

**A Subjective Representationalist Approach to Phenomenal Experience**

by

Neil J. Mehta

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Philosophy)  
in The University of Michigan  
2012

Doctoral Committee:

Associate Professor Eric P. Lormand, Co-Chair  
Associate Professor Andrew M. Egan, Rutgers University, Co-Chair  
Professor Victor M. Caston  
Professor Peter A. Railton  
Assistant Professor George A. Mashour

© Neil J. Mehta 2012

## ACKNOWLEDGMENTS

I am extraordinarily grateful to Eric Lormand and Andy Egan for offering insightful and constructive comments, supplying unwavering encouragement, and demonstrating great patience in reading countless drafts of my work. I also thank Todd Ganson for his helpful and extensive remarks concerning many chapters of this dissertation.

I have benefitted from the assistance of many, many more people, including Jake Berger, David Braddon-Mitchell, Victor Caston, David Chalmers, Billy Dunaway, Matt Evans, Dmitri Gallow, Dave Manley, George Mashour, Angela Mendelovici, Peter Railton, Robert Schroer, Brad Thompson, and anonymous referees for *Pacific Philosophical Quarterly* and *Philosophical Studies*.

I presented portions of this dissertation to audiences at the 6th Annual Gateway Graduate Conference at the University of Missouri, St. Louis; the 2011 Southwest Graduate Student Conference in Philosophy; and a 2008 meeting of the Australian National University Philosophy Society. I thank those participants for their feedback.

## PREFACE

There is, as philosophers like to say, *something it's like* for a subject to see a painting by Jackson Pollock, to hear a basketball thud against the backboard, or to taste an Imperial IPA.<sup>1</sup> More generally, there is something it's like for a subject to have an experience. This “what-it's-like,” or *phenomenal*, aspect of experience has been thought to pose large and troubling philosophical questions.

This dissertation will be plugging a view about phenomenal experience that I dub *subjective representationalism*. As a first pass, the idea behind *representationalism* is that experiencing is to be analyzed in terms of a special sort of *representing*.<sup>2</sup> To have an experience is simply to harbor the right kind of representation. Representations of what, though? I claim that at least some of our experiences are representations of *subjective properties* – roughly, properties that are essentially mind-dependent. Hence the moniker “subjective representationalism.”

In this dissertation, I would like to provide a careful, thorough defense of subjective representationalism as a general theory about the nature of phenomenal experience. Unfortunately, I can't do that. The philosophical literature on experience is so sprawling that even an entire dissertation cannot come close to discussing all of the

---

<sup>1</sup> Since Nagel (1974) used “what-it's-like” talk to characterize phenomenal experience, this has become the dominant way of introducing the subject matter.

<sup>2</sup> Representationalism is now a very popular view with a long list of advocates. My defense of representationalism is most deeply indebted to a version of the approach developed by Tye (1995).

relevant issues. So I will do something far more modest. I will show that subjective representationalism is a well-motivated theory, and that it can explain many far-flung data points about experience. Along the way, I'll exhibit notable advantages that subjective representationalism has over some leading rivals.

Let me provide a short user's guide to this dissertation. The reader may be relieved to learn that the chapters of this dissertation are very nearly self-contained and can thus be read in any order. The only exception is that I define the crucial terms *phenomenal experience* and *phenomenal property* in Chapter 1 and rely on those definitions throughout. Readers wishing to skip the bulk Chapter 1 are encouraged to consult those definitions.

Now for an overview of the chapters. In the first two chapters, I will display some data about experience. Chapter 1 focuses on the thoroughly spatial character of experience. Here I argue against views of experience according to which phenomenal properties – roughly, the properties which constitute “what it's like” to have an experience – are internal to the subject's mind. My targets include standard versions of sense-datum theory and Block's “mental paint” theory. If my arguments succeed, then phenomenal properties are outside the subject's mind. Representationalists typically embrace this claim.

Chapter 2 quickly recapitulates some well-known motivations for representationalism and then considers an important choice point for representationalists: are phenomenal properties objective (essentially mind-independent) or subjective (essentially mind-dependent)? Most representationalists – including Alex Byrne, Fred Dretske, Christopher Hill, and Michael Tye – appear to be objective representationalists, holding that all phenomenal properties are objective. I introduce a

robust set of inversion intuitions and argue that subjective representationalism, which holds that some phenomenal properties are subjective, can better accommodate such intuitions than objective representationalism.

These first two chapters comprise my positive, though defeasible, case for subjective representationalism. The next two chapters are defensive, rebutting what strike me as the most deep and threatening objections to subjective representationalism.

Chapter 3 focuses on the familiar “explanatory gap” problem, originally developed by Joseph Levine (1983). David Chalmers and Frank Jackson (2001) have refined Levine’s explanatory gap argument, and Chalmers (but not Jackson) endorses a version of this argument that concludes that phenomenal consciousness is not reducible to the physical. Since the version of subjective representationalism I hold is part of just such an attempt to reduce phenomenal consciousness to the physical, I spend this chapter developing the resources to respond to Chalmers’ argument.

Chapter 4 considers the intuition that internal twins must have precisely the same kinds of phenomenal experiences. This intuition is widespread, powerful, and recalcitrant. But it is very likely that representationalism, especially of the highly externalist sort that I defend in Chapter 1, is incompatible with such intuitions. So, since I can’t join my opponents, I set out to beat them. I argue that this intuition is unreliable and attempt to debunk it.

Finally, Chapter 5 begins *very* tentatively filling in some details of the subjective representationalist account. Representationalists can be sorted into *first-order* and *higher-order* theorists. Roughly, first-order representationalists hold that a ground-level representation of a certain kind suffices for phenomenal experience, while higher-order representationalists insist that phenomenal experiences also require a higher-order

representation of the ground-level representations. I am tempted by the first-order view on grounds of parsimony. In this chapter, I focus on rebutting just one introspectively-based argument for higher-order representationalism offered by Uriah Kriegel (2009a).

Still, I underscore that the case for subjective representationalism will be provisional and incomplete (and this applies tenfold to the case for first-order representationalism!). For one thing, my arguments in favor of subjective representationalism can helpfully be construed as an inference to the best explanation. But better data, or better explanations of existing data, may always come along. More immediately, a complete defense of an inference to the best explanation requires comparing the preferred theory to *all current* rivals. There are simply too many rival theories of experience for me to do this. I do discuss many prominent rivals along the way, including dualism, Block's "mental paint" theory, and various non-subjective forms of representationalism. But other rivals – most notably, naive realism – receive little discussion. This is for lack of space and time, not lack of interest.

Still, I do hope to show, by the end of this dissertation, that the explanatory power of subjective representationalism is remarkable. Subjective representationalism may not be the only game in town, but it's a very appealing theory indeed. Exploring it more carefully will pay substantial dividends in our understanding of the nature of experience.

## TABLE OF CONTENTS

Acknowledgments .....	ii
Preface .....	iii
Chapter	
1. Beyond transparency: the spatial argument for experiential externalism ...	1
2. Exploring subjective representationalism .....	34
3. The real moral of the ‘explanatory gap’ .....	71
4. How I learned to stop worrying and love phenomenal externalism .....	100
5. Is there a phenomenological argument for higher-order representationalism? .....	123
Bibliography .....	145



# Chapter 1.

## Beyond transparency: the spatial argument for experiential externalism

Suppose I'm at the zoo. I observe a hulking gray rhinoceros as it amiably meanders around its pen. I then muse (as philosophers are wont to do) about the nature of my current visual experience. I introspect "what it's like" for me to see the meandering rhino. In doing so, I attend to certain properties – *phenomenal properties*, defined carefully below. These phenomenal properties determine what my experience is like, so by attending to them I come to know what my experience is like.

Where are these phenomenal properties instantiated? As a first pass, we can characterize *experiential internalism* as the view that all instantiated phenomenal properties are instantiated in the subject's mind. An experiential internalist might say that the phenomenal properties associated with my rhino-experience are instantiated in certain neurons in my visual system.

Also as a first pass, we can characterize *experiential externalism* as the contrary view that at least some instantiated phenomenal properties are instantiated outside the subject's mind. An experiential externalist might say that the phenomenal properties associated with my rhino-experience are instantiated on the rhino.

My primary goal in this chapter is to argue for experiential externalism. (As we shall see, this gets us quite far towards the version of subjective representationalism that I prefer.)

Put crudely, the argument goes like this. Phenomenal experience has a pervasively *spatial* character. That is, in introspection, all phenomenal properties associated with experience seem instantiated in certain spatial locations. This alone does not show that phenomenal properties are instantiated in these spatial locations. For things need not be as they introspectively seem.

Nevertheless, it is reasonable to assume that at least one phenomenal property is as it introspectively seems. So *that* phenomenal property must be instantiated in the relevant spatial location. But the only kind of space is ordinary physical space, and the relevant spatial locations are outside the subject's mind. So at least one phenomenal property is instantiated outside the subject's mind. This is what experiential externalism says.

The remainder of this chapter carefully refines this crude argument.

Two further introductory remarks are in order. First, in service of this argument, I will meticulously describe the pervasive spatial character of experience.<sup>1</sup> Aside from its use in my argument, this description is of intrinsic philosophical interest.

Second, my argument will be structurally similar to familiar arguments based on the "transparency of experience."<sup>2</sup> Transparency-based arguments are often invoked to support specific experiential externalist views. However, I will suggest, phenomenological intuitions about spatiality are considerably more stable than related

---

<sup>1</sup> For other recent discussions of the spatial character of experience, see Briscoe (2009), Burge (2010, esp. pp. 492-518), Prosser (forthcoming), Schroer (2007), Schwenkler (forthcoming), and Thompson (2010).

<sup>2</sup> Contemporary discussion of the transparency of experiences traces to Harman (1990). For some prominent transparency-based arguments in favor of specific versions of experiential externalism, see Jackson (2007), Kennedy (2009), Martin (2002), and Tye (2002).

phenomenological intuitions about transparency. Thus, spatiality considerations should be more dialectically effective than transparency considerations for supporting experiential externalism.

Here is a road map for what follows. §1 proposes a relatively theoretically-unladen framework for classifying experiences and characterizes experiential externalism using this framework. §2 introduces some plausible auxiliary assumptions. §3-4 examine the spatial character of a particular visual experience and argues that the phenomenal properties associated with that experience introspectively seem spatially located. §5-7 ambitiously generalize the argument to all phenomenal properties associated with all experiences. §8 marshals these points into an argument for experiential externalism and speculatively entertains an extension of the argument. §9 compares this spatial argument to the better-known transparency argument(s).

## 1. A framework for individuating experiences

We normally individuate experiences phenomenally, i.e., in terms of *what it's like to have them*. How do we do this? To determine what kind of experience I am having, I *introspect* what it is like for me to have that particular experience. When I do this, I attend to certain properties – properties that seem puzzling in many ways.

These introspectible properties generate many well-known philosophical puzzles, such as the inverted spectrum puzzle, the zombie puzzle, the puzzle concerning what Mary didn't know, and the explanatory gap puzzle.<sup>3</sup> Call properties of this sort – properties (i) that I can become directly aware of via introspection, and (ii) which

---

<sup>3</sup> Influential presentations of these puzzles are offered, respectively, by Shoemaker (1982); Chalmers (1996, pp. 93-171); Jackson (1982); and Levine (1983) and (2001).

generate, in the appropriate way,<sup>4</sup> these well-known philosophical puzzles about experience – *phenomenal properties*. I will sometimes speak of particular phenomenal properties like *phenomenal redness*; this should be understood as the phenomenal property associated with my normal experiences of some highly determinate shade of redness.

This characterization of phenomenal properties is relatively theoretically neutral, for it is a plain sociological fact that these puzzles exist (whether or not these puzzles have solutions, and no matter what form these solutions might take).<sup>5</sup> For all I have said so far, phenomenal properties might be properties of things inside or outside our heads; they might be mental or non-mental properties; and they might be physical or non-physical properties.

It is very plausible that phenomenal properties are *all* that phenomenally individuate experiences.<sup>6</sup> Here, then, is the suggestion:

*Individuation Thesis:* Any experience is phenomenally individuated wholly in terms of which phenomenal properties that experience presents.<sup>7</sup>

---

<sup>4</sup> What is “the appropriate way” of generating these puzzles? The idea is that phenomenal properties are those which zombies putatively lack (or lack awareness of), that Mary putatively doesn’t know about, that distinguish the experiences of spectrum-inverted twins, and that generate the appearance of an explanatory gap.

<sup>5</sup> Thus the definition of phenomenal properties does not presuppose that zombies, inverted spectrum scenarios, etc. are metaphysically possible or even coherently conceivable.

<sup>6</sup> Of course, there are myriad *non-phenomenal* ways one might individuate experiences. One could individuate experiences in terms of what time they begin, how tall the subject of the experience is, etc. The textual claim concerns only what individuates experiences *purely phenomenally*.

<sup>7</sup> Some philosophers (including Tye (2009, ch. 4) and many naïve realists) individuate experiences not just in terms of properties, but also in terms of specific objects. Such philosophers can still accept the Individuation Thesis as long as they grant that phenomenal properties are all that individuate experiences *phenomenally*, per the previous footnote. And this seems right. For, even in optimal conditions, one may have phenomenally identical experiences of distinct objects (e.g., when one successively sees twins). “What it’s like” to see either twin is precisely the same.

The term *presentation* here is a neutral label for the relation that we bear to phenomenal properties in experience. Though different theorists will propose very different construals of this relation, every theorist who isn't an eliminativist about experience should recognize such a relation. Given how neutrally I have characterized both phenomenal properties and the presentation relation, virtually any theorist can endorse the Individuation Thesis.

For example, one might think that experiences are phenomenally individuated solely via their intrinsic physical properties<sup>8</sup>; this is the view that phenomenal properties are intrinsic physical properties, and presentation is simply instantiation. Or one might think, as some *representationalists* do, that experiences are phenomenally individuated solely via the properties they represent.<sup>9</sup> On this view, phenomenal properties are represented properties, and presentation is representation.

*Naïve realists*, meanwhile, think that some experiences are literally partially constituted by bits of external reality; when I see a red flower, the redness of the flower<sup>10</sup> is a constituent of my visual experience.<sup>11</sup> On this view, presentation is something like *constitution* or *containment*. One might even think, as some *dualists* do, that any experience is phenomenally individuated in terms of the non-physical properties which that experience acquaints one with.<sup>12</sup> If this is right, then phenomenal properties are these non-physical properties, and presentation is acquaintance. As such examples show, the present terminological framework is not theoretically loaded.

---

<sup>8</sup> E.g., Hill (1991).

<sup>9</sup> Representationalism is defended by Byrne (2001); Carruthers (2000) and (2005); Dretske (1995) and (2003); Harman (1990); Hill (2009); Lycan (1996); Pautz (2010); Prosser (forthcoming); Tye (1995), (2000b), and (2002); and many others.

<sup>10</sup> And perhaps the flower itself. See footnote 7.

<sup>11</sup> Proponents of naïve realism include Brewer (2008), Campbell (2002), Fish (2009), Kennedy (2009), Martin (2006), and McDowell (2008).

<sup>12</sup> Recent defenses of dualism may be found in Chalmers (2009), Nida-Rümelin (2007), and White (2007).

On some uses of the term “phenomenal property,” it is definitional that experiences instantiate phenomenal properties. This is not definitional as I use the term. The issue is purely terminological, for, on my usage, any experience will instantiate the property of *presenting such-and-such phenomenal properties*.

Now I will use this terminological framework to characterize the view I shall argue for. Theories of experience may be divided according to where they locate phenomenal properties relative to the subject’s mind. (Throughout, the *mind* may be understood to include the brain, nervous system, non-physical soul, or whatever.) There are three possibilities:

(1) *Experiential extremism*: No phenomenal properties are instantiated.

(2) *Experiential internalism*: Some phenomenal properties are instantiated, and all instantiated phenomenal properties are instantiated in the subject’s mind.

(3) *Experiential externalism*: Some phenomenal properties are instantiated, and at least one phenomenal property is instantiated outside the subject’s mind.

I will defend experiential externalism. Experiential externalism may be subdivided into these two views:

(3a) *Moderate experiential externalism*: Experiential externalism is true, but at least one phenomenal property is instantiated in a subject’s mind.

(3b) *Radical experiential externalism*: Experiential externalism is true, and all instantiated phenomenal properties are instantiated outside the subject's mind.

