8-2020

Free Will Experiences and Higher-Order Thoughts

Kyle Alan Hale

FREE WILL EXPERIENCES AND HIGHER-ORDER THOUGHTS

by

KYLE ALAN HALE

Under the Direction of Eddy Nahmias, PhD

ABSTRACT

A naturalist wanting to understand our conscious experience of free will may find it difficult to judge exactly what content is present in that experience. I provide an approach for how naturalistic theories of consciousness can be used to go about that, and then I apply David Rosenthal's HOT theory to that approach. After explaining HOT theory's framework, I review the literature on the experience of free will in order to identify which items of content of the experience are libertarian and which are compatibilist. I then use HOT theory to show how various mental states interact to produce content associated with the experience of libertarian free will. Since the naturalist can't accept libertarian metaphysics, the conclusion is that the experience of free will is illusory. I end by addressing concerns with that conclusion, including the question of why it is that we still feel free.

INDEX WORDS: free will, consciousness, HOT theory, libertarianism, compatibilism, phenomenology, content, conditions of satisfaction, free will skepticism

FREE WILL EXPERIENCES AND HIGHER-ORDER THOUGHTS

by

KYLE ALAN HALE

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2020

FREE WILL EXPERIENCES AND HIGHER-ORDER THOUGHTS

by

KYLE ALAN HALE

Committee Chair: Eddy Nahmias

Committee: Jessica N. Berry

Neil Van Leeuwen

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2020

DEDICATION

For Taylor.

ACKNOWLEDGEMENTS

Thank you to Eddy Nahmias for talking through all permutations of this argument, as well as Jessica Berry and Neil Van Leeuwen for their keen comments. I'm also indebted to Nathan Palmer for our many conversations on these and other topics.

TABLE OF CONTENTS

LIST OF TABLES

1 INTRODUCTION

We continuously experience ourselves as freely choosing to act. While I am out for a walk, for example, I may come upon a path that splits off down a hillside. The choice to walk down that new path has a certain feeling to it, the feeling of the choice itself, in addition to the feeling of the steps or the path beneath my feet. Furthermore, if I choose to kick a rock and send it rolling down the hill in front of me, the relationship between that choice and the movements of my leg feels different to me than the relationship between my shoe and the movements of the rock. What is that difference? Is it that I choose my movements freely while the rock's are caused by an outside force, my shoe? If so, what does it mean to choose them freely? Or are my movements caused by outside forces like the rock's are, such that I do not choose freely at all? If I don't have such freedom, why do I still feel like I do?

These phenomenological questions are about the *content* of a conscious experience. Content is related to, but not the same as, the objects of our experience. In the simple example of a visual experience, if I see a rock in front of me, the phenomenal content is not the object, the rock itself. Rather, the content is my perception of size, color, and shape, all of which are a part of my visual experience that there is a rock in front of me. As this suggests, the experience of seeing a rock involves not just the perception of a rock as a whole but also of its size, color, shape, etc. Thus we can speak of an experience's content as a whole as well as its discrete items of content. Furthermore, a conscious experience has *conditions of satisfaction* (Searle, 1983) which are associated with the content of the experience: if an item of content (such as the visual experience of grayness) corresponds to a feature of an object in the actual world (such as the color of a rock in front of me), that condition is satisfied; otherwise, it is not satisfied. Thus, any conscious experience can go one of two ways: if its conditions of satisfaction are met, it is a

veridical experience; if they are not met, it is an illusory experience. A veridical experience is one in which the phenomenological (how it appears) matches the metaphysical (how it actually is).

A veridical experience of free will may also be divided into two categories: either the experience has wholly compatibilist content or it has at least some libertarian content. The compatibilists think we have free will that is compatible with a deterministic physical world (e.g., Frankfurt, 1969; Strawson, 1962), and the libertarians think we have free will that is incompatible with determinism (e.g., Clarke, 1993; Kane, 1996). This disagreement is a metaphysical one, but it has a phenomenological correlate: while some libertarians and compatibilists agree that the experience of free will is veridical, they disagree about the experience's content and conditions of satisfaction (Bayne, 2008; Horgan, 2015). Thus, they will come to different answers to the questions about my experience to choose to walk down the hill.

A naturalist asking these questions will want to explain conscious experiences in naturalistic terms, where naturalism is the metaphysical view that everything is matter and that all its interactions are due to natural laws and forces. A compatibilist with this naturalistic motivation might say that some libertarian explanations of conscious experiences make metaphysical commitments that can't be accounted for naturalistically, and then go on to try to give such a naturalistic explanation. One thing that the compatibilist can do to provide this explanation is to reject some of the phenomenal content that the libertarian claims to be necessary to the experience of free will. However, this rejection is arbitrary unless there is a standard for judging what content is actually present in the experience of free will and what is not. In order for a naturalist to judge *what* content is present, they need two things: (a) a

naturalistic theory of consciousness that lets them say *when* an item of content is present in an experience, and (b) a list of candidate items of content.

This sets up the first two of three claims of my overall argument:

C1: Given a sufficiently thorough list (b) of candidate items of content of the experience of free will, any theory of consciousness meeting the criteria of (a) can be used to produce a sub-list (c) of actual items of content in the experience of free will.

C2: Sub-list (c) is composed in part of those items of content that the compatibilist rejects, and so the experience of free will has libertarian conditions of satisfaction.

For (a), the theory of consciousness, some naturalists will be inclined to turn to David Rosenthal's Higher-Order Thought (HOT) theory (Rosenthal, 1986, 1997, 2005). It meets the criteria; it is friendly to naturalistic accounts of the world (Rosenthal, 1986, p. 339), and its framework allows one to reason about the phenomenal content of one's conscious experiences. A full defense of C1 would require considering many competing theories of consciousness, but that is outside the scope of my overall argument here. I will instead be defending a narrower version **C1′** where HOT theory is provided for (a). While I do think that the broader claim C1 can be defended using other theories of consciousness, the choice of HOT theory as a starting place for my argument is not arbitrary. In addition to meeting the criteria above, it has strong empirical support relative to other theories of consciousness. I will therefore outline both the explanatory framework that HOT theory provides and the empirical research supporting its use.

To produce (b), the list of candidate items of content, I will consult some recent literature on the phenomenology of free will and agency. In order to provide as thorough a list as possible,

this literature will include both libertarian and compatibilist phenomenological accounts. The literature focuses on two features of the phenomenology: the agent as cause, and the ability to do otherwise. There are candidate items of content for both of these features, some of which are associated with libertarian conditions of satisfaction and some compatibilist. I will filter each candidate item of content through the perspective of HOT theory to produce (c) the list of items of content that are actually present in the experience, and will then show that some of them are associated with libertarian conditions of satisfaction. C2, then, will be defended in light of C1′, by using HOT theory.

It may be surprising that a naturalistic theory of consciousness predicts conditions of satisfaction that, were they met, would require metaphysical commitments that can't be accounted for naturalistically. This outcome of C1 and C2 leads to my third claim:

C3: Because theory (a) predicts libertarian conditions of satisfaction (as claimed in C1 and C2), but the criterion that (a) is naturalistic entails a naturalistic metaphysics such that those libertarian conditions of satisfaction cannot be met, (a) predicts that our experience of free will is not veridical, but illusory.

