As regards modals about human agency, some such scenario-specifications will hold fixed the whole history of the actual world prior to a given agent’s choice/act, will specify some way the agent chooses/acts that differs from the agent’s actual-world choice/act, and will also specify that there are no violations of any actual-world laws of nature. On this approach, some of the scenario-specifications that are “accessible” for a given agent in a given decision situation—perhaps all but one of them—might well fail to be metaphysically possible, even though all these scenario-specifications represent live epistemic possibilities, prior to the moment of choice, for the agent.

¹⁹ A package-deal compatibilist could perhaps also take on board the claim that free agency is a matter of *agent causation*—while yet advocating a construal of agent causation under which it is compatible with state-causal determinism. This idea, which was urged on me by Ned Markosian, seems well worth considering seriously. For a defense of the claim that agent causation theorists should be compatibilists, see Markosian (1999). For a defense of the converse claim—that compatibilists should be agent causation theorists—see Markosian (2012).

²⁰ What about unusual contexts in which (i) the question of moral responsibility is at issue (and perhaps legal responsibility too), but (ii) the question whether free will and determinism are compatible gets explicitly raised and purposefully emphasized? My contextualism predicts that those present will experience at least *some* increased tendency toward withholding attributions of free will and of moral responsibility; but it seems to me too crude to claim that a non-standard, incompatibilist, “score in the language game” *must* be produced by posing the free-will/determinism issue, or that all parties to the conversation *must* allow the contextual parameters for ‘can/could do otherwise’ to take on the maximally strict, incompatibilist, setting even in contexts where, according to contextualist compatibilism, it is not really apt. (The celebrated trial lawyer Clarence Darrow tried this rhetorical ploy in 1924, in the sentencing hearing, before Judge John R. Caverly, in the famous case of the confessed teen-age murderers Leopold and Loeb. Darrow did succeed in saving them from the gallows; Judge Caverly instead sentenced each of them to life in prison.)

²¹ For helpful discussion and feedback, my thanks to two anonymous referees and to Mark Balaguer, Oisin Deery, Yael Lowenstein, Michael McKenna, Ned Markosian, Shaun Nichols, David Pitt, Mark Timmons, Dave Shoemaker, and audiences at California State University Los Angeles, the 2013 NOWAR conference, and the 2014 meeting of the American Philosophical Association Pacific Division. At the APA meeting I gave this paper as the first Marc Sanders lecture in philosophy. I thank the Marc Sanders Foundation for that honor, and for their ongoing support of scholarship in philosophy.