I am cautiously sympathetic to radical experiential externalism, and I will tentatively suggest an argument to support it.

Experiential externalism should not be confused with the more familiar thesis of *phenomenal externalism*. According to phenomenal externalism, two subjects who are exact intrinsic duplicates may differ in their phenomenal experiences. But even if experiential externalism is true, phenomenal externalism may still be false. For suppose that at least one instantiated phenomenal property is instantiated external to the subject, as experiential externalism says. It may still be that any two subjects who are exact intrinsic duplicates necessarily must be presented with the same (possibly external) phenomenal properties, contrary to phenomenal externalism (whether or not those phenomenal properties are actually instantiated). How plausible this is will depend on one's view of the presentation relation. Since experiential externalism does not entail phenomenal externalism, I will not discuss phenomenal externalism here, though I will defend it in Chapter 4.

Why does it matter if experiential externalism is true? Because many popular views about experience are incompatible with experiential externalism. If my argument succeeds, such views are ruled out. For example:

On some views, no phenomenal properties are instantiated. Such views are straightforwardly experientially extremist. On other views, experiences are individuated wholly in terms of their intrinsic physical properties; an extreme version of

a “mental paint” theory might hold this.<sup>13</sup> This is plainly a form of experiential internalism, as it holds that phenomenal properties – the ones that individuate experiences – are all properties of one’s mind.

Next, consider *representationalism*. As I am using this term, representationalists hold that experiences are individuated wholly in terms of their representational content. On this view, experience *presents* phenomenal properties by *representing* them; if these contents concern only properties, then phenomenal properties are simply the represented properties. Most representationalists are experiential externalists.<sup>14</sup> But a representationalist may hold that experiences represent only properties of one’s mind (and that at least some of these properties are instantiated); this counts as experiential internalism.<sup>15</sup> Or she may hold that experiences represent only uninstantiated properties.<sup>16</sup> This counts as experiential extremism.

It is worth taking special note of Sydney Shoemaker’s representationalist view that (to use my terminology) at least some phenomenal properties are mind-dependent properties instantiated in environmental objects – properties like *being disposed to cause experiences of such-and-such type in certain subjects*.<sup>17</sup> This is a form of *subjective representationalism*, a term I will define more carefully in subsequent chapters.

Shoemaker’s subjective representationalism is an experientially externalist view, not an experientially internalist view. For what matters in classifying a view as experientially internalist or externalist is where phenomenal properties are supposed to be instantiated – in the subject’s mind, or outside the subject’s mind? It does not matter whether phenomenal properties are mind-dependent or mind-independent properties.

---

<sup>13</sup> For discussion, see Block (1996).

<sup>14</sup> Paradigmatic examples include Dretske (1995) and Tye (1996) and (2000).

<sup>15</sup> I *very* tentatively interpret Kriegel (2009) as holding such a view.

<sup>16</sup> Chalmers (2006) is naturally interpreted as holding such a view.

<sup>17</sup> See Shoemaker (2003) and (2006); a related view is discussed in Egan (2006).



And Shoemaker holds that at least some phenomenal properties are (mind-dependent) properties instantiated *in environmental objects*, outside the subject's mind.

Finally, according to *sense-datum theory*, experience acquaints us with certain properties of mind-dependent objects (sense-data). On this view, presentation is *acquaintance*, and phenomenal properties are these properties of sense data. Paradigmatic forms of sense-datum theory are experientially internalist views, since they hold that all sense-data are in one's mind.<sup>18</sup>

Thus, it matters a great deal whether experiential externalism is true or not. Specifically, if my argument for experiential externalism is successful, then many leading theories of experience are ruled out, including certain mental paint theories; some forms of representationalism; paradigmatic versions of sense-datum theory; and views on which no phenomenal properties are instantiated.

## 2. Assumptions

My argument requires two auxiliary premises. Both are widely, but not universally, accepted, and both seem eminently defensible to me. However, providing such a defense would be time-consuming and would require discussion of many extraneous issues. So I will simply assume the truth of these premises.

The first premise concerns introspection:

*Weak Charity Principle:* Some phenomenal property, presented by the experience of some actual normal human subject, is as it introspectively seems.

---

<sup>18</sup> But in principle a sense-datum theory could be experientially externalist, since sense-data might be located outside one's mind. For something can be *mind-dependent* without being located *in one's mind*. Jackson (1977) considers a view like this.

The Weak Charity Principle is aptly named; it is a *very* weak principle. It does not say that most phenomenal properties, or even very many phenomenal properties, are as they introspectively seem. It claims only that *at least one* phenomenal property is as it introspectively seems, and thus has considerable initial plausibility. For to deny the Weak Charity Principle is to think that introspection is massively misleading: *no* phenomenal property is as it introspectively seems. This is a radical and arguably unwarranted error theory.

Not all philosophers will accept the Weak Charity Principle. Some, like Schwitzgebel (2008), think that introspection really is massively misleading in major respects. Others, including Boghossian and Velleman (1989) and arguably Chalmers (2006), hold that experience routinely involves massive misrepresentation. A proponent of such a view might think that introspection inherits the unreliability of experience.

I lack space to discuss such views here; I will just take the Weak Charity Principle for granted. Those worried about this principle may construe my argument as having this conditional conclusion: if the Weak Charity Principle is true, then experiential externalism is true. Some experiential extremists will be unimpressed by this conclusion, since some experiential extremists are already committed to rejecting the Weak Charity Principle. However, some experiential extremists will likely want to accept the Weak Charity Principle.<sup>19</sup> Further, many experiential internalists accept the Weak Charity Principle, or at least have given no indication of wanting to reject it. If

---

<sup>19</sup> For example, Chalmers (2006) arguably endorses experiential extremism, but Chalmers (2003) argues that, under certain circumstances, introspection is infallible. Assuming those circumstances have actually obtained at least once, this entails the Weak Charity Principle.

even this conditional conclusion is true, such philosophers would have to give up the Weak Charity Principle. This would be a substantial and surprising philosophical result.

The second premise concerns the nature of space:

*Physical Space Thesis:* There is only one kind of space, namely, ordinary physical space.

The Physical Space Thesis entails that there is no additional mental space, beyond ordinary physical space. It also entails that there is no dualistic, non-physical space.

Physicalists should unhesitatingly endorse the Physical Space Thesis, but one might expect dualists to balk at it. As a sociological matter, however, most contemporary dualists do not posit any non-physical space. Instead, they typically take phenomenal properties to be non-physical properties instantiated in physical objects (e.g., our brains) in physical space, or to be uninstantiated properties.<sup>20</sup> My argument will thus apply even to most actual dualists. And I emphasize again that experiential externalism is compatible with dualism. In any case, those who reject the Physical Space Thesis may again construe my argument as having a conditional conclusion: if the Physical Space Thesis is true, then experiential externalism is true. This is still a powerful philosophical result.

Some radical pessimists about introspection, and some radical dualists, may reject the assumptions introduced in this section. Other philosophers should find these assumptions innocuous. My argument applies with full force against such philosophers.

---

<sup>20</sup> So far as I know, David Chalmers, perhaps the most prominent contemporary dualist, does not posit any non-physical space. And in his reluctant defense of dualism, Lycan (2009) argues that the best form of dualism does not posit non-physical space.

### 3. Spatial appearances introduced

Let's begin the argument for experiential externalism. Focus on this case: I am alert, calm, and otherwise cognitively well-functioning, and I am looking at a mango in good light, carefully attending to its color. I then introspect what it's like for me to experience those colors; I attend to certain phenomenal colors.

I make the following highly intuitive phenomenological claim about this case: when I introspect my mango-experience, the phenomenal colors presented by that experience seem to be instantiated in certain spatial locations, both relative to me and relative to each other. I'll call the relevant type of space *phenomenal space*.

To bring out the appeal of this phenomenological claim, let's flesh out the example a bit more. When I introspect what it's like for me to have the mango-experience, certain phenomenal color properties seem instantiated *in front of me* in phenomenal space. Moreover, these phenomenal color properties seem related to each other in phenomenal space: this speck of phenomenal redness seems instantiated *to the left* of that speck of phenomenal orangeness in phenomenal space. There is nothing special about my mango experience: the apparent spatiality of phenomenal color properties is a pervasive feature of visual experiences.

This phenomenological claim by itself rules out almost no theory of experience. For it does not comment on whether phenomenal space is private or public, mental or non-mental, physical or non-physical, etc. After all, experience does not trade in fancy concepts like the concepts of the public/private, the mental/non-mental, or the physical/non-physical.

I also emphasize that I am focusing here on a case involving introspection, not unreflective visual experience. Of course, in normal cases of unreflective visual

experience, various properties (like *being red* and *being orange*) seem instantiated in various spatial locations relative to me and to each other. One could argue that these properties are not phenomenal properties. But my phenomenological claim concerns the case where I introspect what it is like for me to see the ripe red mango, and introspection (I pointed out in §1) makes me aware of phenomenal properties. It is specifically these *phenomenal* properties, discernable introspectively, that seem instantiated in spatial locations relative to me and to each other.

Thus, certain phenomenal properties introspectively seem instantiated in certain spatial locations relative to me and to each other. (I'll often shorten this by saying that phenomenal properties introspectively seem *spatially located*, though strictly speaking they seem *instantiated in spatial locations*.)

I have assumed the Physical Space Thesis: there is only one kind of space, namely, ordinary physical space. So we may conclude that, if things are as they introspectively seem, then certain phenomenal properties must be genuinely instantiated in certain locations *in ordinary physical space*.

Let us say that something is a *phenomenal object* at some time  $t$  just in case it instantiates a phenomenal property at  $t$ . (I mean to be very generous in what counts as an "object." On my usage, expanses of sky, nonphysical souls, and neural processes would all count as objects.) This conclusion narrows down what phenomenal objects must be for my introspective seemings to be correct. If phenomenal properties are instantiated in certain locations in ordinary physical space, then phenomenal objects must also be located in ordinary physical space. Moreover, introspection provides specific information about the spatial locations of phenomenal properties. When I introspect my mango experience, the phenomenal redness presented by that experience

seems to cover a surface shaped like a mango. So, if introspection is correct, there must be a phenomenal object shaped like a mango.

Could this mango-shaped phenomenal object be in my brain? This thought would be reasonable if, whenever I visually experienced a mango, a cluster of neurons shaped like a mango collectively fired. But this is not the case.

Or perhaps phenomenal objects are on the retina? This suggestion is also unsuccessful. The mango I am looking at is rounded. As I introspect, I notice that the phenomenal properties presented by my experience seem to occupy a space shaped like the seen surface of a rounded mango. These phenomenal properties seem to be instantiated in locations along a three-dimensional surface. But, while my retina is three-dimensional, nothing on my retina is shaped three-dimensionally like the seen surface of a mango.

When I see a mango, the only thing that stands in the right spatial relations is the mango itself. But the mango is an external object in physical space. So we have arrived at this conclusion:

*Singular Spatiality Thesis:* if things are as they introspectively seem, then some of the phenomenal properties presented by my mango-experience are instantiated in an external object (i.e., an object outside my mind) in physical space.

The Singular Spatiality Thesis states *where* some phenomenal properties are located, if things are as they introspectively seem. I reiterate that it entails little about the nature of phenomenal properties themselves. It is silent about whether phenomenal

properties are physical or non-physical; mind-dependent or mind-independent; and intrinsic or extrinsic properties of external objects.

Thus, phenomenological considerations, plus the Physical Space Thesis, strongly support the Singular Spatiality Thesis.

#### 4. Can we get by with less?

This section addresses some potential reservations about the argument for the Singular Spatiality Thesis.

First, one might worry that the argument supports only this conclusion, which is weaker than the Singular Spatiality Thesis: if things are as they introspectively seem, phenomenal properties must stand in *relations somehow isomorphic to spatial relations*, which need not be *spatial relations, simpliciter*. This weaker conclusion is compatible with the claim that phenomenal properties are, e.g., neural properties. For then one can say that it does not introspectively seem that this speck of phenomenal redness is to the left of that phenomenally orange speck. Rather, it introspectively seems that the phenomenally red speck stands in some relation to the phenomenally orange speck that is isomorphic to this spatial relation. There are all kinds of relations that may be isomorphic to spatial relations – perhaps, for instance, neuronal firing speeds are isomorphic to these spatial relations.

This worry misconstrues the phenomenological point. For when I introspect my mango-experience, that phenomenally red speck seems to be quite literally *to the left, simpliciter*, of that phenomenally orange speck. If I were to find out that my introspective awareness of the “to-the-left” relation was merely awareness of neuronal

firing rates, I would think that introspection was misleading about how phenomenal redness and phenomenal orangeness are related.

For a useful contrast, suppose I were to discover instead that my introspective awareness of the “to-the-left” relation was awareness of literal spatial relations in a non-physical space. Though I would be greatly surprised, I would think that introspection was correct in its presentation of the “to-the-left” relation.

This same point handles a related worry. One might claim that, when we introspect, we are aware of *two* types of properties: *phenomenal properties*, which individuate experiences, and *ordinary properties* (like colors, sounds, etc.), which don’t individuate experiences.<sup>21</sup> The proponent of this view might say that while ordinary properties introspectively seem instantiated in some kind of space, phenomenal properties do not introspectively seem instantiated in any kind of space.

To soothe this worry, I need not take a stand on whether introspection makes us aware of two types of properties. All I need is the claim that the phenomenal properties themselves introspectively seem spatially located. And this claim is phenomenologically apt. *Phenomenal* redness itself – the property that I am introspectively aware of, the one that engenders puzzles about zombies, inverted spectra, etc. – seems instantiated to the left of the phenomenally orange speck.

Thus, phenomenological investigation supports the Singular Spatiality Thesis. Let’s see if the considerations supporting the Singular Spatiality Thesis can be pushed further.

## 5. Spatiality and perceptual experience

---

<sup>21</sup> Though see Sundström (2007) for vigorous opposition to such views.



The Singular Spatiality Thesis is quite limited: it speaks only about where *some* phenomenal properties presented by *a single experience* are instantiated if things are as they introspectively seem. I now want to generalize the Singular Spatiality Thesis to all perceptual experiences.

First, consider all phenomenal properties of my mango-experience. *All* such properties, I claim, introspectively seem spatially located. To anticipate a bit: I will draw several of my examples from the literature on the “transparency” of experience. The full significance of this will become clear in §9.

Start with phenomenal colors. One delicate issue is exactly which properties associated with color experience are phenomenal. When I look at wall that is uniformly white, there is a sense in which the whole wall looks the same color. But there is another sense in which certain areas of the wall (e.g., shaded or highlighted areas) look different with respect to color. If there is just one set of phenomenal color properties, do these properties correspond to the sense in which the wall looks uniformly white or the sense in which the wall looks to have different colors? Or are two sets of phenomenal properties in play here?<sup>22</sup>

Fortunately, I need not take a stand on this issue. For both types of properties introspectively seem spatially located. When I introspect my wall-experience, I can identify quite precisely which locations seem to be uniformly white, and which locations seem to be shaded or highlighted. So whichever of these color-related properties are phenomenal properties, they introspectively seem spatially located. (Even reflection on inverted spectrum<sup>23</sup> or inverted earth<sup>24</sup> cases does not undermine this point.)

---

<sup>22</sup> See Chalmers (2006, pp. 84–89) for some discussion.

<sup>23</sup> Shoemaker (1982).

<sup>24</sup> Block (1990).

Let's consider other phenomenal properties presented by my mango experience. When I am looking straight at the mango, I experience at least part of the surface of the mango very sharply; outside of this clear portion, my experience has a more hazy quality. Therefore, one might think that my experience presents *phenomenal haziness*.

If there is such a thing as phenomenal haziness, then it also introspectively seems spatially located. The phenomenally hazy bit *surrounds* the phenomenally sharp bit in phenomenal space. Now, maybe introspection does not determinately say where the phenomenal haziness ends; maybe introspection indicates only approximately where the phenomenal haziness ends. But even if such indeterminacy exists, properties like phenomenal haziness still introspectively seem spatially located.