This claim is a variety of *free will skepticism*, the thesis that we lack metaphysical free will. My overall argument and this skeptical conclusion will help to answer some of my initial phenomenological questions. Defending C3 in light of C1′ establishes that no, I do not in fact choose freely. And why is it that I still feel like I choose freely? That question will be answered by producing the actual items of content in the experience of free will in the course of using HOT theory to defend C1′ and C2. My skeptical conclusion also invites new questions, such as

what it means to be a moral agent who is not free. I will end by addressing such questions and other possible concerns with my argument and its skeptical conclusion, as well as considering advantages of taking it up.

2 ROSENTHAL'S HOT THEORY OF CONSCIOUSNESS

HOT theory provides explanations of the mental states involved in phenomenal conscious experience, as well as a detailed account of the mechanisms involved in making mental states conscious and introspecting about those conscious states. In order to apply this theory to particular conscious experiences, such as the experience of free will, it's important to understand the terminology and tools of its framework for explaining conscious experiences. It's also important to understand the strong empirical support behind HOT theory, which motivates my use of it here to partially defend the first claim of my argument. I'll provide explanations of both HOT theory's framework and its empirical support in this section.

2.1 Understanding HOT Theory

The basic unit in David Rosenthal's theory of consciousness is a mental state such as a thought, belief, or percept. He defines a mental state as any state that has intentional properties, phenomenal properties, or both (Rosenthal, 1986, pp. 332–333). Phenomenal properties are what give certain mental states their unique experiential qualities, or what it's like to be in that mental state (Nagel, 1974). When I see a passing bicyclist with a bike whose paint reflects wavelengths corresponding to the color red, I mentally experience redness; that redness is a phenomenal property of the mental state(s) involved. Intentional properties are those of directedness or aboutness. If I have a thought directed at or about an object, that thought has an intentional property. In Searle's (1983) parlance, intentional properties have conditions of satisfaction. Just as with my discussion of conditions of satisfaction above, if I have an intentional mental state that is directed at or about an object that isn't there, the mental state fails to meet its conditions of satisfaction.

With this definition of a mental state, Rosenthal is able to define conscious states as "simply mental states we are conscious of being in." Furthermore, "our being conscious of something is just a matter of our having a thought of some sort about it" (Rosenthal, 1986, p. 335). When we have a higher-order thought about mental states in this way, we say that the higher-order thought is *representing* the other mental states.

By virtue of this representation, these higher-order thoughts, or HOTs (Rosenthal, 1997, 2005), are able to make a variety of first-order mental states conscious to us. Saying that the HOT *makes* the first-order state conscious shouldn't be taken as implying some action on the part of a HOT. As Giustina and Kriegal put it,

> in Rosenthal's theory this higher-order thought does not do anything to the lower-order state in order to make it conscious. It does not bring about any *intrinsic change* in that state that *renders* that state conscious. Rather, the higher-order thought makes the lower-order state conscious simply by *being there*. It is in this sense that consciousness is, in Rosenthal's theory, a *relational* rather than intrinsic property of conscious states. (Giustina & Kriegel, forthcoming, p. 2)

First-order mental states can be visual percepts, memories, thoughts, etc. When I see the red bike in front of me, I will have first-order mental states involving specific shades of red associated with underlying physical wavelengths of light, contours of the edges of the object, conceptual ties between those shades and contours representing certain colors and shapes, and perhaps short-term or long-term memories of experiences with other objects having similar properties. Some of these first-order states have phenomenal properties, some intentional properties, and some both. If a mental state is not represented by a HOT it is not conscious; thus

first-order states with either phenomenal or intentional properties may be either conscious or unconscious.

A single HOT can also represent groups or "bunches" of first-order states. In this way "HOTs often unify into a single awareness a large bunch of experiences" (Rosenthal, 2003, p. 328). I don't just experience a mess of lower-visible-spectrum shades of varying intensity gathered in curves against a background, I experience these various first-order states as a red bicycle. That unity of experience requires that the lower-order states which are represented by a single HOT include not just raw percepts but also appropriate concepts and memories related to past percepts. For someone who has ridden a bicycle before, the phenomenology of seeing the red bike is not just that it is a red framework of metal but that it is a thing to sit on and ride.

When states with phenomenal properties are made conscious, we have a conscious phenomenal experience; for example, if a HOT represents first-order visual percepts caused by a red object in front of me, I'll consciously experience redness. So, the content of an experience, its conscious phenomenology, is the result of a HOT representing first-order states with phenomenal properties. It may seem easy at this point to map HOTs to conscious experiences and first-order states to the content of the experiences, but this would be a mistake. Conscious experiences and their content explain mental processes at a more abstract level than a theory of consciousness such as HOT theory does. That is, HOT theory is a more fine-grained explanation of mental processes, in that a conscious experience and its content are a matter of a HOT representing first-order states. Saying that a conscious experience has certain phenomenal content is saying that first-order states with certain phenomenal properties are made conscious by a HOT's representation. This relationship between content and first-order states is what allows

HOT theory to serve as our theory (a): phenomenal content is present in an experience when first-order states with corresponding phenomenal properties are represented by a HOT.

In addition to the HOTs I've described so far, we can have even higher-order thoughts when we introspect. During introspection, a third-order HOT represents the second-order HOT (Rosenthal, 1997, p. 730). While a second-order HOT makes me aware of being in a mental state, a third-order introspective HOT makes me aware that I am aware of it. That third-order HOT makes the second-order HOT conscious to me, including its representation of first-order HOTs and their properties and relations. HOT theory thus clearly distinguishes between our in-the-moment stream of consciousness and the more focused introspective consciousness that we sometimes employ (Rosenthal, 1986, pp. 336–337).

An important point that this account of introspection brings out is that all mental states are unconscious without an accompanying HOT. A first-order state alone is unconscious until it is represented by a HOT, and when the second-order HOT represents its first-order states, it is making them and not itself conscious. That second-order HOT itself is unconscious until it too is represented by a HOT, as during introspection (Rosenthal, 1986, pp. 337–338). This is another reason why HOTs can't be mapped to conscious experiences: HOTs are not usually conscious.

This economy of conscious states makes sense if we consider the plethora of first-order mental states that must guide our everyday actions. While some actions may require that we be conscious of them, we also get a lot done by having things offloaded to unconscious habit or instinct; walking or riding a bike is a largely unconscious action, leaving us able to consciously enjoy our surroundings as we do so. Similarly, it is impossible to be introspectively aware of everything we are conscious of, and so not all of our second-order HOTs can be conscious at once. Only by employing introspection and its third-order HOTs are we aware of those second-

order HOTs. This tendency for HOTs to be unconscious is what gives consciousness its immediate, transparent quality: the mental mechanisms producing the conscious state are themselves usually unconscious, and we are normally only aware of that immediate, flowing stream of consciousness, the consciousness of first-order states.

The taxonomy of HOT theory, then, includes:

- mental states: those states that are intentional and/or phenomenal
- first-order state: a lowest-order mental state such as a visual percept or a memory
- conscious states: mental states that are represented by higher-order thoughts
- higher-order thought (HOT): a mental state at a higher order than the state (or group of states) it represents and thereby makes conscious; usually itself unconscious
- introspective thought: a HOT representing another HOT (or group of HOTs) thereby making the latter conscious

2.2 Empirical support for HOT theory

There are many theories of consciousness, some naturalistic and some not. While there are likewise many conceptions of naturalism, I've taken naturalism to be the view that all of reality is the interaction of matter according to natural laws and forces, that "reality is exhausted by nature" (Papineau, 2020). Non-naturalistic theories might try to explain consciousness in terms of special mental substances, but the kind of naturalist I've described would likely deny this and say that there are only physical substances, and so mental states must be described entirely in physical terms. Theories that adopt this physicalism about consciousness generally aim to explain mental states in terms of brain states. There are several theories that do this with

varying degrees of success, where we measure that success by the confirmation of experimental studies of the brain. While there have been debates about exactly what the empirical data means for HOT theory and its competitors,[1] and much more needs to be done to design empirical studies that will test these theories, there is work that strongly supports HOT theory as the leading theory among naturalistic theories of consciousness.

I'll focus especially on the work of Hakwan Lau and his collaborators. Lau and Rosenthal co-authored a study (Lau & Rosenthal, 2011) that compares a variety of experiments which distinguish HOT theory from its competitors. The study assumes, first of all, that higher-order representations are processed by the prefrontal cortex, while first-order states are the domain of lower-level processes, such as the processing of visual inputs by the visual cortex. These experiments then measure prefrontal activity in scenarios where conscious awareness is required but successful task performance is not, "because these cases most crucially distinguish the higher-order view from its alternatives" (Lau & Rosenthal, 2011, p. 371). For example, first-order theories say that the presence of first-order states are sufficient for consciousness. We know experimentally that successful task performance is possible without the help of the prefrontal cortex, by relying on lower-level processes and their first-order states alone. So, one possible prediction of first-order theories is that conscious awareness goes hand in hand with successful task performance. However, there is also strong evidence that conscious awareness is associated with prefrontal activation. This suggests that consciousness should not be required for successful task performance. This is better explained by HOT theory's reliance on higher-order representations for consciousness (Lau & Rosenthal, 2011, p. 366–367). So, while they encourage additional work to verify their conclusions, the authors find the experimental data to

[1] See especially the 2011 exchange between Block and Rosenthal cited below, which is also a good summary of some general objections to HOT theory and Rosenthal's responses to those objections.

support HOT theory's prediction (Lau & Rosenthal, 2011, p. 371). In a similar manner, Joseph LeDoux and Richard Brown (2017) find an adaptation of HOT theory to better support LeDoux's empirical work on emotion than competing theories.

Lau and Brown (2019) address one of the primary criticisms of HOT theory which is that it fails to account for scenarios where a subject has a conscious experience without corresponding first-order states, a logical possibility under the theory. First-order theories, on the other hand, are argued to avoid this problem, since under those theories conscious experiences consist only of first-order states (Block, 2011b, 2011a; Rosenthal, 2011). In response, the authors cite clinical cases involving irregular visual experiences, such as the Rare Charles Bonnet Syndrome cases in which the subject has visually phenomenal experiences in the form of hallucinations but without an intact primary visual cortex. Because the competing first-order theory might rely on first-order mental states corresponding to an intact visual cortex, it can't account for such experiences. While it may seem counterintuitive to say that a HOT can represent a first-order state that is not present, the authors consider a variety of ways it can indeed represent such a state, and conclude that "a version of the standard higher-order approach should be considered less problematical than the first-order view" (Lau & Brown, 2019, p. 25).

Finally, Brown, et al. (2019) give an even deeper look at the relationship between HOTs and the prefrontal cortex, as well as describing how HOT theory offers a good middle ground between theories that they think are too conceptually simple or too complex compared to what we know about the brain's processes. On the too-simple end are first-order theories; on the too-complex are theories like Global Workspace Theory (GWT). GWT says that conscious experiences are a matter of first-order states being globally broadcasted, where that "conscious sensory content […] is distributed widely to a decentralized 'audience' of expert networks" such

as memory, cognition, etc. (Baars, 2005, p. 47). This view of consciousness requires then that successful task performance be linked to consciousness in the form of global broadcasting of the first-order states in question. There are empirical considerations that make this view problematic. For example, the authors discuss blindsight patients, who are patients that have no visual phenomenal experiences but are still able to reliably detect visual stimuli. If GWT were correct, the first-order visual stimuli used to successfully complete the task would be globally broadcast, creating a conscious experience (Brown, et al., 2019, p. 8). Since such experiences do not occur for blindsight patients, GWT cannot be correct. HOT theory on the other hand is able to explain such phenomena since successful task performance and conscious experience are decoupled.

This empirical support gives us reasons to further develop HOT theory, as well as reasons to feel encouraged to apply it to specific areas in consciousness research. That is what I'll be doing here, by using it as the theory of consciousness (a) that allows me to examine the content of the experience of free will. I'll do so by going through a list (b) of candidate items of content and saying which content is compatible with the structure of consciousness as given by HOT theory. My next step, then, will be to gather that list of candidate items of content.

3 THE FEATURES OF FREE WILL EXPERIENCES

Thus far I have been assuming that our phenomenology of free will involves a rich variety of content for us to introspect and reason about. That makes my account an example of what Tim Bayne calls a liberal account of the phenomenology of agency. A strictly conservative account, on the other hand, says that the experience of agency lacks the content to support this richer phenomenology, but instead only includes minimal content related to an agent's action (Bayne, 2008). I want to not only assume but also defend the liberal position. So, one way to read my overall argument is that it is an argument in favor of a liberal account of agential phenomenology, that regardless of the metaphysical status of free will, our experience of free will does involve rich content.

To provide this account, I'll first look to the recent literature to identify two primary features of the free will experience. I'll borrow their names from Nahmias, et al. (2004), although they intersect closely with the distinctions drawn by Terry Horgan (2015). The first is the idea that we as agents are the causal source of our choices, and the second is our ability to do otherwise when we choose to act. Both libertarians and compatibilists offer interpretations of each feature, and both propose different phenomenal content for each feature. So, I'll first describe each feature in neutral terms that members of both camps generally agree with, and then I'll offer both a libertarian and a compatibilist interpretation of the content of the feature at hand. The presence of the content given by these interpretations will then be considered in §4.

3.1 Agent as cause

When you choose to do something, you experience yourself as doing it; when you move, your experience is of "your arm, hand, and fingers as being moved *by you yourself*," as Horgan puts it (2015, p. 35). That is, you experience yourself, the agent, as the cause of your choice to

move. This feature includes the perception of our bodies, but also of other parts of ourselves, including our memories, dispositions, or whatever else we take to be required for defining selfhood. We experience ourselves and our choices as involved in the causation of the action in some way, and so we experience ourselves as a part of the action (Bayne, 2008).