Next, consider phenomenal shape properties. When I look at the mango, there is a sense in which it looks three-dimensional, roughly like an ellipsoid. But there is another sense in which I am apparently aware of something two-dimensional, roughly like an ellipse.<sup>25</sup> (It is easier to get this sense when one thinks about how to draw a mango.) Still, both of these phenomenal shape properties introspectively seem spatially located. This should be clear, since *shapes* of any kind are individuated by their (geometric) spatial properties. Moreover, these phenomenal shape properties introspectively seem to be instantiated in certain locations relative to other phenomenal properties. The phenomenal ellipsoid I see when I look at the mango introspectively seems to be below *this* phenomenal cylinder and to the right of *that* phenomenal box-shape.

---

<sup>25</sup> Peacocke (1983) famously raised this issue with the “two trees” case. For further discussion, see Hill (2009, ch. 5), Millar (2010), Siewert (2004), and Tye (1996b).

Well, here is a property presented by my visual experience that one might think is phenomenal but not apparently spatial: the property of *being a visual experience*.<sup>26</sup> My experience instantiates this property, and arguably I can introspectively attend to this property, too. I can, for example, introspectively tell that I am having a visual experience, not a tactile experience, of the mango's shape.

But the fact that I can introspectively tell that my experience is visual rather than, say, tactile does not show that the property *being a visual experience* is a phenomenal property. For though I've assumed that phenomenal properties are properties of the type that (among other things) we are introspectively aware of, this is quite different from the assumption that all introspectible properties are phenomenal properties.

Notice that I can attend quite easily to what it is like for me to experience phenomenal redness or phenomenal squareness. By contrast, it is much harder to follow this instruction: "attend to what it is like for you to have a visual experience." Insofar as I can follow this instruction, I do so by attending to phenomenal colors, phenomenal shapes, etc. This suggests that my knowledge that my mango-experience is visual comes, not purely from introspecting what my experience is like, but also from certain background knowledge – for instance, the background knowledge that experiences that present phenomenal colors are visual experiences. So my introspective knowledge that my mango-experience is visual need not threaten the thesis that all phenomenal properties presented by that experience introspectively seem spatially located.

I know of no other phenomenal properties presented in my mango-experience that pose a threat to this thesis. Moreover, in the argument I just outlined, while I

---

<sup>26</sup> Lycan (2002) raises a similar point.

pointed out many distinctive features of visual experience, I did not rely on distinctive features of my mango-experience (rather than some other visual experience). The same points apply to any visual experiences of normal subjects.<sup>27</sup>

These points even extend to non-standard visual experiences, such as blurry experiences, double-vision experiences, and after-image experiences. It's not clear exactly what the relevant phenomenal properties are for cases of blurry vision. When my eyes water and I have a blurry experience of a frog, am I merely presented with points of phenomenal greenness, shaped something like a cloud? Or is there some proprietary phenomenology of blurriness, so that I am presented with *phenomenal blurry-colors*, whatever those are?

Regardless of what one wants to say, the relevant properties introspectively seem spatial. This cloud-shaped set of phenomenally green points, or these phenomenal blurry-green specks, or whatever, introspectively seem to stand in *some* kinds of spatial relations to me and to other phenomenal colors (or phenomenal blurry-colors). This holds even if one thinks that these spatial relations introspectively seem imprecise.

Similarly for double-vision cases: if I see double when I look at my left hand, *this* phenomenally presented ring-finger-shape introspectively seems to be to the left of *that* one. The phenomenal properties associated with after-images, too, introspectively seem to occupy spatial positions: the phenomenal greenness of this after-image seems to be in front of certain other phenomenal colors. Thus, we can generalize the Singular Spatiality Thesis to all phenomenal properties presented by any visual experience of any normal subject.

---

<sup>27</sup> Schroer (2007, p. 408) endorses a similar conclusion, though his argument differs substantially from mine. It is also worth mentioning that this is the strongest phenomenological claim about spatiality that Schroer (2007) endorses. I will defend a much stronger claim below.

Generalizing further: we need not restrict our conclusions to *visual* experiences. Plausibly, the same is true for any perceptual experience at all. Since the considerations here are parallel to those in the visual case, discussion will proceed briskly. (I invite readers persuaded that this generalization succeeds to skip to the Perceptual Spatiality Thesis at the end of this section.) Consider auditory experiences. Phenomenal loudness, phenomenal timbre, and so on are introspectively seem to come from various directions in space. One might worry that it's possible to experience a sound (say, in an echoing room) without having any introspective sense of where the sound is coming from. But it's plausible that, even in such a case, phenomenal sounds introspectively seem to come from all around, or from locations that are changing rapidly, or from locations that aren't precisely specified.

And similarly for other perceptual experiences. Phenomenal hardness, phenomenal slipperiness, phenomenal roughness, etc., introspectively seem to be variously located in space. This is especially obvious in a case where I touch two different objects, one with each hand. The phenomenal properties presented by my tactile experiences associated with each hand seem instantiated in particular spatial locations relative to one another. But the point also applies in cases where I touch only one object.

For similar reasons, phenomenal tastes introspectively seem spatially located. To help latch onto this idea, consider a case where I swish wine inside my mouth. The phenomenal tastes presented by my wine-experience seem instantiated in various rapidly changing spatial locations. The point generalizes to all gustatory experiences. Even phenomenal odors introspectively seem located, though perhaps diffusely.

I just noted that phenomenal properties associated with tactile, gustatory, and olfactory experiences introspectively seem spatially located in certain ways. Earlier, I argued that, if things are as they introspectively seem, certain phenomenal properties presented by my mango-experience must be instantiated on the (physical) mango. A parallel argument shows that, if things are as they introspectively seem, then the tactile, gustatory, and olfactory phenomenal properties that I experience must be located in certain ways with respect to my (physical) body. Since this argument is parallel to the earlier one, I present it swiftly in the next paragraph.

Introspection on perceptual experience provides a wealth of detail about how tactile, gustatory, and olfactory phenomenal properties are related in some kind of space. But the only things that stand in anything like the correct spatial relations are objects appropriately related to my physical body – perhaps my skin or objects touching my skin, my tongue or objects touching my tongue, and my nose or bodies of air related in certain ways to my nose.

To offer a representative example: when I dip my hands in warm water, I experience phenomenal warmth. This phenomenal warmth introspectively seems to occupy a spatial region shaped roughly like the outer surface of my hands. What object might instantiate such phenomenal warmth? The only good candidates are my hands themselves, or the layer of water surrounding my hands. For there will not be any hand-shaped cluster of neurons firing in my brain, nor are there any other obvious hand-shaped candidates. Similar examples can be constructed for other perceptual modalities.

In light of these observations, I propose this generalized version of the Singular Spatiality Thesis:

*Perceptual Spatiality Thesis:* For any introspected perceptual experience E of any actual normal human subject S, if things are as they introspectively seem, then each phenomenal property presented by E is instantiated in an external object (i.e., an object outside S's mind/brain) in physical space.

## 6. Spatiality and non-perceptual experience

So far, we've focused on perceptual experiences. But, as I'll argue in this section, the phenomenological point applies to *all* experiences. As before, readers already persuaded of this point may go directly to the Spatiality Thesis at the end of this section without loss.

Consider the phenomenal properties associated with pains, pleasures, and bodily sensations. These phenomenal properties all introspectively seem (precisely or vaguely) located in regions shaped roughly like parts of one's body. There is no such thing as a pain, pleasure, or tickle that seems to be nowhere. The relevant phenomenal properties may seem diffuse (as when my whole arm aches) or moving rapidly (as when my arm tingles), but nevertheless these phenomenal properties introspectively seem spatially located. For reasons that should now be familiar, the only good candidates for objects that might instantiate such phenomenal properties (assuming that things are as they introspectively seem) are parts of my physical body. For only parts of my physical body have the right three-dimensional shape.

The argument even applies to experiences associated with emotions or moods. The precise relationship between experiences and emotions or moods is highly contentious. Still, all parties will grant that certain emotions are typically accompanied

by distinctive bodily sensations. Think of what it's like to feel the burn of anger, for instance: one experiences the clenching of one's jaw, the tensing of various muscles, a surge of heat in one's chest, etc. Since these are *bodily* sensations, the relevant phenomenal properties introspectively seem instantiated in specific spatial locations. The only plausible candidates for such locations are within, or on the surface of, my physical body.

Are there additional phenomenal properties, beyond those associated with bodily sensations, that may accompany emotions or moods? Perhaps. One might think that emotional experiences present some phenomenal properties that introspectively seem instantiated outside one's body.<sup>28</sup> Perhaps when I see a bear in the woods, that phenomenally brown, phenomenally fuzzy thing introspectively seems *phenomenally dangerous*. Or perhaps, in love, my beloved introspectively seems *phenomenally radiant*.

Even if this is right, such phenomenal properties introspectively seem spatially located. I argued earlier that visually experienced properties like phenomenal brownness and phenomenal fuzziness introspectively seem spatially located. And when I see a bear, phenomenal dangerousness introspectively seems instantiated by the same object that is introspectively seems phenomenally brown and phenomenally fuzzy. So phenomenal dangerousness, if such there be, seems spatially located. Similarly for phenomenal radiance.

To change topics: recently, many have suggested that there is a distinctive phenomenology of thought.<sup>29</sup> For example, Horgan and Tienson (2002, p. 523) claim to introspect a phenomenal difference between "hearing ... 'Time flies' as a cliché about the passage of time, vs. ... hearing it as a command at the insect races." They suggest that

---

<sup>28</sup> As Masrour (forthcoming, p. 5) suggests.

<sup>29</sup> Including Bourget (2010), Horgan and Tienson (2002), Kriegel (2011), and Loar (2003).



this phenomenal difference goes hand in hand with the intentional difference between the two readings of “Time flies.” If this is so, do these special *intentional phenomenal properties* seem spatially located?

Yes. For this phenomenology of intentionality, if it exists at all, is associated with episodes of (e.g.) hearing words, sentences, etc. This involves ordinary auditory experiences. But I’ve already argued that auditory phenomenal properties seem instantiated in spatially located phenomenal objects. (Recall again that the term “object” is being used very loosely, to apply to physical things, non-physical things, processes, etc.) Intentional phenomenal properties introspectively seem to be properties of these auditorily experienced phenomenal objects, which are presented spatially. So these intentional phenomenal properties themselves seem spatially located – specifically, located wherever in space those auditory phenomenal objects appear to be.

This discussion supports a general phenomenological thesis about our actual experiences:

*Spatiality Thesis:* For any introspected experience E of any actual normal human subject S, if things are as they introspectively seem, then all phenomenal properties presented by E are instantiated in external objects (i.e., objects outside S’s mind) in physical space.<sup>30</sup>

The Spatiality Thesis does not claim that the apparent spatiality of phenomenal properties in phenomenal space is a necessary feature of all introspected experiences, so thought-experiments about aliens with space-free phenomenology cannot undercut it.

---

<sup>30</sup> Aside from cases where I might, via some grisly process, see the color of my brain. Such exceptions will not threaten the main argument.

This phenomenological claim is restricted only to experiences of actual normal human subjects.

## 7. A complication: phenomenal spatial properties

Consider this interesting wrinkle: at least some spatial or space-like properties associated with experience are plausibly phenomenal properties. After all, there is something it's like to see an orange speck on my left; what that's like differs from what it's like to see the same speck on my right. A natural explanation is that these experiences present different phenomenal space properties, just as experiences of red and orange present different phenomenal color properties. Does this affect the argument? It seems not. Phenomenal space properties, as this label suggests, are obvious candidates for phenomenal properties that introspectively seem instantiated in spatial locations.

Matters are complicated, however. To see why, consider two subjects, Rightside Up and Upside Down.<sup>31</sup> Rightside Up has the same sorts of visual experiences as normal human beings. Upside Down has the sorts of visual experiences that Rightside Up would have were he to wear up-down inverting goggles: goggles that, intuitively, “flip everything upside down.” Upside Down's visual experiences are perfectly normal for her; she lives in a community of beings who have evolved to have such experiences. Moreover, Rightside Up and Upside Down are equally adept at maneuvering in space in all relevant ways; they are functionally identical.

Such a case seems perfectly conceivable. Moreover, if these characters take their experiences at face value, then they will acquire true beliefs about the spatial locations of

---

<sup>31</sup> Thompson (2010) and Prosser (forthcoming) construct similar cases.

things.<sup>32</sup> This is so despite the fact that if Rightside Up were abruptly to start experiencing things like Upside Down, and if he were to take his experiences at face value, then he would form false beliefs about the spatial locations of things.<sup>33</sup>

If this is possible, then two individuals can have phenomenally identical and wholly appropriate experiences, even though things introspectively seem different to them with respect to space. (Imagine Rightside Up views a scene, and Upside Down views a similar scene while hanging upside down from the ceiling.) Call these *spatial inversion cases*. I take no stand on whether spatial inversion cases are really possible. If they are not, then there is not even an apparent threat to the Spatiality Thesis. Even if they are, though, the Spatiality Thesis remains secure.

The Spatiality Thesis says that, for any introspected experience E of any actual normal human subject S, if things are as they introspectively seem, then all phenomenal properties presented by E are instantiated in external objects (i.e., objects outside S's mind) in physical space. Subjects like Upside Down are not a counterexample to this thesis. When Upside Down and Rightside Up introspect their visual experiences, both will agree that all the phenomenal properties presented by those experiences introspectively seem spatially located. This is so despite the fact that, if Rightside Up were to have the same experience as Upside Down, things would introspectively seem very different to him with respect to space than they would to Upside Down. The phenomenal properties associated with a particular experience will all seem spatial, though specifically how they seem to be spatially located may differ from subject to subject.

---

<sup>32</sup> So it would be natural for those who think that experiences have contents to say that the experiences of Rightside Up and Upside Down have veridical contents with respect to space.

<sup>33</sup> So it would be natural for those who think that experiences have contents to say that the experiences of Rightside Up and Upside Down differ in content.

An apparently more worrisome threat to the Spatiality Thesis comes from this potential possibility. Perhaps there could be a subject whose experiences are just like Rightside Up, but where the properties I've been calling "phenomenal space properties" are normally correlated with some non-spatial property – perhaps color. (This might be thought of as a radical form of synesthesia.)

Even if this is possible – and intuitions here are very murky – the Spatiality Thesis remains unthreatened. For it concerns only *actual normal human* subjects. It is very plausible that, for any actual experience E that *we* have, all phenomenal properties presented by E introspectively seem spatially located. This is so even if, for other *actual abnormal* or *merely possible* individuals undergoing the same experience E, those phenomenal properties wouldn't introspectively seem spatially located.

This completes my phenomenological project of characterizing the spatial character of experience. In the next section, I will show how the spatial character of experience supports experiential externalism.

## 8. The path to experiential externalism

From here, the path to experiential externalism is smooth. I have assumed this principle:

*Weak Charity Principle:* Some phenomenal property, presented by the experience of some actual normal human subject, is as it introspectively seems.

I have also spent much time arguing for this thesis via phenomenological considerations and the Physical Space Thesis:

*Spatiality Thesis:* For any introspected experience E of any actual normal human subject S, if things are as they introspectively seem, then all phenomenal properties presented by E are instantiated in external objects (i.e., objects outside S's mind) in physical space.

These premises immediately entail that at least one phenomenal property is instantiated outside the subject's mind. But this is simply experiential externalism. So experiential externalism is secured by the Spatiality Thesis plus the Weak Charity Principle.

This argument for experiential externalism is compatible with the thought that some phenomenal properties are instantiated in the subject's mind, or are simply uninstantiated. I argued only for the claim that *at least one* phenomenal property is instantiated outside the subject's mind. For all I've said, plenty more phenomenal properties may be instantiated in the subject's mind or not instantiated at all. Thus, my view is compatible with views according to which phenomenal properties presented by some experiences are not instantiated anywhere.<sup>34</sup>

I would like to tentatively suggest a possible extension of this argument. Many will find this principle appealing:

---

<sup>34</sup> Representationalists standardly think that illusory and hallucinatory experiences often present uninstantiated phenomenal properties. For different reasons, Johnston (2004) also thinks that hallucinatory experiences present uninstantiated phenomenal properties (which he calls "sensible profiles").

*Uniformity Principle:* Either all instantiated phenomenal properties are instantiated in the subject's mind, or all instantiated phenomenal properties are instantiated outside the subject's mind.

A possible rationale for the Uniformity Principle is that giving up this principle makes it very difficult to retain a unified theory of phenomenal properties.