I said above that some libertarians are seen as making metaphysical claims that can't be accounted for naturalistically. This is especially true of *agent-causal* libertarians, and this is the type of libertarianism I'll be addressing unless otherwise noted. This libertarian has a richer, more unitary view of the self compared to the compatibilist, and attributes to that self strong powers of causation. In keeping with Bayne's (2008) distinction between liberal and conservative agential phenomenology, I'll refer to the phenomenally richer self as a *liberal self*. In this view, we have "an experience of the self as a cause, where the causal role of the self is not to be understood as derivative on or reducible to the causal role of the self's mental states" (Bayne, 2008, p. 193). This experience of the causal role of the agent includes the sense that the agent possesses the almost godlike powers of being an uncaused cause, that "freedom consists in being an undetermined determinant of one's action" (Clarke, 1993, p. 194). One item of content of the libertarian experience, then, is that the agent is a liberal self. Another is that this self seems to have a primary causal role. Furthermore, that the agent's experience of having free will includes the impression of initiating a new causal chain of events in a way that couldn't have been caused or predicted before the agent chose to act.

Horgan says that "such terminology seems *phenomenologically* apt regardless of what one thinks about the intelligibility and credibility of metaphysical libertarianism" (Horgan, 2015, p. 36). This apparent aptness makes it difficult for the compatibilist to offer a satisfying interpretation of the phenomenology involved. One compatibilist approach is to "analyse free

actions (roughly and with caveats) as actions appropriately caused by the agent's beliefs and desires" (Nahmias, et al., 2004, p. 167). These beliefs and desires of the agent would then be experienced as having the primary causal role, rather than the agent themselves. The compatibilists can thereby put the burden back on the libertarian "to show that we actually have a 'thick' experience of ourselves as agent-causes that goes beyond our experience of our mental states causing our decisions and actions" (Nahmias, et al., 2004, p. 167). This "thick" experience is the experience of a liberal self, and so this compatibilist response is another form of the conservative vs. liberal debate about the content of free will experiences.

A more liberal compatibilist response would be to accept aspects of libertarian phenomenology of the self but reject libertarian metaphysics. For example, what we might call "liberal self skepticism" would allow the compatibilist to accept that we experience a liberal self even though the conditions of satisfaction of that experience are not met. This would still allow the compatibilist to reject other aspects of libertarian phenomenology, and to reject free will skepticism. In turn, a compatibilist's acceptance of liberal self phenomenology does not pose a problem for my defense of libertarian conditions of satisfaction and of free will skepticism, as I'll explain in §4–5.

3.2 Ability to do otherwise

When we make a choice, we experience ourselves as choosing among more than one option. Included in that experience is the notion that we could have done otherwise than we have done. In Horgan's (2015) words, our experience contains a "core optionality." This feature is also described as "openness" about our future in the moment before and during our action (Deery & Nahmias, forthcoming). Robert Kane has used Borges's (1944/1993) imagery to describe this openness as a "garden of forking paths" from which we choose (Kane, 2005, p. 7).

A libertarian might describe this experience of openness as one in which "an action is free only if, *given all conditions as they are at and up until the moment of choice*, the agent is able to act or choose in more than one way" (Nahmias, et al., 2004, p. 165). This description introduces two items of libertarian content: the feeling that we have the ability to choose from among more than one option, and the feeling that our choice could be different even if all conditions are held fixed right up until we choose. Furthermore, one has the sense that it is "up to oneself whether or not one does the thing one does" (Clarke, 2019, pp. 756–757). This feeling of the choice being up to oneself is a third libertarian item of content for the feature of the ability to do otherwise. So, the experience of libertarian free choice is that it is "*up to you* whether or not to perform" the chosen action, such that "you *could have done otherwise*" (Horgan, 2015, p. 36, see also 2007b; Kane, 1996, p. 44ff.). Some studies on folk intuitions, the intuitions that everyday non-philosophers have about free will experiences, support such interpretations of this feature (see, e.g., Deery, et al., 2013, p. 143).

The compatibilist disagrees about the content associated with the feature of the ability to do otherwise. For example, one might instead offer a conceptual interpretation of the word 'ability' in terms of dispositions (Deery, et al., 2013, p. 144; Vihvelin, 2004). A compatibilist with this dispositional view might claim that there is an item of content for the sense that we have a disposition to do otherwise than we have done. Compatibilists might similarly say that the feeling of our choices being up to us is actually the feeling that we're appropriately "responsive to reasons" for making the choice (Nahmias, et al., 2004, p. 165n5). Finally, they might take the conservative stance toward the phenomenal content and deny the libertarian claim that this feature requires the experience of all conditions being held fixed until the agent has made the choice (Nahmias, et al., 2004, p. 166).

4 HOT THEORY AND LIBERTARIAN PHENOMENOLOGY

In order to better analyze the findings of the literature I've reviewed, I've organized them into table form, where the rows are the items of content of the two features of the experience of free will, and the columns correspond to the libertarian vs. compatibilism interpretations of each item of content:

Table 1: Features of the experience of free will

	Libertarian	**Compatibilist**
Agent as cause	Agent is a liberal self	Agent is either a liberal or non-liberal self
	Agent has primary causal role	Agent's beliefs and desires have primary causal role
	Agent is an uncaused cause	No content of being an uncaused cause
Ability to do otherwise	Choice is up to the agent	Choice is a response to agent's reasons
	Agent has ability to choose among multiple options	Agent has disposition to do otherwise
	Choice could be different even if conditions hold fixed	No content of choice being able to be different

There are a few things to note about these items of content and the way I've structured them. First, the compatibilist column contains fewer claims of candidate items of content. This relative sparseness isn't surprising given the compatibilist's tendency to take up a more

conservative stance about content. Having more items of content does, however, provide more opportunities for the libertarian to show that an item of content is present.

In order to use HOT theory to demonstrate that the experience of free will has libertarian conditions of satisfaction, I don't need to show that there are no compatibilist items of content present. The libertarian would likely agree that at least some of the content in the compatibilist column is present in the experience of free will, but they would hold that none of the content in the compatibilist column is sufficient to say that someone has had a free will experience. The compatibilist in turn might try to say that none of the content in the libertarian column is necessary for a free will experience. So, to defend a libertarian free will experience against such a compatibilist, I only need to show that one item in the libertarian column is present in the experience of free will.

However, the compatibilist might also accept portions of the libertarian content, such as the liberal self, but deny the rest. As such, I'll show that all six items of libertarian content are present in the experience. My case, though, will be especially underpinned by the first item in the libertarian column, the experience of being an agent as a liberal self. This is in part because there is a close relationship between the two features of the experience of free will, the agent as cause and the ability to do otherwise. As Horgan puts it, the two features are "intimately bound up" together (2015, p. 37). HOT theory will help us to see why that is. So, I'll account for the first item in the list by way of HOT theory's support for experiencing oneself as a liberal self, and then proceed to account for the rest of the content in light of those findings.

4.1 Agent as cause

Agent as a liberal self: When we choose to act we experience ourselves as being an agent, and the libertarian takes up a rich, unitary view of this agential self, which I've called a

liberal self. HOT theory provides support that we do indeed experience a liberal self. Just as a HOT is able to represent a unified conscious awareness of anything by representing a group of first-order states, it is able to represent a unified sense of self.