To mention some influential views about experience: phenomenal properties have been thought to be environmental/bodily properties; or neural properties; or properties of private mental sense-data; or non-physical properties of one's mind. On the first view, all instantiated phenomenal properties will be instantiated outside the subject's mind. On the remaining views, all instantiated phenomenal properties are instantiated in the subject's mind. It is difficult, to say the least, to name a reasonably unified class of properties that (i) may be instantiated inside or outside the subject's mind, and (ii) are plausible candidates to be phenomenal properties.

One may, of course, flout the Uniformity Principle by claiming that phenomenal properties are fundamentally disunified: perhaps some are neural properties, and some are environmental properties. Prima facie, such a schizophrenic theory of phenomenal properties is theoretically costly.

I have already defended experiential externalism, which says that at least one phenomenal property is instantiated outside the subject's mind. But if one accepts the Uniformity Principle, then either all instantiated phenomenal properties are instantiated in the subject's mind, or all instantiated phenomenal properties are instantiated outside the subject's mind. So experiential externalism plus the Uniformity Principle together entail that all instantiated phenomenal properties are instantiated outside the subject's

mind (and that at least one phenomenal property is instantiated). This thesis is simply radical experiential externalism.

In sum: I have argued, with the help of the relatively mild assumptions from §2, for experiential externalism. In a more speculative spirit, I showed that, if we further accept the plausible Uniformity Principle, then radical experiential externalism follows. I am unsure whether the Uniformity Principle is true, but it deserves serious consideration.

## 9. Spatiality and transparency

In this section, I will compare this spatial argument for experiential externalism to a different phenomenological consideration that has often been advanced to support experientially externalist theories (like certain forms of representationalism and naïve realism): the *transparency* consideration. Exactly how to state this consideration is much disputed, and I will discuss this in more detail in the next chapter. For now, let me simply quote a formulation offered by Michael Tye in an article where he attempts to develop this consideration in detail:

When you introspect your visual experience, the only particulars of which you are aware are the external ones making up the scene before your eyes. You are not aware of those objects *and* a further inner object or episode. (2002, p. 139)

Tye's phenomenological claim is that introspection on experience reveals awareness only of external particulars. Coupled with a suitably formulated charity principle about the reliability of introspection, this supports the thought that experience involves awareness only of external particulars. And this is just what Tye's favored version of representationalism holds.

My argument is structurally similar. I advance the phenomenological claim that phenomenal properties introspectively seem instantiated in some kind of space. Given the Weak Charity Principle, phenomenal properties must actually be instantiated in some kind of space. And given the Physical Space Thesis, some phenomenal properties must be instantiated in physical space. This is what experiential externalist views, including Tye's form of representationalism, hold.

However, the phenomenological intuitions behind the transparency argument are much less robust than the phenomenological intuitions behind the spatial argument. Tye's phenomenological point is plausible when we focus on some phenomenal properties. But many have worried that his phenomenological point seems much less plausible when we consider phenomenal properties made salient by other cases: cases of blurry vision, afterimages, etc.<sup>35</sup> In these cases, to put the worry cautiously, it is not clear whether the relevant phenomenal properties seem to be instantiated in external or internal particulars. For instance, when philosophers have considered whether phenomenal blurriness even *introspectively seems* to be external or internal, opinion has been divided.

By contrast, as I argued in §5, our phenomenological sense of spatiality is robust in all of these cases, as well as in more typical cases. Indeed, I made a point in that section of considering many of the cases that allegedly make trouble for the phenomenological point articulated by Tye and others: I considered phenomenal haziness and blurriness, “two-dimensional” and “three-dimensional” phenomenal shape properties, etc.<sup>36</sup>

---

<sup>35</sup> See, for example, Block (1996), Kind (2003), Peacocke (1983), and Siewert (2003). For responses to some of these concerns, see Lycan (1996a) and (1996b) and Tye (1996a) and (2002).

<sup>36</sup> Schroer (2007) holds that something like the spatial phenomenology I've described is, in fact, the phenomenon that transparency theorists like Harman and Tye were trying to describe all along. For



If the phenomenological data concerning spatiality are less controversial than the phenomenological data concerning transparency, then spatiality considerations provide stronger support for experiential externalism than transparency considerations, at least dialectically. Still, there is no reason to think of these as *competing* considerations. One may think that both spatial considerations and transparency considerations favor experiential externalism, and perhaps even that these considerations mutually support one another.

In conclusion, robust phenomenological intuitions about spatiality support experiential externalism; this rules out many currently popular views about experience. Given this result, a natural question is whether radical experiential externalism is true. I express guarded optimism that radical experiential externalism may successfully be defended via the Uniformity Principle.

Subjective representationalism, as I pointed out earlier, is compatible with experiential externalism (and radical experiential externalism). But so are many other theories of experience. So, in the next chapter, I'll begin highlighting the distinctive advantages of subjective representationalism over other experiential externalist views.

---

instance, he says, "Properly understood, the so-called 'transparency' of visual experience ... consists of the fact that all the relevant visual phenomenal features are experienced as being located in a spatial field" (p. 414).

My argument puts pressure on this claim. For there are cases (like blurry vision cases) where the relevant spatiality intuition is robust, but the transparency intuition is fragile. It is not immediately obvious how to explain this on Schroer's view.

## Chapter 2.

# Exploring subjective representationalism

This chapter considers *representationalism*, a currently popular view about the nature of experience. The core idea of representationalism is that experiencing is to be analyzed wholly in terms of representing. More precisely, representationalism is the view that a mental state is an experience in virtue of being an appropriate type of representational state, perhaps in conjunction with playing a certain functional role; experiences are phenomenally similar or different wholly in virtue of having similar or different representational contents.<sup>1</sup>

Representationalism has several notable virtues. As I pointed out in the last chapter, it is compatible with experiential externalism, and we have strong reasons to accept experiential externalism. It is also compatible with radical experiential externalism, and we have at least some reason to accept this view, too.

Moreover, representationalism can readily explain what's going on in illusory and hallucinatory experiences – which are *prima facie* puzzling. Suppose that scientists

---

<sup>1</sup> This view is sometimes called “strong representationalism,” as there are related but substantially weaker views that go under the label “representationalism.” For example, Byrne (2001) argues that phenomenal character *supervenes on* representational properties, without making the stronger claim that mental states have their phenomenal character *in virtue of* having certain representational properties. As I will not be concerned with these weaker views in this chapter, I'll stick with the terminology in the text. Advocates of representationalism, as I use the term, include Carruthers (2000) and (2005); Dretske (1995) and (2003); Harman (1990); Hill (2009); Lycan (1996); Pautz (2010); and Tye (1995), (2000b), and (2002). Tye (2009) endorses a view very close to representationalism.

directly stimulate my brain so that I have a visual experience as of a red apple before me, even though no apple is there. I experience phenomenal redness. How can this be, since (we may suppose) there is nothing phenomenally red nearby?

The representationalist may reply that to experience phenomenal redness is to (appropriately) *represent* phenomenal redness. It is a platitude about representation that one can represent what doesn't exist. So my hallucinatory experience is simply a representation of an uninstantiated property; the puzzle is dissolved.

Third, representationalism offers the appealing prospect of a physicalist reduction of experience. As long as representation itself can be explained in wholly physicalistically acceptable terms, experiencing (which is analyzed in terms of representation) will also be explicable in wholly physicalist terms.

Finally, representationalism meshes well with our best current theories of perception. It is utterly commonplace for cognitive scientists to speak of what, e.g., vision represents: colors, textures, shapes, etc. Admittedly, it is much disputed exactly which properties are perceptually represented. But that some properties or other are perceptually represented is a very widespread assumption that has borne much fruit.

These are *prima facie* powerful motivations for representationalism. So I propose to set aside other experiential externalist views – most notably, naive realism – for present purposes. I do this, not for lack of interest, but for lack of sufficient time and space. Instead, my main goal in this chapter is to see what kind of representationalist theory is most plausible.

Specifically, a pressing question for representationalism is just what goes into the contents that determine the phenomenal similarities and differences among experiences. Contrast two sorts of properties. *Objective properties* are robustly mind-

independent in this sense: whether or not something instantiates an objective property does not depend essentially on whether or not that thing is appropriately related to the minds of subjects. Meanwhile, *subjective properties* are mind-dependent in a related sense: whether or not something instantiates a subjective property does depend essentially on whether or not that thing is appropriately related to the minds of subjects. (The subject in question might exist at some time other than when the subjective property is instantiated.)<sup>2</sup> An obvious question is whether the contents of experience concern only objective properties, or whether they concern at least some subjective properties.

To sharpen this question, recall that *phenomenal properties* are the properties (a) that we are apparently directly aware of via introspection, and (b) which generate, in the appropriate way, certain well-known philosophical puzzles about experience, such as the explanatory gap puzzle, the inverted spectrum puzzle, the zombie puzzle, and the puzzle concerning what Mary didn't know.

*Objective representationalism* is the view that all phenomenal properties are objective properties. As objective representationalism is usually developed, phenomenal colors are simply colors, and these in turn are something like dispositions to reflect light, or perhaps the categorical grounds of such dispositions. Objective representationalists say that phenomenal sounds, phenomenal tastes, phenomenal smells, etc. are also all objective properties of some sort.

An alternative view is *subjective representationalism*, which denies objective representationalism while affirming representationalism simpliciter. The core subjective representationalist idea is that at least *some* phenomenal properties (like phenomenal

---

<sup>2</sup> I admit that I'm not sure that this is quite the right way of drawing the objective/subjective distinction. That said, it's pretty clear that there is some important distinction of this sort to be drawn, and no one knows quite how to draw it. The distinction I propose in the text provides a good enough approximation to go on.

colors, phenomenal sounds, phenomenal tastes, etc.) are subjective properties.<sup>3</sup>

Subjective representationalism is not the much stronger view that *all* phenomenal properties are subjective properties, though it is compatible with this view.

Many prominent representationalists appear to be objective representationalists. It's difficult to tell for sure, since few explicitly consider the question. Still, when representationalists give examples of the sorts of properties represented in experience, they typically list exclusively objective properties. For instance, according to Michael Tye:

The most fundamental level of representation in visual experience, then, consists in what is represented in the array prior to any grouping.... for example, distance away, orientation, determinate color, texture, whether a discontinuity in depth is present there, and so on.<sup>4</sup>

These are all objective properties, at least as Tye construes them. Texture is manifestly an objective property, and Tye holds that colors are also objective properties.<sup>5</sup> Properties like distance away, orientation, and having such-and-such discontinuity in depth are also objective, though relational. For whether or not something is such-and-such distance from me, or oriented a certain way with respect to me, or whatever, does not essentially depend on whether that thing bears a certain relation to *my mind*. Rather, it depends on whether or not it bears a certain relation to *me*.<sup>6</sup>

---

<sup>3</sup> Sydney Shoemaker has explored many views in the vicinity of subjective representationalism (1994a, 1994b, 2000, 2001, 2003, 2006). However, Shoemaker disavows reductivist ambitions, so strictly speaking he does not qualify as a representationalist, as I'm using the term.

<sup>4</sup> Tye (1996, pp. 122-123).

<sup>5</sup> E.g., see his (2000b, ch. 7).

<sup>6</sup> I thank Brad Thompson for helpful discussion on this point.

Tye also discusses phenomenal properties associated with non-visual experiences. He conjectures that the contents of such experiences may concern properties like “pitch, tone, loudness, pungency, muskiness, sweetness, saltiness, [and] sourness.”<sup>7</sup> He makes much of the fact that such experienced properties as sounds and smells are “publically accessible,”<sup>8</sup> and makes a similar point about taste: “We taste things by tasting their tastes. One and the same taste can be tasted by different people.”<sup>9</sup> It is possible for a subjective property to be publically accessible, but given Tye’s objectivism about colors, it would be natural for him to embrace a similar view about pitch, tone, loudness, etc. So these comments suggest, though they don’t demonstrate, that Tye thinks of these phenomenal properties as objective.

Alex Byrne pins down a thesis which he calls (CV) – roughly, the thesis that experiences have representational content. He then says:

... if (CV) is supported by an inference to the best explanation of illusions, then one might expect perceptual content to be relatively thin. Visual illusions, as the object of study in the visual sciences, concern properties like shape, motion, colour, shading, orientation and the like.<sup>10</sup>

Byrne later endorses this approach, saying, “We may provisionally conclude that perceptual content is relatively thin.”<sup>11</sup> (By saying that perceptual content is “relatively thin,” Byrne means to exclude kind properties – like the property of *being a pine tree* or *being a lemon* – from perceptual content.) Byrne never backs down from this provisional

---

<sup>7</sup> Tye (2000b, p. 50).

<sup>8</sup> Ibid, p. 49.

<sup>9</sup> Ibid.

<sup>10</sup> Byrne (2009, p. 449).

<sup>11</sup> Ibid, p. 450.

conclusion in that paper.<sup>12</sup> Byrne holds that color is an objective property,<sup>13</sup> and all of the other properties he lists are manifestly objective.

Similarly, leading representationalist Fred Dretske says this that “... *what* (properties) one is [aware] *of* in having the [experience of seeing a pumpkin] are color, shape, texture, distance, and movement ...”<sup>14</sup> So far as I know, Dretske has not published a view about whether colors are objective, but the other properties he lists certainly are.

And consider Christopher Hill, who analyzes experiencing in terms of the representation of what he calls “A-properties” – he uses this term more or less as I have been using the term “phenomenal properties.” Hill suggests that the A-properties associated with visual experiences of size and shape “are the values that are obtained when certain computable functions are applied to angular properties (together with various other quantities).” Hill elaborates that

(i) the functions are in effect constancy transformations, (ii) they each take a number of arguments in addition to angular sizes and angular shapes, all of which are relevant to computing constancies, (iii) their values fall short of being genuine constancies, and (iv) their values are A-sizes and A-shapes.<sup>15</sup>

Hill states that angular sizes are “defined with reference to the nodal point of the lens of the observer’s eye” and that angular shapes “can for present purposes be identified with the set of all visual angles that are subtended by pairs of points on the boundary of the

---

<sup>12</sup> Strictly speaking, Byrne does not endorse representationalism (as I’ve defined it) in this paper, but only a weaker thesis – see footnote 1. Still, his view is compatible with representationalism, and it’s notable that the contents he attributes to experiences involve exclusively objective properties.

<sup>13</sup> Byrne and Hilbert (2003).

<sup>14</sup> Dretske (1999, p. 112).

<sup>15</sup> Hill (2009, p. 165).

object's facing surface."<sup>16</sup> He proceeds to generalize this account to all A-properties – that is, to all phenomenal properties – not just to those phenomenal properties associated with visual experiences of size and shape. All A-properties are produced by constancy transformations, but fall short of being genuine constancies.<sup>17</sup>

Hill's proposal is a form of objective representationalism.<sup>18</sup> For angular sizes and shapes are objective properties in our sense: while they are sensitive to the location of the observer, they are wholly mind-independent. (Similarly, while the property *being within a mile of a coffee table* is sensitive to the location of coffee tables, it is a wholly mind-independent property.) And any property that is a computable function of only mind-independent properties will itself be a mind-independent property.

Indeed, Hill even considers the suggestion “that A-properties should be seen as involving or depending constitutively on internal factors” and rejects this proposal as “largely unmotivated.”<sup>19</sup> This remark rules out neural properties and the like from serving as input to the computable functions which constitute A-properties; Hill shows no sympathy for subjective representationalism here.

So, among many representationalists, objective representationalism seems to be orthodoxy. I underscore once more that none of the remarks quoted above provide conclusive evidence that Tye, Byrne, Dretske, or Hill are objective representationalists. The point is only that all candidate phenomenal properties that they cite seem objective; objective representationalism is the most natural extension of their views.

---

<sup>16</sup> Ibid, p. 162.

<sup>17</sup> Ibid, pp. 165-168.

<sup>18</sup> Hill says several times that A-properties are not “objective,” but it's clear from context that his use of the term differs from ours. Roughly, Hill uses the term to refer to something like *observer*-independence, not *mind*-independence.

<sup>19</sup> Ibid, p. 167.



My aim in this chapter is to provide reasons for preferring subjective representationalism to objective representationalism. I'll show that objective representationalism cannot accommodate the possibility of illusion-free phenomenal inversion, while subjective representationalism can. Moreover, the intuition that illusion-free phenomenal inversion is possible is robust.

Additionally, a dominant motivation for objective representationalism is its ability to naturally accommodate the *transparency phenomenon*, which we considered in the previous chapter. But I will develop a particular version of subjective representationalism that holds that at least some phenomenal properties have the form *appropriately causing mental state M in me*. I will show that this version of subjective representationalism is just as well-placed as objective representationalism to accommodate the transparency phenomenon.

To be clear, my claim is not merely that subjective representationalism can *accommodate* the transparency motivation for objective representationalism. My claim is that it can accommodate this motivation *just as well* as objective representationalism. To the extent that this motivation supports objective representationalism, it also supports subjective representationalism.

Here is a map of the terrain ahead. Section 1 describes one brand of subjective representationalism and argues that it handles phenomenal color inversion intuitions better than objective representationalism. Section 2 generalizes the argument to other phenomenal properties. Section 3 shows that the transparency motivation for objective representationalism equally motivates subjective representationalism, and Section 4 rebuts a recent objection. Closing remarks appear in Section 5.

## 1. Capturing spectrum inversion intuitions

My main goal in this section is to demonstrate one substantial advantage of subjective representationalism over objective representationalism. We have a robust intuition that there are possible cases of *phenomenal inversion without illusion*. Objective representationalism plainly cannot accommodate such intuitions. Subjective representationalism can.

Here is a familiar phenomenal inversion case. Consider two subjects, Jack and Jill, who are both looking at a ripe (red) bell pepper. What it's like for Jack to see tomatoes, cherries, and ripe bell peppers is (in a salient respect) the same as what it's like for Jill to see limes, frogs, and unripe bell peppers; these experiences share a common phenomenal character. Similarly, what it's like for Jack to see unripe bell peppers is the same as what it's like for Jill to see ripe bell peppers; those experiences, too, share a common phenomenal character. Call this the *Color Inversion Case*. This case seems perfectly conceivable.

It also seems perfectly conceivable that Jack's experience and Jill's experience could be wholly veridical – that neither of them are misperceiving this ripe red bell pepper. We can imagine that their experiences are normal for them: for example, Jack's experiences of things we call “red” have always been like Jill's experiences of things we call “green,” and vice versa. Both use color terms standardly (they both call ripe bell peppers “red” and unripe ones “green”). Their sorting behavior is identical – they agree completely on which items go in which piles when asked to sort things by color. Given these stipulations about the case, it seems arbitrary to say that one of their experiences of the ripe red bell pepper is veridical and the other falsidical (and perhaps even more counterintuitive to say that both of their experiences *must* be falsidical).

While there is an extensive literature attempting to show that Jack's experience and Jill's experience cannot both be veridical, it is safe to say that none of these arguments have been found widely compelling; our intuition that cases like the Color Inversion Case need not involve any illusion is robust.<sup>20</sup> The relevant intuition is not merely that Jack and Jill both have correct *beliefs*, but also that their *experiences* are not falsidical.

Objective representationalism, as it is ordinarily developed, is incompatible with the possibility of illusion-free spectrum inversion. Objective representationalists analyze what it's like to have color experiences in terms of the properties that those experiences represent. Typically, objective representationalists say that the represented properties are simply colors, and that these in turn are either (roughly) dispositions to reflect light or the categorical grounds of such dispositions.<sup>21</sup> So when Jack and Jill have different experiences while looking at a ripe bell pepper, Jack represents the bell pepper as having the color property *phenomenal F-ness*, and Jill represents it as having the color property *phenomenal G-ness*.

But perhaps the bell pepper can instantiate both color properties, so that Jack's experience and Jill's experience are both veridical? Unfortunately not. For it's very plausible (and objective representationalists normally accept) that the phenomenal property associated with Jack's experience of ripe bell peppers (phenomenal F-ness) is incompatible with the phenomenal property associated with his experience of unripe bell peppers (phenomenal G-ness). These properties are incompatible in this sense: nothing can wholly instantiate both phenomenal F-ness and phenomenal G-ness at a given time.

---

<sup>20</sup> For further defense of this point, see Chalmers (2006), Shoemaker (1982, 1994a, 1994b, 2000, 2003), and Thompson (2009).

<sup>21</sup> See, for example, Byrne (2001) and (2006); Byrne and Hilbert (2003); and Tye (1995) and (2000b).

So if the bell pepper wholly instantiates any phenomenal color property, then it can instantiate at most one of the two incompatible properties attributed to it by Jack's and Jill's experiences.

The problem is generated by the following commitments of the objective representationalist concerning the Color Inversion Case<sup>22</sup>:

(1) No object can wholly instantiate phenomenal F-ness and phenomenal G-ness at a given time.

(2) Jack's experience of the ripe bell pepper represents something as wholly instantiating phenomenal F-ness, and Jill's experience of the ripe bell pepper represents something as wholly instantiating phenomenal G-ness.

(3) Jack's experience and Jill's experience attribute these respective properties to the same thing at the same time.

(1) is common ground to almost all theories of experience. The phenomenal property paradigmatically associated with my experiences of red things seems deeply incompatible with the phenomenal property paradigmatically associated with my experiences of green things; it seems unintelligible that they could be co-instantiated.<sup>23</sup>

As for (2), we stipulated that Jack and Jill have different experiences in the Color Inversion Case. Since objective representationalists analyze differences in experience via differences in objective properties represented, they must accept something like (2).

---

<sup>22</sup> Roughly this problem is also noted by Chalmers (2006), Egan (2006), and Thompson (2007).

<sup>23</sup> But cf. Harman (2001) for a contrary opinion.

Usually, objective representationalism is motivated partly via the transparency consideration (more on this in section 2). For now, we can simply note that the usual form of objective representationalism says that phenomenal color properties are represented as being instantiated in ordinary environmental objects (like bell peppers). So objective representationalists normally accept (3).

It follows from (2) and (3) that Jack's experience represents an object as wholly phenomenally F, and Jill's experience represents the same object as wholly phenomenally G. But (1) says that no object can be both wholly phenomenally F and wholly phenomenally G. So Jack's experience and Jill's experience cannot both be veridical. Since the (canonical) objective representationalist is committed to (1) - (3), she must deny that Jack's experience and Jill's experience are both veridical in the Inversion Case. The point readily generalizes: objective representationalism cannot countenance illusion-free color inversion.

Can subjective representationalism do better? At first glance, it appears not. Everyone, subjective representationalists included, should accept (1). And any representationalist, objective or not, must accept something like (2). For Jack and Jill have phenomenally different experiences, and representationalists are committed to understanding such differences in terms of representational differences. Finally, subjective representationalists (just like objective representationalists) can deny (3) by denying the transparency consideration. But this is a compelling datum, and I'd like to see whether we can keep it.

Egan (2006) provides a neat solution to the problem. The solution is motivated in part by the point that, to understand the contents of certain propositional attitudes,

we need more fine-grained objects than possible worlds.<sup>24</sup> The following example illustrates why possible-worlds content isn't fine-grained enough. Suppose that from noon to midnight I am sitting on the bus to Chicago, with my eyes closed the entire time. Sometime in the middle of the trip, I wonder what time it is *now*. No matter how much information I acquire about which possible world is actual, this alone will not tell me what time it is now. To learn this, I must learn something about – to put things intuitively – *where I am located* in the world.

A natural thought is to introduce *centered possible worlds*. There are many ways of understanding what these are, but I'll take a centered possible world to be a world with a "marked" individual and time. So here is a centered possible world: <the actual world, Barack Obama, December 1 2010>. Propositional attitudes may be assigned centered possible world contents. Since centered possible world contents are strictly more fine-grained than possible world contents, we don't lose any modeling capabilities when we switch to them. But we do gain some powerful new modeling capabilities. For example, we can now say what I'm wondering when I wonder what time it is now. I'm not wondering which world is actual, but rather which temporal location I'm in now.

Possible worlds contents determine functions from possible worlds to truth values. By contrast, centered possible worlds contents determine functions from *centered* possible worlds to truth values. If we are modeling contents with ordinary possible worlds, then any two actual individuals who represent incompatibly can't both be correct. If I believe that p and you believe that not p, we can't both be right. But

---

<sup>24</sup> For compelling arguments that we need something more fine-grained than possible worlds to understand content, see Perry (1979) and Lewis (1979). But note that Perry does not endorse a centered-contents approach to explaining the relevant phenomena.

introducing centered possible worlds permits for cases in which two actual individuals represent incompatibly, and both are correct.

To see this, suppose that at time  $t$  Sammy is in Chicago and Sally is in New York. And, at time  $t$ , it's raining in Chicago but not in New York. Sammy thinks the thought he would express by saying, "It's raining here," and Sally thinks the thought she would express by saying, "It's not raining here"; these thoughts have quite different truth conditions. The truth of Sammy's thought requires that it be raining in some salient location in *Chicago*, while the truth of Sally's thought requires that it be raining in some salient location in *New York*. Still, both think something true.

Still, there is an important intuitive sense in which Sammy's thought and Sally's thought represent incompatibly, even though both of them think true thoughts. The centered worlds theorist can capture this sense of representational incompatibility by saying that Sammy's thought and Sally's thought have incompatible centered contents: one has the content that it is raining in some salient location near the marked center, and the other has the content that it is not raining in some salient location near the marked center. Assessed at the actual world centered on Sammy at  $t$ , only the former content is true. Assessed at the actual world centered on Sally at  $t$ , only the latter content is true.

Let's define an *ordinary function* as a function from possible worlds to extensions. Standardly, it is thought that any property (like *being a dog* or *being red*) determines an ordinary function. The property *being a dog*, for example, determines the function that takes as input any possible world and delivers as output every dog in that world. The possible worlds theorist will likely model predicates (like "is a dog") in terms of ordinary functions.

But now consider *centered functions*, which are functions from centered possible worlds to extensions. The centered possible worlds theorist can model predicates (like “is a dog” or “is now red” or “is near me”) in terms of centered functions instead of ordinary functions.

There is a centered function corresponding to every ordinary function, since, for every function from possible worlds to extensions, there is a corresponding function from centered possible worlds to extensions. (Intuitively, these are functions from centered possible worlds to extensions that ignore the marked center.) So any theoretical work done by ordinary functions can also be done by centered functions.

But there are also centered functions corresponding to no ordinary functions. (Intuitively, these are functions from centered possible worlds to extensions that do not ignore the marked center.) So there may be theoretical work that centered functions can do and ordinary functions cannot do.

As I said, properties are normally thought of as determining ordinary functions. But, from here on out, I’ll instead talk as though properties determine centered functions. (This use of the term “property” is non-standard, but it is a natural extension of the standard use.) This lets us distinguish between *ordinary properties* and *centered properties*. An ordinary property is a property that determines a centered function corresponding to an ordinary function, while a centered property is a property that determines a centered function corresponding to no ordinary function. (Thus every property is either ordinary or centered, and no property is both.)

Some examples will help. The property of *being a dog* is ordinary; it determines a centered function from any centered possible world to the dogs in that world. (The function ignores any information about the marked centers of these worlds, so it



corresponds to an ordinary function.) Meanwhile, the property *being me* is centered; it determines a centered function from any centered possible world to the individual marked at the center of that world. There is no ordinary function corresponding to this centered function, since the centered function puts to use information about the marked centers of these worlds.

Notice that centered properties are not instantiated (or uninstantiated) *simpliciter* in the actual world, or in any other possible world; they are instantiated (or uninstantiated) only relative to a center in the actual world, or relative to a centered possible world. (Analogy: we cannot evaluate the thought that New York is here *simpliciter*, but only given a context.)

The pivotal suggestion of Egan (2006) is that the subjective representationalist can exploit centered properties to explain what's going on in the Inversion Case. The thought is that Jack's experience and Jill's experience represent things as being the same way – they have the same centered worlds content, ascribing the same properties – even though the veridicality conditions of their experiences differ given their different locations in the world. Here ends my recapitulation of Egan (2006); I'll spell out my preferred version of the view below.

I should mention that, while some have used the label “subjective” to describe content modeled in terms of centered possible worlds, that is not how I'm using the term in this chapter. I reserve the term “subjective” to refer to any property whose instantiation depends essentially on whether or not the thing instantiating it is appropriately related to the minds of subjects. I'll speak of “*de se* content” to describe content modeled in terms of centered possible worlds.

Pressing forward: what property do both Jack and Jill ascribe to the ripe bell pepper? Let's start with a simple subjective representationalist proposal according to which phenomenal color properties are centered properties of the form *causing mental state M in me* (where "me" picks out the marked individual of the centered world). Let's remain neutral on what the mental state in question is. It might be the experience itself (perhaps picked out via its neural properties), or some sub-personal perceptual state, or some other mental state.

On this view, phenomenal redness is the centered property of causing some specific type of mental state – call it  $M_{\text{red}}$  – in the marked individual at some centered world. There is something it's like for me to experience phenomenal redness, and what it's like is to be explained wholly in terms of the fact that my experience of phenomenal redness represents something as causing a mental state of type  $M_{\text{red}}$  in me.

There is an immediate but misguided objection to this view. The objection goes like this: when I introspect on my experiences involving phenomenal redness, I don't seem to be aware of a complex property, at least not of the sort *causing mental state  $M_{\text{red}}$  in me*. I can't introspectively separate a causal element when I consider my experience of phenomenal redness.

The objection is defused by noting that our experiences need not represent this complex property –*causing mental state M in me*–as a complex property. Experience may represent the property while *misrepresenting* it as simple.<sup>25</sup> Alternatively, experience may represent the property while remaining silent on whether the property is simple or

---

<sup>25</sup> It is a cost of this view that it ascribes misrepresentation to experience; it's not clear to me how large this cost is. In any case, the same problem afflicts objective representationalism. Objective representationalists typically say that phenomenal colors are enormously complicated dispositional properties, but phenomenal colors certainly don't seem to be enormously complicated dispositional properties when I introspect them.

complex. Introspection alone doesn't refute the view that phenomenal properties have such hidden complexity.

Here is a better objection. Consider the state of affairs obtaining at the moment of the Big Bang, which presumably caused everything afterwards. (If one doesn't think that states of affairs are the relata of causal relations, no matter. The point can easily be rephrased in terms of *events*, or *properties*, or whatever one thinks the relata of causal relations are.) Specifically, that state of affairs has caused every mental state I've ever undergone. So the Big Bang instantiates every phenomenal color property that I've ever experienced (relative to me now).<sup>26</sup>

This is counterintuitive on two counts. First, it's just implausible that the Big Bang really does instantiate all of those phenomenal colors relative to me now. Second, and more seriously, the instantiation of certain phenomenal colors excludes the instantiation of other phenomenal colors. For instance, nothing can be both phenomenally red and phenomenally green (at the same time and place, relative to the same center). But, on the view we're considering, the Big Bang instantiates these and many more incompatible phenomenal colors relative to me now.

To deal with such problems, we can try restricting the causal relation in question. To be phenomenally red, we might say, a thing must do more than simply cause an  $M_{\text{red}}$ -state in me in any way at all. A thing must further stand in the kind of causal relation to my  $M_{\text{red}}$ -state that is typical of cases of veridical perception. The idea is that, when I successfully perceive a red mango, the mango causes an  $M_{\text{red}}$ -state in me in a particular way. I'll abbreviate this by saying that the thing must *appropriately* cause the  $M_{\text{red}}$ -state. The new proposal – and the one I will explore in the rest of this chapter

---

<sup>26</sup> I owe this example to Eric Lormand.

– is that phenomenal colors are centered properties of the form *appropriately causing mental state M in me*.

Appropriate causation is incompatible with all kinds of “deviant” causal chains. It is also incompatible with causal chains that are, to phrase the idea intuitively, “too long” or “too short.” When I successfully perceive the mango, both the Big Bang and a certain state of my retina are causally implicated in the production of the  $M_{\text{red}}$ -state. But neither causes the  $M_{\text{red}}$ -state in the appropriate way, so, on the present proposal, neither is phenomenally red.

We now have the resources to handle the Inversion Case without imputing misrepresentation. When Jack looks at a ripe bell pepper and Jill looks at an unripe bell pepper, their experiences both ascribe the following property to the entire surface of the respective bell peppers: *appropriately causing mental state  $M_1$  in me*. What it’s like to have an experience quite generally is determined by the content of that experience – in this case, by what properties it ascribes. Since Jack’s experience and Jill’s experience ascribe the same property, their experiences are phenomenally the same (at least with respect to phenomenal color). Moreover, the ripe bell pepper does appropriately cause  $M_1$  in Jack, and the unripe bell pepper also appropriately causes  $M_1$  in Jill. So both of their experiences are veridical in this respect.