First, each HOT involves a "minimal concept of self" in that it represents a relationship between some concept of self and the conscious state in question (Rosenthal, 1997, p. 741). This minimal concept of self need not be very sophisticated and is compatible with the kind of consciousness we sometimes want to attribute to infants or non-human animals. On its own this minimal self doesn't give us the liberal self required by the libertarian view. When combined with other tools that HOT theory gives us, however, this minimal concept of self contributes to the experience of having that richer sense of a liberal self.

We use a variety of personal markers to identify ourselves. These self markers are based on our memories, beliefs, physical makeup, etc., and are first-order mental states with phenomenal properties that are available to conscious thought when needed. So, when I choose to go for a walk, the experience of that choice is a HOT which represents and connects a variety of first-order self markers, perhaps my past choices to go on walks, my body's energy level at that moment, memories of where I have and haven't walked before, etc. When I choose to turn down a new path on my walk, new self markers come into play, associated with the others via memory and the self markers still in play.

In other words, when a HOT represents a mental state as being conscious, that representation includes a reference to the minimal concept of self in the form of some minimal set of these self markers, but also to a larger group of whichever of these markers are relevant to the conscious experience at hand (such as choosing to go for a walk). This larger group will include concepts necessary for that liberal self, including a sense of continuity over time and

space, at least in the form of associations between our memories and our present mentally-represented bodily states. The combination of the phenomenal properties of each first-order state involved will then lend to every conscious act a rich phenomenal experience of being a self. This sense of self doesn't need to include every possible self marker, but only those relevant to the task at hand. As the task evolves or as we take on new tasks, these markers will be united with additional self markers and slough off unnecessary ones. As Rosenthal puts it,

> For any new first-person thought, the reference that thought makes to oneself is secured by appeal to what other, prior first-person thoughts have referred to, and this process gradually enlarges the stock of self-identifying thoughts available to secure such reference. (Rosenthal, 2003, p. 335)

The set of self markers can thus be seen as a continuous and evolving set, whose common members are at least those markers comprising the minimal concept of self. The inclusion of this set of markers in every conscious experience lends to us the experience of being a rich, unitary self: a liberal self.

When I make the choice, then, the HOT doesn't merely represent the choice as a conscious mental state; it represents that I am the one in that mental state, that I, as a liberal self, am the one making the choice. This description makes no metaphysical commitment about the nature of the self (Rosenthal, 2003, p. 351). In particular, it should not be taken as a claim that there is actually some simple, unitary self making the choice. In fact, it claims that the content associated with the experience of selfhood is complex, in that it is comprised of many first-order states, the self markers. What HOT theory does is explain why it is that I *experience* myself as a unitary self, and it explains why I experience *myself* as the one having my conscious experiences:

I "identify the individual to which each HOT assigns its target as being the same from one HOT to the next" (Rosenthal, 2003, p. 337).

Since all of my conscious mental states contain a reference to the minimal concept of self, and since all conscious states have access to the liberal self via the full set of first-order self markers, all of our choices involve the experience of a liberal self in a way that is consistent with the libertarian phenomenological claim. That is, this liberal self is the libertarian agent.

This means that, according to HOT theory, compatibilists who hold something like a Humean bundle theory about their experience of selfhood are misdescribing their phenomenology. However, as explained in the previous section, the compatibilist can instead accept the liberal self view while still resisting the remaining libertarian items of content.

Primary causal role: Another libertarian phenomenological claim is that this liberal self has a primary causal role, in that the *agent* causes the choice and subsequent action rather than just the agent's mental states (such as beliefs or desires). This isn't to say that mental states like beliefs and desires play no causal role, but that they only do so as a part of a liberal self.

When I choose to go for a walk, I might say that I do so purely because I want to, and that nothing else caused me to do so. However, for the libertarian it isn't sufficient to say that it is only my desire to go for a walk that causes me to do so. That desire functions only as a property of an agent, myself. It is because I as an agent desire to go for a walk that I choose to go for a walk.

Every time we consciously choose to act, the HOT making the choice conscious represents a group of mental states that also includes markers of the self, some of which are our beliefs and desires. It is never the case, however, that we make a choice wherein mere beliefs or desires are represented by a HOT in the absence of the liberal self. Since the agent's liberal self

will always be present in the experience, and since the desires belong to the agent, the agent's liberal self is experienced as being primary rather than the agent's desires. Therefore, the phenomenology will always be that of the agent as the primary cause of the choice in question.

Uncaused cause: There is an even stronger sort of content associated with the phenomenology of being an agent with a primary causal role, which is that of being an uncaused cause. To be a cause that is not itself caused is to make a break in the unfolding of the causal chain of the universe, to start a new causal chain. Prior to my kicking the rock while walking down the hill, I might imagine it to have moved from place to place only ever in response to states and events that started at the Big Bang and unfolded (at least macroscopically) deterministically since then. When I choose to kick it, a strong version of libertarianism would claim, I experience myself as starting a new causal chain of events.

HOT theory predicts that this is exactly what we would experience. In the moment of choice we are consciously aware of only those self markers required for the action at hand, including beliefs and desires immediately relevant to the choice, and assume a primary causal role for that liberal self while being unconscious of other first-order states having a causal role. In Rosenthal's words,

> Because our mental states are not all conscious, we are seldom if ever conscious of all the mental antecedents of our conscious states. And conscious desires and intentions whose mental antecedents we are not conscious of seem to us to be spontaneous and uncaused. (Rosenthal, 2003, pp. 350–351)

The resulting phenomenal content is of ourselves being an uncaused cause: our causal role is conscious for us, but any antecedent causes of that causal role are not conscious for us.

This explanation of conscious and unconscious states applies equally to any outside forces that empirical work might tell us are causally influencing our choices. If those processes are a result of a deterministic unfolding of events including my genetic history and the environment in which I was raised, then ultimately my choice was entirely determined. However, that set of deterministic influences would not be represented in my experience. Due to HOT theory's economy of conscious mental states, in order to be conscious of a choice the only states I would need to have represented are those necessary for the experience of the choice itself. In the case of going for a walk, all I need to be conscious of is the desire to go for a walk, my energy level, etc., as appropriate to the task at hand.

While this explanation covers a significant portion of the phenomenology of being an uncaused cause, it misses a nuance. As Terry Horgan puts it, "*Not* presenting one's behavior *as* causally determined by prior conditions" is different from "Presenting one's behavior *as not* causally determined by prior conditions" (Horgan, 2015, p. 56). I haven't addressed the latter. To do so will require some concepts I haven't yet touched upon, and so I will return to the defense of this item of content at the end of the following section.

4.2 Ability to do otherwise

Choice is up to the agent: With these insights into the phenomenology of agent causation, it is easier to understand the phenomenology of our ability to do otherwise. When I choose to go for a walk, I do so in part because the choice is mine, and in this way I experience the choice as up to me. The libertarian asserts that it is not sufficient to say that in choosing to go for a walk I feel like I am responding to reasons for going for a walk, where those reasons are that I want to stay fit, need some fresh air, am feeling bored, etc. For the libertarian, this responsiveness does not cover the phenomenal content of the choice being up to me.