In what sense, then, is phenomenal redness incompatible with phenomenal greenness? I remarked above that objects instantiate ordinary properties simpliciter, but objects do not instantiate centered properties simpliciter. Rather, objects instantiate centered properties *relative to a marked center* (a marked individual and time). So the idea is that no object can wholly instantiate phenomenal redness relative to a subject at a time *and* wholly instantiate phenomenal greenness relative to the same subject and time.

But an object can wholly instantiate phenomenal redness relative to a given subject at a given time and wholly instantiate phenomenal greenness relative to *another* subject at that time (or relative to the same subject at *another* time). That is the sense in which phenomenal redness and phenomenal greenness are incompatible properties.

One might worry that talk of properties being instantiated “relative to a marked center” is incoherent, or at least that it is bad metaphysical manners to speak this way. I offer three responses to this worry. First, we can easily pick out centered properties using perfectly acceptable notions: a centered property determines a function from any possible world with a marked individual and time to a set of objects. This way of talking doesn’t lead to any obvious technical problems. So we have a way of translating centered-property talk into perfectly respectable talk. Second, we seem to need centered properties to deal with lots of propositional-attitude contents.

Finally, and most relevantly for present purposes, centered properties permit us to capture two deep but apparently incompatible intuitions. Consider again the case where Jack and Jill are both looking at the same ripe bell pepper and having experiences that differ with respect to phenomenal color. Intuitively, both of their experiences are veridical with respect to phenomenal color. But, intuitively, the properties that they ascribe to the bell pepper seem to be incompatible in some important sense.

Surprisingly, our current proposal reconciles these intuitions. Jack correctly ascribes to the bell pepper the property *appropriately causing mental state  $M_1$  in me*, and Jill correctly ascribes to the bell pepper the property *appropriately causing mental state  $M_2$  in me*. So both of their experiences are veridical with respect to phenomenal colors.

At the same time, Jack could not correctly ascribe to the bell pepper both the property *appropriately causing mental state  $M_1$  in me* and the property *appropriately causing*

*mental state M<sub>2</sub> in me.* For, plausibly, no patch of the bell pepper's surface could appropriately cause both mental state M<sub>1</sub> and mental state M<sub>2</sub> in a given subject at a given time. So these two phenomenal colors are incompatible in an important sense: they cannot both be wholly instantiated in anything at a given time, relative to a given subject. So subjective representationalism can accommodate the whole raft of intuitions about the Color Inversion Case.<sup>27</sup>

One might wonder whether objective representationalism can do just as well by using centered properties instead of ordinary properties to analyze the content of color experiences. For just as there are both objective and subjective ordinary properties, there are both objective and subjective centered properties. I've already given examples of subjective centered properties, so here's an example of an objective one: the centered property "being far away." The sun instantiates the property *being far away* relative to the earth now, and its instantiation of this property does not depend essentially on its relation to the minds of any subjects.

But even if objective representationalists make use of centered properties, they cannot accommodate our intuitions about illusion-free color inversion. We have the robust intuition that when Jack and Jill both look at the same ripe bell pepper and have different experiences with respect to phenomenal color, both experiences may be veridical. The intuition remains even if the case is one where the bell pepper has all the same objective centered properties (or near enough) "for Jack" and "for Jill." Jack and Jill might be standing right next to each other at the same time; though their eyes can't

---

<sup>27</sup> I think that this kind of subjective representationalist account can easily handle examples like Ned Block's Inverted Earth case (1990) and Brad Thompson's cases involving spatial inversions (2010), though I don't have room to walk through such cases here. Such cases seem deeply problematic for objective representationalism.

literally be in the same place, there need not be any relevant difference between their positions.

So the objective centered properties of the bell pepper are the same whether we take *Jack at noon* or *Jill at noon* as the marked center. The only properties of the bell pepper that differ “relative to Jack at noon” and “relative to Jill at noon” are mental: Jack and Jill are in different mental states. So the introduction of centered properties does not help the objective representationalist capture our intuitions about illusion-free color inversion; it helps only the subjective representationalist.

Observe that this proposal is silent about colors *simpliciter*. My subjective representationalist proposal characterizes only phenomenal colors. One may further hold either that colors *simpliciter* are phenomenal colors, or that they are not. On the latter view, one may also hold either that colors themselves are objective or subjective properties. The brand of subjective representationalism I’ve defended does not commit one to any of these particular views, though it may accompany some views about the nature of color better than others.

As Billy Dunaway pointed out to me, the introduction of *indexical* content is what does the heavy-lifting in this account. But there are lots of ways of analyzing indexical contents: via *de se* content (as I do), via Kaplanian characters (see Kaplan (1989a) and (1989b)), via Fregean senses, etc. So why do I discuss only the *de se* analysis?

For two reasons. First, as Brogaard (2010) emphasizes, there are several other potential further uses of *de se* content in understanding experience. Second, and more importantly, it seems to me that this is the only approach compatible with the determinacy of experience. I return to this issue in §4, especially footnote 39.

In sum, subjective representationalism that invokes *de se* contents with centered properties of the form *appropriately causing mental state M in me* can accommodate a set of robust but apparently incompatible intuitions about phenomenal color inversions. Objective representationalism can't do this, with or without centered properties. This is a substantial advantage of subjective representationalism over objective representationalism, at least in explaining color experiences. In the next section, I examine whether the argument extends to phenomenal properties besides phenomenal colors.

## 2. Extending the argument

The argument thus far has been restricted to phenomenal colors. But it extends to any phenomenal determinable of type  $Q$  with determinates  $Q_1 \dots Q_n$ , such that we can coherently conceive of an illusion-free inversion case meeting these conditions (analogous to (1), (2), and (3) above):

(1\*) Nothing can wholly instantiate  $Q_1$  and  $Q_2$  at a given time.

(2\*) Jack's experience represents  $Q_1$ , and Jill's experience represents  $Q_2$ .

(3\*) Jack's experience and Jill's experience attribute these respective properties to the same thing at the same time.

In this section, I'll attempt to generalize the argument. It's plausible that there are cases of illusion-free sound inversion, taste inversion, touch inversion, smell inversion, bodily



sensation inversion, etc., that meet these conditions. If there are such cases, then the argument generalizes to phenomenal sounds, tastes, touches, smells, bodily sensations, etc.

Here is a situation much like the Color Inversion Case, but involving phenomenal sounds; call it the *Sound Inversion Case*. Suppose Jack and Jill both hear a loud sound but have different experiences. For, in general, the experiences Jack has when he hears loud sounds are just like the experiences Jill has when she hears quiet sounds, and vice versa. These experiences are typical for them, and Jack and Jill are behaviorally alike. For example, they call the same sounds “loud” and “quiet,” and respond alike when asked to sort sounds by how loud they are.<sup>28</sup>

According to the representationalist, Jack’s experience of a particular loud sound and Jill’s experience of the corresponding quiet sound have the same phenomenal character because they both represent the same phenomenal property – call it *phenomenal loudness*. Similarly, Jack’s experience of a particular quiet sound and Jill’s experience of the corresponding loud sound have the same phenomenal character because they both represent another phenomenal property – call it *phenomenal quietness*. It seems perfectly conceivable that both of their experiences are perfectly veridical.

The Sound Inversion Case meets conditions (1\*), (2\*), and (3\*), but there are some nuances in seeing how. For (1\*) to be satisfied, there must be two phenomenal properties  $Q_1$  and  $Q_2$  such that nothing can wholly instantiate  $Q_1$  and  $Q_2$  at a given time. The natural proposal is that  $Q_1$  and  $Q_2$  are phenomenal loudness and phenomenal quietness, respectively. But one might worry that something can instantiate both of

---

<sup>28</sup> Those who have trouble coherently conceiving of such a case may instead substitute a “shifted experience” case where Jack’s experience of a sound of  $n$  decibels is, in general, just like Jill’s experience of a sound of  $0.8n$  decibels.

these properties at a given time. For suppose I hear a sound created by a loud guitar and a quiet piano. Plausibly, something (perhaps my experience, or a body of air, or whatever) then instantiates both phenomenal loudness and phenomenal quietness in such a case.

This isn't problematic. While it might be that a single thing instantiates both phenomenal loudness and phenomenal quietness in the Sound Inversion Case, nothing *wholly* instantiates both of these properties. Part of the object – here I'm using the term "object" very permissively – instantiates phenomenal loudness, and part of it instantiates phenomenal quietness. I can readily swap my attention between these parts and the whole object. But I cannot attend to anything that wholly instantiates both properties.<sup>29</sup>

The representationalist must also accept this version of (2\*): Jack's experience represents phenomenal loudness, and Jill's experience represents phenomenal quietness. Representationalists analyze similarities and difference in what-it's-like to have any experiences in terms of similarities and differences in the representational contents of those experiences. So they must analyze the differences in Jack's and Jill's experiences via a difference in the representational contents of their experiences, and talk of "phenomenal loudness" and "phenomenal quietness" is merely a convenient way of labeling the relevant phenomenal properties figuring in these different contents.

Finally, representationalists sympathetic to transparency considerations, to be discussed in the next section, must accept (3\*): Jack's experience and Jill's experience attribute these respective properties to the same thing at the same time. Transparency theorists, as we shall see, think that phenomenal sounds are features of objects outside the subject. And we may build into the Sound Inversion Case that Jack and Jill are

---

<sup>29</sup> The same point holds even if a single entity – say, a stereo system – reproduces the sound of a loud guitar and a quiet piano. Even so, we can experientially distinguish between the phenomenally loud part and the phenomenally quiet part.

appropriately related to the same environmental objects, or at least environmental objects of the same type. (Again, “object” is being used loosely, so that bodies of air count as objects).

Since we have an illusion-free inversion case that satisfies (1\*), (2\*), and (3\*), the arguments of section 1 apply directly. As before, I propose that we analyze properties like *being phenomenally loud* as properties of the form *appropriately causing mental state M in me*.

The argument extends readily to phenomenal properties associated with all perceptual experiences. For example, it applies to phenomenal tastes (like phenomenal sourness and phenomenal sweetness), phenomenal touches (like phenomenal softness and phenomenal hardness), and phenomenal smells (like phenomenal vanilla-scentedness and phenomenal cinnamon-scentedness). It also applies to hedonic experiences (like phenomenal burning-pains and phenomenal freezing-pains) and bodily sensation experiences (like phenomenal stillness and phenomenal dizziness). Running through the above considerations in each case would be tedious and space-consuming, so I will not do so. It is clear how the arguments would go.

As in the previous section, this argument characterizes only certain *phenomenal* properties. It is silent about the nature of sweetness *simpliciter*, softness *simpliciter*, vanilla-scentedness *simpliciter*, etc.

Are there any phenomenal properties that escape the net of this argument? Perhaps phenomenal properties associated with experiences of space do, like the phenomenal property associated with my seeing something as circular or far away. Thompson (2010) provides extensive arguments that illusion-free inversion of even

these properties is possible. If Thompson's arguments succeed – and here I remain neutral on whether they do – then my argument extends to them, too.

Another candidate for phenomenal properties that elude this style of argument are those phenomenal properties, if there are any, associated with the “phenomenology of intentionality.”<sup>30</sup> It is very contentious whether there are such phenomenal properties. Indeed, the arguments here may bear on the question of whether there is a distinctive phenomenology of intentionality. If all uncontroversial examples of phenomenal properties are susceptible to illusion-free inversion, then that provides at least some reason for thinking that a property which resists such inversion is not really a phenomenal property. I won't explore this line of reasoning further; it's fine by me if there are objective phenomenal properties associated with intentionality.

I raise these examples of candidate objective phenomenal properties to highlight a crucial point. Subjective representationalism is not the view that *all* phenomenal properties are subjective. It is not even the view that *most* phenomenal properties are subjective. It is merely the view that *some* phenomenal properties are subjective. And, as the quotes I assembled in the introduction strongly suggest, many prominent representationalists appear to think that all phenomenal properties are objective. So if I've successfully argued that even a few phenomenal properties – say, just the phenomenal colors – are subjective, this would constitute progress. Of course, I think that my arguments extend considerably further than that.

### 3. Transparency and representationalism

---

<sup>30</sup> The literature on this has grown explosively, but a good starting point is Horgan and Tienson (2002).

Perhaps the most commonly cited consideration in favor of objective representationalism is the *transparency* consideration, which I discussed briefly in the previous chapter. Objective representationalists don't agree on exactly how to formulate this consideration, but their formulations do have much in common. I will argue that subjective representationalism can account for the transparency phenomenon at least as well as objective representationalism does, however that phenomenon is best articulated.

I will focus on Gilbert Harman's and Michael Tye's articulations of transparency. I select Harman (1990) because it is the most cited recent discussion of transparency; I select Tye (2002) because it is an especially thorough elaboration of how the transparency intuition supports objective representationalism.

Let's begin with Harman's seminal discussion. Here is the pivotal passage:

Look at a tree and try to turn your attention to intrinsic properties of your visual experience. I predict that you will find that the only properties there to turn your attention to will be properties of the tree, including relational properties of the tree 'from here'.<sup>31</sup>

It's worth highlighting that Harman does not hesitate to talk about the "relational properties of the tree"; the transparency consideration, whatever it is, is not a consideration about whether experience presents us with relational or non-relational features.

---

<sup>31</sup> Harman (1990, p. 39).

I assume that, to follow Harman's instruction to try to turn one's attention to intrinsic properties of one's visual experience, we use introspection. Harman makes both (i) the negative claim that introspection turns up no intrinsic properties *of visual experience*, and (ii) the positive claim that introspection turns up properties *of the tree*.

But what is the general type of object whose properties we can introspectively attend to? Is it the class of *external objects, environmental objects, non-mental objects*, or what? Harman's discussion is none too clear on this point.

Now consider Tye's remarks. Here are some representative passages:

Whatever the nature of the qualities of which we are directly aware when we focus upon how the surfaces before us *look*, these qualities are not experienced as qualities of our experiences but rather as qualities of the surfaces.<sup>32</sup>

When you introspect your visual experience, the only particulars of which you are aware are the external ones making up the scene before your eyes. You are not aware of those objects *and* a further inner object or episode.<sup>33</sup>

Like Harman, Tye makes both a negative claim and a positive one. Tye's negative claim is that attending to how things look via introspection does not seem to turn up anything inner – it reveals no experiences, inner objects, or inner episodes. Tye's positive claim is that it does turn up (apparent) qualities of external things, like surfaces.

The common phenomenological point that we can extract from Harman and Tye is this: introspection seems to reveal *only* properties of certain kinds of *outer stuff* – perhaps external or environmental or intentional objects. It does *not* seem to reveal properties of *inner stuff*, like the experience itself. I'm using the term "stuff" because of

---

<sup>32</sup> Tye (2002, p. 138).

<sup>33</sup> Ibid, p. 139.

its pliability. Expanses of sky, experiential events or processes, and ordinary objects all count as *stuff* in my sense.

It is clear how this motivates objective representationalism. Unless introspection is massively unreliable – which we may reasonably think it's not – these introspected phenomenal properties (like phenomenal colors) *are* properties of outer stuff, not inner stuff. This is just what objective representationalism says. According to a common version of objective representationalism, phenomenal colors are something like dispositions to reflect light, or the categorical grounds of those dispositions. Outer stuff – like apples, expanses of sky, pitchers of beer, and so on – instantiates such properties. (Inner stuff does too, but presumably we're not normally aware of *those* properties.) The objective representationalist may hold that other phenomenal properties are also objective properties of outer stuff.

Many kinds of experiences are allegedly problematic for the transparency argument, including (to select only a few examples) experiences associated with blurry vision or double vision, experiences involving size and shape constancy, and experiences associated with bodily sensations.<sup>34</sup> I set aside such objections for four reasons.

First, I am tentatively sympathetic to the idea that such examples do not really pose problems for the transparency thesis. Second, this chapter is largely aimed at those sympathetic to objective representationalism. Since many such sympathizers accept a suitable transparency thesis, my argument will be dialectically effective for them.