Again, this libertarian interpretation of the content is supported by the idea of the liberal self that HOT theory predicts is involved in all conscious choices. As with beliefs and desires, when my reasons are conscious but their causes are not, I will experience my liberal self as being primary, and so I will experience the choice to act one way or another, or whether or not to act, as being ultimately up to me and not merely a matter of my responsiveness to reasons.

Ability to choose among multiple options: Although the HOT representing a choice will not encompass all of the causal factors going into the decision and all their possible outcomes, it can represent the availability of an option itself. When I come to a literal fork in the path and need to choose which way to go, it may be that the brain states corresponding to a decision to act are engaged before I become conscious that I have chosen, and then I begin walking (Rosenthal, 2003, p. 350n38). Upon walking in one direction, the HOT representing my choice will represent the option chosen, the other possible options (here, the fork not taken), and all my liberal self markers, so that my experience is of myself as having the ability to choose from among the available options. In other words, to consciously experience a choice, all we need be conscious of is the optionality and not the true causal source of the choice. This source may be the agent, some event in the agent's brain, or causally deterministic factors outside and inside the agent. It's not manifest in the phenomenology which of these is the case, but the experience of optionality is manifest.

As mentioned earlier, one way for the compatibilist to explain away the metaphysical ability to do otherwise is that we instead possess dispositions to do otherwise (Vihvelin, 2004). However, when it comes to the phenomenology, one study by Deery, et al. denies that this is how the layperson experiences choice. Rather than saying that "an ability to act is a disposition, or a bundle of dispositions," participants in the study "tended to interpret their agentive

experience in terms of an ability to do otherwise, and they interpreted that ability incompatibilistically" (Deery, et al., 2013, pp. 144–145). That subjects do not experience the ability to do otherwise dispositionally can be explained in terms of HOT theory. A dispositional account of ability does not satisfy the phenomenal experience because the items of content related to the experience of optionality must also be accounted for. I do have certain dispositions to choose amongst options, but when it comes down to the choice, those dispositions will be represented by a HOT alongside other first-order states, including the availability of multiple options as described above. Because of this representation of optionality itself, when making a choice I can't help but be presented with the phenomenal content of multiple options being available to me, and so the dispositional account alone fails to do justice to the phenomenology.

Choice could be different even if conditions hold fixed: When we choose, the libertarian claims, we can imagine that if the choice could be replayed several times over with the same input conditions, the output choice may be different from one iteration to the next. If I my entire life history, and indeed the history of the entire universe, were held fixed, I would still be able to choose something different each time. When I choose, I cause that choice, and I could have caused a different choice even with all conditions holding fixed until the moment of choice.

The ability of HOTs to represent the optionality of a choice, as described above, is an important part of this item of content. Again, the group of first-order states involved in my conscious choice does not include anything more than it needs to include, but it does include that optionality. It also includes first-order states of memories and concepts that represent my current state of affairs and life history. A HOT's representation of both this history and this optionality produces the phenomenology of a choice with perceived initial conditions that do not change to account for a difference in choice, but also the phenomenology of all relevant options being open

to me equally. At the moment of choice I resolve upon a single option and experience a change in my life history, but up until that point I experience all conditions as holding fixed.

It is this explanation, also, that allows me to continue my defense of the phenomenology of being an uncaused cause. As explained above, Horgan (2015) emphasizes the difference between not experiencing our behavior as caused, and experiencing our behavior as not caused. I can now show that both are present in the experience of free will. Because I experience conditions as holding fixed until the moment of choice, when I experience myself as being the cause of a choice between *x* and *y* I also experience myself as not being caused to choose *x* or *y*. So, the HOT involved in the choice doesn't represent first-order states associated with the prior deterministic factors that cause me to choose, which means that *I fail to experience myself as caused*. At the same time, the HOT involved represents first-order states associated with the perceived unchanging initial conditions of my choice, as well as first-order states associated with my liberal self. Since I experience the initial conditions as unchanging, I also experience myself as unaffected by them; that is, they do not cause me to do anything. I experience myself as the only determining factor in the choice between *x* and *y*, and so *I experience myself as uncaused*. Thus I am an uncaused cause of my choices both in Horgan's (2015) first and second senses.

4.3 Summary of the experience of free will

The discussion above defends my claims C1′ and C2. The output of the defense of C1′ is a list of items of content present in the experience of free will according to HOT theory. That list includes the following:

1. The agent is a liberal self
2. The agent has the primary causal role in making choices

3. The agent's beliefs and desires have a lesser causal role
4. The agent is an uncaused cause
5. A choice is up to the agent
6. A choice is a matter of responsiveness to reasons, but only as a part of the agent
7. An agent has the ability to choose among multiple options
8. A choice could be different even if all conditions hold fixed until the moment of choice

Items 1, 2, 4, 5, 7, and 8 are my six initial libertarian claims of content. There are also some compatibilist items of content are present, as expected. As I pointed out earlier, I only need to show that one of the libertarian items of content is present in the experience of free will in order to defend C2, the claim that the experience of free will has libertarian conditions of satisfaction. In addition, for the compatibilist who accepts item 1 but rejects the rest, I have also defended the presence of all six libertarian items. Therefore, I have defended C2.

At the moment of conscious choice, then, a HOT will represent a group of first-order mental states including the following: self markers relevant to the task at hand, such that I experience myself as choosing; first-order states relevant to the causation of the event, such as my beliefs and desires; visual, auditory, and other percepts relevant to the successful performance of the action. What are not rendered conscious are the following: first-order beliefs, desires, or instincts that primarily cause the choice; causes prior to those unconscious first-order states; changes in the conditions of those first-order states that lead to different choices. "The sense we have of free agency results from our failure to be conscious of all our mental states," Rosenthal says (Rosenthal, 2003, p. 351). I have used HOT theory to show that that sense we have is a libertarian sense. That is, the product of a phenomenally liberal self interacting with

these conscious and unconscious states in the way that HOT theory predicts is an experience of choice with a richly libertarian phenomenology.

5 SKEPTICISM: CONCERNS AND ADVANTAGES

The defense of my skeptical claim C3 follows naturally from the above defenses of my first two claims. We assume that we are in a world described by naturalism. HOT theory meets the criteria of (a), including being naturalistic. HOT theory is also able to be used to defend C1 and C2, which establishes that we have libertarian conditions of satisfaction, and so a libertarian phenomenology of free will. However, libertarian metaphysics do not hold in a world described by naturalism. If we have a libertarian phenomenology of free will but lack libertarian metaphysics, then we do not have free will. Therefore, C3 holds, and the experience of free will is illusory.

There have been many claims of free will skepticism, and with them come difficult questions such as those I mentioned in the introduction. The most immediate of these, if not the most critical, is the question of why it is that we still feel free. My defense of C1′ and C2 provides an explanation: HOTs represent us as being in first-order states that produce content for experiences with libertarian conditions of satisfaction. To say that we have libertarian conditions of satisfaction is to say that we have a libertarian phenomenology. To say that we have this phenomenology is to say that we feel like we are free in the liberal sense described by libertarianism.

My skeptical conclusion runs counter to both compatibilism *and* libertarianism: compatibilism in that the conditions of satisfaction are libertarian, and libertarianism in that the conditions of satisfaction cannot be met. As such, they will both offer responses to my argument.

Many libertarians will agree with my conclusion about the phenomenology but will tend to reject the naturalistic metaphysical goals of my argument. This is especially true of the agent-causal libertarian. That position will result in an impasse. For the naturalist, unless the libertarian

can give either a cohesive naturalistic metaphysics that matches the phenomenology or a stronger argument for a non-naturalistic metaphysics, my skeptical conclusion holds. Setting aside agent-causal libertarianism, there are naturalistic libertarian accounts that may be able to avoid my skeptical conclusion (see, e.g., Kane, 1996). However, there are further reasons for doubting that such accounts are able to provide us with the kind of freedom we feel like we have (see Pereboom & Caruso, 2017, p. 199).

One possible compatibilist concern is that while I have discussed free will as a matter of making choices, we in fact experience various types of choices, each with a different phenomenology. For example, Nahmias (2006) describes the process of making a choice after we've deliberated about our options. The outcome of such deliberation, he says, is either a "close call" or a "confident decision" (Nahmias, 2006, p. 629). A close call is a choice where we're torn between the options at hand, and equally desire to choose each. A confident decision is one where we have reasons to be completely confident about the option we choose. While my account is able to address close calls, the compatibilist might argue, a confident decision does not have the phenomenology I've described. In particular, "the confident agent would never feel the need for [an] unconditional sense of alternative possibilities" (Nahmias, 2006, p. 636). This is another way of saying that the experience of free will does not require the content of the ability to choose among multiple options. However, I've established in my defense of that item of content that every choice will contain the representation of optionality. Even the highest degree of confidence will not remove this optionality from the experience. This is especially clear in light of my defense of the choice being able to be different even if conditions hold fixed. The confident agent may feel like they had chosen freely even though their deliberation left them

with no other reasonable options, but such an agent would still have an experience that they could have, for example, chosen against their best reasons.

Horgan (2015) offers a compatibilist response to the idea that the conditions of satisfaction of free will experiences are libertarian. Horgan describes the *introspectability thesis*, which is that we have the capacity for introspection necessary to say what are the conditions of satisfaction for libertarian free will experiences. My argument would be an example of an account that relies on the introspectability thesis. Horgan argues against this thesis by attempting to show various ways in which our conceptual powers are not great enough to introspect about the conditions of satisfaction of free will experiences (Horgan, 2015, p. 55ff.). To explain this lack of conceptual power, he describes a geometry problem which we are supposed to see as obviously having a definite, calculable answer, but one which we can't arrive at by introspection alone. Horgan thinks that answering the geometry problem is analogous to the attempt to say what are the conditions of satisfaction of free will experiences: no matter how much we sit and think about the questions, their answers will evade us, even though there are definite answers (Horgan, 2007a, pp. 17–18). Horgan's argument might sound like the conservative stance about the content of our experiences that I described above, but it is slightly different. Rather than being an argument against the *existence* of certain kinds of content, Horgan's argument is against our *ability to introspect about the conditions of satisfaction* of the content of our experiences.

HOT theory, however, shows us that we do have conceptual powers great enough for that kind of introspection. First, it gives us a clear picture of how introspection works in terms of the interrelation of mental states. Second, it sets up that economy of mental states, wherein we only need to be conscious of those mental states that are appropriate for the task at hand. The geometry problem in Horgan's example is indeed difficult to introspect about, but that is not

what our conceptual powers evolved for. We evolved to introspect about our environment and our actions relative to it. That is, if there's anything we do have the conceptual powers to introspect about, it is our ability as agents to choose among options. Since we do have those introspective powers, Horgan's argument against the introspectability thesis fails, and what that introspection reveals, according to HOT theory, is a phenomenology with libertarian conditions of satisfaction.

But Horgan might answer that this misses the heart of his concern, which could be put like this: How is it that any theory that is developed via introspection can capture what it is to consciously experience something, including the experience of making a choice freely? Or, more to the point, let's say a compatibilist argues that their experience of free will isn't libertarian, that they just have a different experience of free will than others claim to. That is, the content of the compatibilist's free will experiences would point to a compatibilist metaphysics were they met. In that case, what reasons would the compatibilist have for thinking that a theory developed via HOT theory's version of introspection would tell them they are mistaken?[2]

The first part of the answer to that is a repetition of what I've said above: HOT theory provides a rich account of the mechanisms of our introspection, such that if we accept the theory we can trust the results of our introspection, including the account of the two features of the experience of free will given here. For the concern that the compatibilist is just experiencing something different, my answer is the one I gave at the beginning of §4. It isn't that there aren't compatibilist conditions of satisfaction in a given person's free will experience, it's just that there are additional conditions as well. The mistake of the compatibilist is not in what they have experienced overall, but in what content they have reported as being part of the experience of

[2] Personal communication with Eddy Nahmias, October 2019.

free will. Here I have given two features of the experience of free will, each with a variety of content that HOT theory requires. If HOT theory is accurate, then, it says that the compatibilist is failing to properly describe the experience of free will and its conditions of satisfaction, while the libertarian is properly describing those conditions.

Broader objections might be taken to my skeptical conclusion; more questions might arise. For example, if we do not have free will, have our past choices meant anything? How about our present or future ones? Without free will, can we have the moral status we thought we had? If not, what happens to our value system? Could we survive a collapse of our values?

These questions have an edge of anxiety, and they aren't answerable by HOT theory, or by defenses of claims C1–3. There are, however, some reasons to take up the free will skepticism without anxiety, and reasons to see that it is in fact a more desirable view of our agency.

First, I don't believe that a collapse of values is something to be worried about. It's feasible that a society aware of its lack of free will would continue to have something like Peter Strawson's reactive attitudes (Strawson, 1962), and so we just would continue to hold each other morally responsible, whether or not the theorists say that moral responsibility requires free will. Such a position would say that we're moral creatures, and will continue to be regardless of the metaphysics or our understanding of it. But it is also possible that our understanding of the lack of free will would indeed have downstream effects that alter our moral status, and I won't take on such a debate here.

There is a related argument for not being concerned about free will skepticism, however. There's a deeper kind of moral responsibility, an ultimate responsibility that is impossible to have without free will. This kind of responsibility is what Galen Strawson said would make us deserving, were heaven and hell real, of eternal blessing or torture (Strawson, 1994, p. 9f.). One

doesn't have to believe in heaven or hell to feel like some of our methods of punishment are too terrible for creatures that lack free will. An answer to this sensibility is found in Pereboom & Caruso (2017). They refer to the ultimate responsibility I've described as *basic desert moral responsibility* and define free will in terms of it. They review both compatibilist and libertarian possibilities for a metaphysical basis for that free will, and rule out each one. The remaining option, they say, is their own version of free will skepticism called *hard incompatibilism*. However, they also acknowledge that

> it is not the philosophical arguments for free will skepticism that are the problem, it is the existential angst they create and the fear that relinquishing belief in free will and basic desert moral responsibility would undermine morality, negatively affect our interpersonal relationships, and leave us unable to adequately deal with criminal behavior. (Pereboom & Caruso, 2017, p. 199)

In response to this angst, they then give us a thorough tour of what our society's justice system might look like if we lacked free will and that ultimate moral responsibility. In lieu of harsh punishments they offer a system based on medical quarantine, wherein those who have committed crimes are rehabilitated if possible. While there are many concerns about this model yet to be worked out,[3] it offers a glimpse into an improved society as a result of accepting free will skepticism.