Third, in the last chapter, I defended a separate argument that supports the same conclusion as the transparency argument – the conclusion that, if experience is veridical,

---

<sup>34</sup> Among anti-transparency theorists, Boghossian and Velleman (1989) and Smith (2008) focus on blurry-vision and double-vision cases; Peacocke (1983) and Siewert (2004) emphasize size- and shape-constancy phenomena; and Block (1996) discusses bodily sensations.

phenomenal properties are properties of outer stuff. But this argument, I claim, is not vulnerable to the kinds of objections that have been raised for transparency theses like those defended by Harman and Tye.

My fourth comment is for those unsympathetic to the conclusion of the transparency argument. One can easily develop a version of subjective representationalism according to which many or all phenomenal properties represented by experience are properties of inner stuff. So even if transparency enthusiasts are getting the phenomenology wrong, subjective representationalism more broadly wouldn't be in trouble, though the specific subjective representationalist proposal I've sketched here would then lose an important source of support.

Setting aside worries about whether experience is transparent, then, here is the payoff. At best, transparency considerations support only a conclusion about *where the stuff that instantiates (e.g.) phenomenal colors is located*: it's located *outside* the subject.<sup>35</sup> These transparency considerations do not support any conclusion about exactly which properties of outer stuff are relevant.

To be sure, this conclusion does cut against lots of views about experience. If Harman and Tye have correctly articulated the phenomenological point, then this cuts against views according to which we are directly aware of intrinsic properties of our experience,<sup>36</sup> including representationalist views that hold that phenomenal properties are instantiated by our experiences or by other states in our heads.

But it does not cut against all forms of subjective representationalism. In particular, it does not cut against the form of subjective representationalism sketched in the previous section, according to which at least some phenomenal properties have the

---

<sup>35</sup> At least when the experience is veridical, and when the subject isn't viewing her insides.

<sup>36</sup> Like the "mental paint" view suggested by Block (1996).



form *appropriately causing mental state M in me*. These properties are not instantiated by subjects' experiences or minds.

On this subjective representationalist view, phenomenal properties like phenomenal color may certainly be instantiated by outer stuff. When I successfully visually perceive a red mango, the mango – a paradigmatically outer entity – instantiates the property of appropriately causing a certain mental state in me. Indeed, the subjective representationalist who wishes to capture strong transparency theses of the sort articulated by Harman and Tye may say that experience represents *all* phenomenal properties as instantiated in outer stuff, never in inner stuff.

In short, if Harman and Tye have properly articulated the transparency phenomenon, then any view according to which phenomenal properties are properties of outer stuff is equally well-placed to explain this phenomenon. The version of subjective representationalism I've developed falls into this category.

Still, one might worry that representationalist views that invoke *de se* content face special worries. In the next section, I consider one such worry.

#### 4. Determinacy and centered content

Color experience seems to present objects as *being a certain specific way*. When I see a red strawberry and experience phenomenal redness, my experience presents the strawberry as being a highly determinate way – phenomenally red. Call this the *determinacy phenomenon*. (The determinacy phenomenon might or might not be related to the transparency phenomenon.)

Thompson (2007, fn. 28) worries that a content involving merely a centered property cannot concern any specific way things are, and thus can't accommodate the

determinacy phenomenon. In this section, I'll try to develop the worry (since Thompson mentions it only in a footnote) and then relieve it.

Before I elaborate on the worry, I should separate it from a different worry. One might worry that paradigmatic phenomenal properties seem to be intrinsic properties of objects rather than (as subjective representationalism has it) relational properties of objects. This worry applies to any form of subjective representationalism, whether or not the view posits *de se* contents. I'm not dealing with this worry here, for others have replied to it persuasively.<sup>37</sup> The worry I'm engaging applies specifically to views that individuate experiences via *de se* content (content that invokes centered possible worlds).

An example might help bring out the worry. If I believe that Chicago is west of New York, there is a specific way I take things to be: I take things to be such that Chicago is west of New York. But if I believe that Chicago is west of *here*, it seems that there is not a specific way I take things to be, for I needn't have any view about where I am. In the latter case, one might think, my belief places some constraints on how I take things to be, but it doesn't concern a specific way that I take things to be. The worry, stated more generally, is that *de se* content cannot specify how things are; it can only place constraints on specifications of how things are.

To show that this worry is misguided, let's look more carefully at propositional attitudes whose contents do not essentially involve centered properties – say, my belief that Chicago is west of New York. This belief plainly concerns how Chicago is specifically. Why is that? What does this specificity amount to?

Well, on one approach, the content of my belief that Chicago is west of New York corresponds to a set of possible worlds. If my belief is correct, then I can't be in

---

<sup>37</sup> See Shoemaker (1994a, p. 28).

certain possible worlds: I can't be in those possible worlds in which Chicago isn't west of New York. It would be natural for advocates of this approach to say that my belief concerns how Chicago is specifically because it rules out specific possibilities concerning Chicago.

On another approach, the content of my belief that Chicago is west of New York involves a structured Russellian proposition containing Chicago itself and the ordinary property *being west of New York*.<sup>38</sup> Advocates of this view also have a natural account of why my belief concerns specifically how Chicago is: it attributes the specific ordinary property *being west of New York* to Chicago.

But *de se* content is deeply analogous to ordinary content. Consider again my belief that Chicago is west of *here*. We can model the *de se* content of my belief with a set of centered possible worlds – possible worlds with a marked individual and time. If my belief is correct, then I can't be in certain marked locations within possible worlds: I can't be in any location in a possible world in which Chicago isn't west of that location in that world. In other words, when I believe that Chicago is west of here, I *do* take things to be a specific way. I take myself to be in a certain type of location within a possible world: one in which Chicago is to the west of *me now*.

Alternatively, we can model the *de se* content of this belief with a structured Russellian proposition containing Chicago itself and the centered property *being west of here*. While this centered property doesn't determine an extension given a possible world alone, it does determine an extension given a centered possible world. So, as

---

<sup>38</sup> Actually, a more natural thing to say is that the proposition contains Chicago itself, New York itself, and the relation *being to the west of*. But I am not primarily concerned with the right way of thinking about belief contents. I develop the view in the text only to draw certain analogies with the content of experience. That's also why I don't talk about Fregean views of content here. Since I'm not proposing a Fregean account of the content of experience, the parallels between a Fregean account of ordinary content and a Fregean account of centered content aren't relevant here.

before, it's natural to say that my belief concerns specifically how Chicago is. For my belief attributes the specific centered property *being west of here* to Chicago.

Now for the payoff: the same point applies to *de se* contents that involve properties like *appropriately causing  $M_{\text{red}}$  in me*. Suppose my experience of phenomenal redness involves such a *de se* content. This content cannot be understood just in terms of which worlds it rules out; nor can it be understood just in terms of which properties it attributes. Nevertheless, it can be understood just in terms of which centered worlds it rules out; alternatively, it can be understood just in terms of which centered properties it attributes.

Understood either way, the content of such an experience does concern how things are specifically. In our example, my experience presents something as appropriately causing  $M_{\text{red}}$  in me. This centered content concerns a specific feature of the object represented in a way tightly analogous to the way that the content featuring the property *appropriately causing  $M_{\text{red}}$  in  $S$  at time  $t$*  concerns a specific feature. Thus, there is no special worry here for subjective representationalism.

Put generally, my point is this. Consider a belief about an object, where this belief has ordinary content featuring only ordinary properties. It should be uncontroversial that such a belief can concern how, specifically, that object is. But *de se* content is theoretically very much like ordinary content; in this section and in section 1, I've meticulously documented the deep parallels between these two approaches to modeling content. Any reasons for thinking that ordinary content can concern the

specific features of objects are equally reasons for thinking that *de se* content can concern the specific features of objects.<sup>39</sup>

Pulling together the results from the previous section and this one: objective and subjective representationalism (of the kind outlined above) explain the transparency and determinacy phenomena equally well.

## 5. Conclusion

The case for subjective representationalism is beginning to come together. I argued in the previous chapter for experiential externalism, and subjective representationalism is compatible with that view. Moreover, at the beginning of this chapter I mentioned many appealing features of representationalism: it can explain illusions and hallucinations, it promises a physicalist reduction of experience, and it meshes very well with our best cognitive science theories of perception.

In this chapter, I've sketched a version of subjective representationalism that has a substantial advantage over objective representationalism: it can accommodate our deeply held intuitions about the possibility of illusion-free phenomenal inversions. Moreover, I've argued that the transparency phenomenon, arguably the single most sociologically influential motivation for objective representationalism, is just as powerful a motivation for this version of subjective representationalism. Finally, I rebutted an objection based on the determinacy of experience.

This concludes my initial case for subjective representationalism. In the remaining chapters, I will shift to the defensive. I will rebut a number of prominent

---

<sup>39</sup> Following up on my earlier remarks: other ways of analyzing indexicality (e.g., Kaplanian or Fregean ways) cannot, I think, capture the determinacy of experience, but the *de se* analysis can. This point deserves further discussion, which I hope to provide on another occasion.

objections to the version of subjective representationalism I've sketched here and suggest that the resilience of subjective representationalism in the face of these objections is further evidence for the theory.

# Chapter 3.

## The real moral of the ‘explanatory gap’

I mentioned in the previous chapter that representationalism is partly motivated by the fact that it offers the prospect of a purely physical, reductive account of phenomenal experience. Many philosophers think that this is impossible, and a variety of arguments have been adduced to support this view. This chapter responds to one of the most influential and compelling arguments of this sort, the *explanatory gap argument*.

Many philosophers have thought that there is an “explanatory gap”<sup>1</sup> between the phenomenal and the physical. For no matter how much physical information one receives, it seems sensible to ask, “But how could [e.g.] phenomenal redness merely be such-and-such physical property?” This explanatory gap is often taken to show that phenomenal experience cannot be reductively explained in physical terms, so physicalism<sup>2</sup> is false. I will construct a compelling version of this argument, the *Argument from Strong Inferences*, which is a sharpened version of an argument suggested by David Chalmers and Frank Jackson (henceforth, CJ).<sup>3</sup>

---

<sup>1</sup> See Levine (1983) and (2001).

<sup>2</sup> Very roughly, physicalism is the view that everything concrete is reducible to the physical. Nothing in my discussion should hinge on how to characterize physicalism more precisely.

<sup>3</sup> In their (2001).

I will show that this argument does not support any anti-physicalist conclusion. Rather, the explanatory gap data reveal a distinction that I suspect is of paramount importance in semantics, metaphysics, and epistemology: the distinction between basic and non-basic concepts. As I make this case, I will also sketch an account of a priori conceptual knowledge that is compatible with naturalistic accounts of intentionality.<sup>4</sup>

## 1. The Argument from Strong Inferences

The Argument from Strong Inferences is motivated by the thought that there is an important epistemological difference between phenomenal truths (like “I am now experiencing phenomenal redness”) and ordinary macroscopic truths (like “water is wet”). It seems that we can infer such truths about water from other truths containing no explicit mention of water. Indeed, it seems that such inferences are both *conclusive* and *a priori* (call such inferences *strong inferences*). By contrast, we cannot strongly infer truths about phenomenal redness from other truths containing no explicit mention of phenomenal redness.

To see how we are supposed to be able to strongly infer macroscopic truths from truths couched wholly in non-macroscopic terms, consider five statements. Let *P* be a truth containing all the fundamental physical information – plausibly, information about

---

<sup>4</sup> After independently developing the main ideas of this chapter, I came across Joseph Levine’s “The Q Factor: Modal Rationalism vs. Modal Autonomism” (2010), which has many points of contact with this chapter. It is worth drawing out four substantial differences between this chapter and Levine’s paper: (1) Levine’s central point is that, on a certain natural reading, CJ’s argument in their (2001) is circular. I agree, but think that (as Levine also notices) there is another reconstruction of CJ’s argument that is not circular. I focus on this version of the argument, whereas Levine discusses it only briefly. (2) Levine and I agree that CJ’s argument, properly understood, supports a certain semantic conclusion rather than the metaphysical one that CJ actually draw. I develop an argument for this understanding of CJ’s argument that Levine does not consider (see my third reason for preferring the Semantic Explanation to Global Inferentialism in section 5). (3) A central contribution of this chapter (in sections 3-4) is to develop a detailed physicalist theory that accommodates the semantic conclusion of CJ’s argument. Levine does not pursue this project. (4) Explaining how the physicalist accommodates CJ’s semantic conclusion also leads me (again in sections 3-4) to develop a physicalist account of a priori conceptual knowledge. Again, Levine is not concerned with this matter.



the fundamental entities and properties of physics, stated in the language of completed physics. By stipulation, fundamental mental properties and entities, if such there be, are excluded from  $P$ . Let  $Q$  be the conjunction of all phenomenal truths, including truths about what phenomenal properties and states are instantiated at all times. Let  $I$  be an indexical truth which contains information like “I am entity  $X$ ” and “It is now time  $Y$ ,” where “ $X$ ” and “ $Y$ ” contain descriptions solely in terms of fundamental physical features of the sort discussed in  $P$ . Let  $T_D$  be a statement which says that the world is a *minimal* world satisfying  $D$ . Here, a minimal world satisfying  $D$  is a world in which  $D$  is true and nothing else is true except whatever must be true for  $D$  to be true. Finally, let  $M$  be the conjunction of all truths about ordinary macroscopic entities and properties.  $M$  will include truths like *water is  $H_2O$* , *the earth is smaller than the sun*, and *I am now north of the equator*. By stipulation,  $M$  excludes fancy metaphysical claims, normative claims, and mathematical claims.

With these definitions in hand, we can now understand the following pivotal claim:

*PQIT-Inferentialism*: There is a strong inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$ .

CJ defend this claim vigorously. Consider, for example, the truth that water is  $H_2O$ , which is one of the many truths in  $M$ .  $P$  would likely contain or entail information about the distribution and behavior of  $H_2O$ ; it would entail that  $H_2O$  is located in certain regions and behaves dynamically in certain ways. Combining this information with  $Q$ , which contains all the phenomenal information, we could further conclude that  $H_2O$  looks, tastes, and feels like water. The indexical truth  $I$  would rule out the possibility

that, say, XYZ was water rather than H<sub>2</sub>O. For this truth would let me figure out that the stuff that causes *me* to have experiences as of water, and which is distributed a certain way in *my* environment, is H<sub>2</sub>O, not XYZ. Finally, T<sub>P&Q&I</sub> would rule out the possibility that there is, in addition to H<sub>2</sub>O, some other physical stuff that causes me to have experiences as of water, that fills the rivers and the lakes, etc. It guarantees that H<sub>2</sub>O is the *only* entity playing this role. It seems that one can strongly infer from (P & Q & I & T<sub>P&Q&I</sub>) to the truth that water is H<sub>2</sub>O, since (P & Q & I & T<sub>P&Q&I</sub>) permits one to conclusively rule out any hypothesis according to which water is not H<sub>2</sub>O.<sup>5</sup> CJ think that similar considerations show that there are strong inferences from (P & Q & I & T<sub>P&Q&I</sub>) to any macroscopic truth at all.

If PQIT-Inferentialism is true, then the following claim looks appealing:

*Global Inferentialism:* For any global truth E, there is a strong inference from (E & I & T<sub>E&I</sub>) to M.

Here, a truth is *global* just in case it includes all the information about the concrete metaphysically fundamental entities and properties in the actual world. (A truth that includes more information than that will still count as global.) So if E is a global truth about the actual world, then any minimal world satisfying E will be just like the actual world.

Global Inferentialism is an appealing explanation of PQIT-Inferentialism. It's quite striking that there is an inference (if there is one) from the physical and

---

<sup>5</sup> What about someone who lacks some of the concepts required to entertain the claim that if (P & Q & I & T<sub>P&Q&I</sub>) is true, then M is true? Such a person is no counterexample to PQIT-Inferentialism. For when considering whether or not some claim is knowable a priori, it is standard to restrict attention to subjects who possess all the concepts required to understand that claim.

phenomenal truths (plus the minor additions of the totality and indexical truths) to M, an enormous conjunction of macroscopic truths; it's very surprising indeed that such an inference is both conclusive and a priori! Global Inferentialism explains why this strong inference would be available. According to Global Inferentialism, PQIT-Inferentialism is just a special case of a more general phenomenon. Quite generally, Global Inferentialism says, there is a strong inference from any global truth E, conjoined with I and  $T_{E\&I}$ , to M. Since it's quite reasonable to think that the complete truths about the fundamental physical and phenomenal aspects of the world together comprise one such global truth about the world,<sup>6</sup> we get a strong inference from  $(P \& Q \& I \& T_{P\&Q\&I})$  to M. One might further suggest that one has a full reductive explanation of some phenomenon just in case one can produce at will any piece of the strong inference from  $(E \& I \& T_{E\&I})$  to that phenomenon.