My parting answer to concerns about free will skepticism is a rhetorical extension of the other responses I've given. Bernard Williams said that the ancient Greek tragedies and other "stark fictions" can say something about our moral status that moral philosophy can't, that there is a limit on "the tireless aim of moral philosophy to make the world safe for well-disposed

[3] See Pereboom (2017) for an attempt to defend against such concerns.

people" (Williams, 2006, p. 59). There's more than a bit of censure in this statement, since he sees much of moral philosophy as covering up, for the sake of well-disposed people, the real horrors that much of humanity faces. Moral theories tend to ignore "the very plain fact that everything that an agent most cares about typically comes from, and can be ruined by, uncontrollable necessity" (Williams, 2006, p. 54). However, I have some hope that moral philosophy can overcome this limit, by facing the reality of our situation as starkly as Sophocles. We can acknowledge that our actions and those of others are a matter of necessity, that they are determined by other factors. My solution to this limit of moral philosophy is in contrast to people like Smilansky (2000) who believe that it is critical to maintain the illusion of freedom, that without the comfort of that illusion we would buckle under the weight of reality's necessity. Reliance on that comfort is just another way of covering up humanity's suffering. While we may not be capable of countenancing all of that suffering at every moment, our other cognitive limitations, such as the economy of mental states that HOT theory describes, already take care of that for us. We're better off introspecting about reality as much as we're able to, so that we might more accurately address the moral issues at hand.

REFERENCES

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. In *Progress in Brain Research* (Vol. 150, pp. 45–53). https://doi.org/10.1016/S0079-6123(05)50004-9

Bayne, T. (2008). The Phenomenology of Agency. *Philosophy Compass*, *3*(1), 182–202. https://doi.org/10.1111/j.1747-9991.2007.00122.x

Block, N. (2011a). Response to Rosenthal and Weisberg. *Analysis*, *71*(3), 443–448. https://doi.org/10.1093/analys/anr036

Block, N. (2011b). The Higher Order Approach to Consciousness Is Defunct. *Analysis*, *71*(3), 419–431. https://doi.org/10.1093/analys/anr037

Borges, J. L. (1993). The Garden of Forking Paths. In *Ficciones*. Everyman. (Original work published 1944)

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.06.009

Clarke, R. (1993). Toward A Credible Agent-Causal Account of Free Will. *Noûs*, *27*(2), 191–203. https://doi.org/10.2307/2215755

Clarke, R. (2019). Agent Causation and the Phenomenology of Agency. *Pacific Philosophical Quarterly*, *100*(3), 747–764. https://doi.org/10.1111/papq.12275

Deery, O., Bedke, M. S., & Nichols, S. (2013). Phenomenal Abilities: Incompatibilism and the Experience of Agency. In D. Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility* (pp. 126–150). Oxford University Press.

Deery, O., & Nahmias, E. (forthcoming). Experience of Free Agency. In *Companion to Free Will*. Wiley.

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, *66*(23), 829–839. https://doi.org/10.2307/2023833

Giustina, A., & Kriegel, U. (forthcoming). Two Kinds of Introspection. In J. Weisberg (Ed.), *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*. CUP.

Horgan, T. (2007a). Agentive Phenomenal Intentionality and the Limits of Introspection. *Psyche*, *13*(1), 29.

Horgan, T. (2007b). Mental Causation and the Agent-Exclusion Problem. *Erkenntnis*, *67*(2), 183–200. https://doi.org/10.1007/s10670-007-9067-9

Horgan, T. (2015). Injecting the Phenomenology of Agency into the Free Will Debate. In D. Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility: Volume 3*. Oxford University Press UK.

Kane, R. (1996). *The Significance of Free Will*. Oxford University Press USA.

Kane, R. (2005). *A Contemporary Introduction to Free Will*. Oxford University Press, USA.

Lau, H., & Brown, R. (2019). The Emperor's New Phenomenology? The Empirical Case for Conscious Experience Without First-Order Representations. In A. Pautz & D. Stoljar (Eds.), *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness*. Retrieved from https://philpapers.org/archive/BROTEN.pdf

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

LeDoux, J. E., & Brown, R. (2017). A Higher-Order Theory of Emotional Consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016–E2025. https://doi.org/10.1073/pnas.1619316114

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435. https://doi.org/10.2307/2183914

Nahmias, E. (2006). Close Calls and the Confident Agent: Free Will, Deliberation, and Alternative Possibilities. *Philosophical Studies*, *131*(3), 627–667. https://doi.org/10.1007/s11098-005-4542-0

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2004). The Phenomenology of Free Will. *Journal of Consciousness Studies*, *11*(7-8), 162–179.

Papineau, D. (2020). Naturalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020). Retrieved from https://plato.stanford.edu/archives/sum2020/entries/naturalism/

Pereboom, D. (2017). A Defense of Free Will Skepticism: Replies to Commentaries by Victor Tadros, Saul Smilansky, Michael McKenna, and Alfred R. Mele on Free Will, Agency, and Meaning in Life. *Criminal Law and Philosophy*, *11*(3), 617–636. https://doi.org/10.1007/s11572-017-9412-2

Pereboom, D., & Caruso, G. D. (2017). *Hard-Incompatibilist Existentialism*. 30.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, *49*(3), 329–359. https://doi.org/10.1007/BF00355521

Rosenthal, D. M. (1997). A Theory of Consciousness. In N. Block, O. J. Flanagan, & G. Guzeldere (Eds.), *The Nature of Consciousness*. MIT Press.

Rosenthal, D. M. (2003). Unity of Consciousness and the Self. *Proceedings of the Aristotelian Society (Hardback)*, *103*(1), 325–352. https://doi.org/10.1111/j.0066-7372.2003.00075.x

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford ; New York: Oxford University Press.

Rosenthal, D. M. (2011). Exaggerated reports: Reply to Block. *Analysis*, *71*(3), 431–437. https://doi.org/10.1093/analys/anr039

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Retrieved from http://books.google.com?id=nAYGcftgT20C

Smilansky, S. (2000). *Free Will and Illusion*. Oxford University Press.

Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies*, *75*(1), 5–24. https://doi.org/10.1007/BF00989879

Strawson, P. (1962). Freedom and Resentment. *Proceedings of the British Academy, Volume 48: 1962*, 1–25.

Vihvelin, K. (2004). Free Will Demystified: A Dispositional Account. *Philosophical Topics*, *32*(1/2), 427–450. https://doi.org/10.5840/philtopics2004321/211

Williams, B. (2006). *The Sense of the Past: Essays in the History of Philosophy* (M. Burnyeat, Ed.). Retrieved from http://books.google.com?id=pudkdYVhjKEC