Let's consider the bearing of Global Inferentialism on physicalism. Physicalists think that P, which contains the complete truth about the fundamental physical aspects of the world, is a global truth. So Global Inferentialism and physicalism together entail this claim:

*PIT-Inferentialism:* There is a strong inference from  $(P \& I \& T_{P\&I})$  to M.

Is PIT-Inferentialism true? Consider again the truth "water is H<sub>2</sub>O," which is one of the many conjuncts of M. It seems clear that the phenomenal truths in Q would be necessary to strongly infer this truth, for some of our most fundamental beliefs about water are beliefs about how water phenomenally looks, tastes, etc. So the issue seems to

---

<sup>6</sup> Notice that this is compatible with physicalism. If P alone is a global truth, then a fortiori so is  $(P \& Q)$ .

be whether or not there is a strong inference from  $(P \ \& \ I \ \& \ T_{P\&I})$  to  $Q$ . If there is, then given  $(P \ \& \ I \ \& \ T_{P\&I})$ , one can strongly infer to  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  and then (if CJ are correct) strongly infer as before to any macroscopic truth  $M$ . If not, then a strong inference from  $(P \ \& \ I \ \& \ T_{P\&I})$  to  $M$  will not be available.

Chalmers, but not Jackson, further argues that we cannot infer from  $(P \ \& \ I \ \& \ T_{P\&I})$  to certain phenomenal truths – say, the truth that I am now experiencing phenomenal redness (let this truth be  $Q_r$ ). His argument runs thus.<sup>7</sup> We can conceive, in arbitrarily much detail, of a world in which  $(P \ \& \ I \ \& \ T_{P\&I})$  is true and  $Q_r$  is false. Imagine a *Zombie World* that is just like this one in all respects captured by microphysics, but where there is no phenomenality at all. In this scenario, someone microphysically just like me exists, but he isn't now having any phenomenal experience at all. But, since this world is just like the actual world in all fundamental physical respects,  $(P \ \& \ I \ \& \ T_{P\&I})$  holds (we may suppose that "I" picks out the appropriate time and individual).<sup>8</sup> If Chalmers is correct that  $(P \ \& \ I \ \& \ T_{P\&I})$  does not permit a strong inference to  $Q$ , then PIT-Inferentialism is false.

We've arrived at the following anti-physicalist argument<sup>9</sup>, which I dub the *Argument from Strong Inferences*:

(1) PQIT-Inferentialism: There is a strong inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$ .  
(Supported by examples like "water is  $H_2O$ .")

---

<sup>7</sup> See his (1996) and (2009).

<sup>8</sup> Chalmers also considers some other arguments for the same conclusion, but these arguments are closely related to the *Zombie World* argument.

<sup>9</sup> I think that this is the most charitable reconstruction of the arguments presented in CJ's (2001), and I offer textual evidence for this attribution in Section 6. However, my primary concern is with the argument itself, whether or not Chalmers or Jackson endorse it.

(2) Global Inferentialism: For any global truth E, there is a strong inference from (E & I & T<sub>E&I</sub>) to M. (Best explanation of PQIT-Inferentialism.)

(3) According to physicalism, P is a global truth. (Definition)

(4) Therefore, if physicalism is true, then so is PIT-Inferentialism: there is a strong inference from (P & I & T<sub>P&I</sub>) to any macroscopic truth M. (Follows from (2) and (3))

(5) PIT-Inferentialism is false. (Supported by the Zombie World argument)

(Conclusion) Therefore, physicalism is false. (Follows from (4) and (5))

The Argument from Strong Inferences converts the intuitive phenomenal-physical explanatory gap into a formidable argument against physicalism. How are we beleaguered physicalists to respond?

Some physicalists avoid the Argument from Strong Inferences by denying PQIT-Inferentialism. If PQIT-Inferentialism is false, then Global Inferentialism is unsupported, and the Argument from Strong Inferences is correspondingly undermined; Block and Stalnaker (1999), Byrne (1999), and Diaz-Leon (forthcoming), among others, explore this approach. Other physicalists think that PIT-Inferentialism is true and thereby duck the Argument from Strong Inferences. Lewis (1966), for example, argues that we can quite straightforwardly strongly infer from the physical truths to the phenomenal truths.

I want to bracket those kinds of moves for present purposes. I think that PQIT-Inferentialism may be true and PIT-Inferentialism may be false for just the reasons given above. Anyhow, I'll assume so for the duration of this chapter. Even with these generous concessions, I think that the Argument from Strong Inferences should be resisted. PQIT-Inferentialism leads us, not to an anti-physicalist conclusion, but to an important semantic result.

Here is how I will proceed. In §2, I will examine more carefully what it would take for PQIT-Inferentialism to be true. In §3-4, I'll offer an explanation of PQIT-Inferentialism that is compatible with physicalism; these sections will also include an account of a priori conceptual knowledge that is compatible with physicalism. In §5 I'll argue that this explanation is better than the explanation provided by Global Inferentialism. §6 contains textual evidence that CJ are committed to Global Inferentialism, a crucial premise in the Argument from Strong Inferences. (Readers uninterested in exegetical issues may skip this section.) I make concluding remarks in §7.

## 2. How could PQIT-Inferentialism be true?

In this section, I want to consider what anyone (physicalist or not) who accepts PQIT-Inferentialism will be committed to. It's astonishing that anyone could perform the inferences described by PQIT-Inferentialism at all, but how on earth can one make such inferences *conclusively and a priori*?<sup>10</sup>

---

<sup>10</sup> There has been much interesting work done recently on the non-conclusive a priori (e.g., on what we might have *some a priori justification* for believing). However, since PQIT-Inferentialism concerns something that we can conclusively know a priori – that if  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  holds, then so does  $M - I$  – I focus here exclusively on conclusive a priori knowledge.

Think about what I would have to be able to do to strongly infer from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$ .  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  describes the universe using fundamental physical, phenomenal, and indexical concepts. It describes the locations (at all times) of all of the fundamental physical particles, forces, fields, etc.; it describes the distribution of all the phenomenal properties and the bearers of those properties; it specifies who one is and what time it is; and, finally, it says that the world is a minimal world satisfying these truths. I can allegedly strongly infer from this description to a description of, say, birds, trees, water, experiences, or what have you. How could I do this?

Well, I would have to possess conclusive a priori knowledge of at least some logical and mathematical truths. For  $P$  will likely contain mostly descriptions of low-level entities like electrons or quarks, but  $M$  will contain mostly statements concerning high-level entities like birds and chairs. I would need to deduce high-level structural properties of entities in the world from  $P$ ,  $Q$ ,  $I$ , and  $T_{P\&Q\&I}$  by performing various logical and mathematical transformations (primarily on  $P$ ).

But logical and mathematical knowledge would not suffice. No matter how many logical or mathematical transformations one performed on  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$ , one wouldn't get a description in terms of macroscopic entities like birds and chairs. So one would additionally have to know some truths about birds, chairs, etc. to move from the physical-phenomenal-indexical description to  $M$ . These truths might be truths like "If  $X$  has a certain shape and size, and is used by person-shaped things to support their weight, and is produced in the right way, and is causally related to me in the right way, and looks enough like entities  $Y$  and  $Z$ , and ..., then  $X$  is a chair."

What sorts of truths about macroscopic entities are these? Not empirical truths, for we're considering what can be known a priori about macroscopic entities. The only

plausible claim is that these truths which we can know conclusively and a priori are semantic. Moreover, this semantic knowledge does not concern the meanings of our ordinary-language terms. Strong inferences from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$  are inferences in thought, conducted with concepts; I can make such inferences even if they involve concepts for which I know no corresponding public-language expression. Perhaps I can even make such inferences without knowing any public language at all. *Conceptual* knowledge, of some sort or other, must be the semantic knowledge we're after; conclusive a priori conceptual knowledge must be available if there is a strong inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$ .

So anyone committed to PQIT-Inferentialism, whether physicalist or anti-physicalist, must accept two further claims:

*Analyticity:* There are conceptual truths which are required for the inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$ .

*Conceptual Apriorism:* We are in a position to know those conceptual truths conclusively and a priori.

Do these claims pose problems for physicalism? I will argue that they do not.

### 3. An alternative account of PQIT-Inferentialism, part 1

In this section and the next, I will propose two hypotheses, compatible with physicalism, that together explain the truth of PQIT-Inferentialism via Analyticity and



Conceptual Apriorism. To motivate the first of these hypotheses, I'll make some general remarks about the nature of reference.

In general, if a concept  $C^{11}$  refers to  $C$ 's, then *something* must sustain the reference relation between concept  $C$  and  $C$ 's. It can't just be a brute fact that concept  $C$  bears the reference relation to  $C$ 's.<sup>12</sup> Distinguish between two ways that a reference relation between some concept  $C$  and  $C$ 's might be sustained. The reference relation might be sustained *at least partially independently of concept  $C$ 's relations to other concepts*, in which case I'll say  $C$  is a basic concept. Or the reference relation might be sustained *wholly via concept  $C$ 's relations to other concepts*; then I'll call  $C$  a *non-basic* concept.

How exactly might a basic concept work? Here I want to draw on the dominant physicalist account of concepts, according to which many concepts get their reference via nomological relations to things in the world.<sup>13</sup> On such accounts, why does my concept RED refer to redness? Because red things *normally/ideally cause*, or *cause in a way resulting in asymmetric dependence*, or *have the function of causing*, my application of the concept RED. I'll say that my concept RED *tracks* redness, where "tracking" is a placeholder for the relevant nomological relation. A plausible suggestion is that basic concepts get their reference by tracking their referents. The tracking relation may be sustained via one's disposition to act immediately in certain ways upon deploying the concept, or via the relationship between subconscious modules (e.g., language processing modules or perceptual processing modules) and deployment of the concept.

---

<sup>11</sup> As is common, I use small caps to denote concepts – e.g., DOG is the concept that refers to dogs. Sentences in brackets denote beliefs – [DOGS BARK], for instance, is the belief that dogs bark.

<sup>12</sup> The issue I'm focusing on is what mechanisms currently sustain the reference relation, not what past (e.g., evolutionary) factors might have resulted in the existence of such mechanisms.

<sup>13</sup> Some influential accounts of this sort may be found in Dretske (1995), Fodor (1990), Millikan (1989), and Tye (1995) and (2000).

Next, consider how a non-basic concept might function. Take, for instance, my concept WATER. (For ease of exposition only, I'll assume for now that WATER is a non-basic concept.) I have a vast stock of beliefs about water: that it looks, tastes, and feels a certain way, is the predominant substance in rivers, oceans, and lakes, has an underlying structure that is responsible for its other properties, stands in certain causal relations to me, is called "water," fills the glass in front of me, etc. This stock of water-beliefs is frequently revised and updated.

On perhaps the simplest view, non-basic concepts, just like basic concepts, refer by tracking their referents. Then one might say that my concept WATER refers to water because it tracks water, and it tracks water solely because it figures in my (constantly updated) stock of water-beliefs.

But other views on how non-basic concepts refer are available. For instance, one might say that metaphysical constraints on interpretation (e.g., constraints involving charity) apply to non-basic concepts. On this view, my concept WATER refers to water solely because interpreting my concept WATER as referring to water best meets the constraints on interpretation. But, since these constraints get their bite via my concept WATER's relations to other concepts – because of how WATER figures in my stock of water-beliefs – this is still a view according to which the concept WATER is non-basic. There are many other possible views on how non-basic concepts get their reference from basic concepts.

This distinction between basic and non-basic concepts suggests the following picture. Basic concepts get their reference by hooking up directly to their referents in some way or other – plausibly, via tracking. Non-basic concepts get their reference via their relations to other concepts; those other concepts might in turn be basic or non-

basic. Any of these other non-basic concepts will themselves get their reference via further basic or non-basic concepts. But, ultimately, this process must bottom out in basic concepts – it cannot be non-basic concepts all the way down.<sup>14</sup> Thus, fixing the reference of all of one’s basic concepts will suffice to fix the reference of one’s non-basic concepts, too.

Now for the first of my two hypotheses. Call the concepts we use to think about  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  our *PQIT concepts*, and call the concepts we use to entertain the macroscopic truth  $M$  our *macroscopic concepts*. Then I hypothesize that this holds:

*PQIT-Determination*: Our PQIT concepts determine the reference of all our macroscopic concepts.

Note that PQIT-Determination entails that it is a conceptual truth, in some good sense of that phrase, that if  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  holds, then so does  $M$ . Assuming that the strong inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$  relies on such conceptual truths, PQIT-Determination entails Analyticity.

Let me explain in some detail how PQIT-Determination could be true. Plausibly, there are basic macroscopic concepts (perhaps SPACE and CAUSE) and non-basic macroscopic concepts (perhaps WATER and WET). Similarly, there are basic PQIT concepts (PHENOMENAL REDNESS, for example) and non-basic PQIT concepts (perhaps

---

<sup>14</sup> It *could*, however, be basic concepts all the way down; many philosophers think that all concepts are basic. However, such philosophers invariably deny PQIT-Inferentialism, too. Since we’re examining what it would take for PQIT-Inferentialism to be true, we must (as far as I can tell) grant that there are some non-basic concepts.

It could also be that some sets of concepts mutually constrain one another’s reference. Perhaps WATER gets its reference partially via its relation to WET, and WET also gets its reference partly via its relation to WATER. Nothing I have said is incompatible with this possibility.

PROTON or CHARGE). Moreover, there will be some overlap between our macroscopic concepts and our PQIT concepts – examples might include SPACE and CAUSE.

To entertain M, we will use both basic and non-basic concepts. Now, it's plausible that all basic macroscopic concepts – concepts like SPACE, TIME, CAUSE, PHENOMENAL REDNESS, I, NOW, etc. – are already required to think about (P & Q & I & T<sub>P&Q&I</sub>). So fixing the reference of our PQIT concepts will quite trivially fix the reference of all basic macroscopic concepts, for our PQIT concepts include all basic macroscopic concepts.

Next, I'll argue that all non-basic macroscopic concepts get their reference via the PQIT concepts. This argument is similar to CJ's argument that one can strongly infer from (P & Q & I & T<sub>P&Q&I</sub>) to M. Consider some macroscopic concepts – say, WATER, BIRD, and GOLD. Like most macroscopic concepts, these concepts are intimately associated with the phenomenal information contained in Q: I recognize water, birds, and gold largely by how they look, feel, etc. Of course, such information is insufficient to determine the reference of my concepts WATER, BIRD, and GOLD, for several reasons.

First, I may conceive of water, birds, and gold as having a certain hidden structure. Information about the shape, motion, and behavior of microphysical particles provided by P will pin down this hidden structure. Second, I conceive of water, birds, and gold as being causally related to me in certain ways. The indexical truth I will provide this kind of information.

Third, I pick out some of these entities (most plausibly, birds) partly via their macroscopic properties (like having wings, having a certain ancestral history, or being called “birds” by the experts). So to determine whether or not something is a bird, I may

consider whether or not it has these properties. But, ultimately, non-basic concepts like WINGS, OFFSPRING, and EXPERTS must get their reference from basic concepts. And, as before, the physical, phenomenal, and indexical concepts look like the best candidates to do this. So it is reasonable to think that the PQIT concepts determine the reference of all macroscopic concepts, just as PQIT-Determination says.

PQIT-Determination gets further support from the fact that Analyticity must be true (assuming, as we are, that PQIT-Inferentialism is true). It must be some kind of conceptual truth that if  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  holds, then so does M. It's plausible that this is a conceptual truth because the basic PQIT concepts include the basic macroscopic concepts, and the PQIT concepts together determine the reference of the non-basic macroscopic concepts.

#### 4. An alternative account of PQIT-Inferentialism, part 2

Here's my second key hypothesis for explaining PQIT-Inferentialism:

*Restricted Conceptual Apriorism:* If the reference of the concepts in some set  $S_1$  is wholly determined by the reference of some concepts in set  $S_2$ , then, given information about the reference of the concepts in  $S_2$ , I can determine the reference of the concepts in  $S_1$  conclusively and a priori.

Let me illustrate what Restricted Conceptual Apriorism says. Suppose that the reference of my concept WATER is fixed by my beliefs about water – e.g., the beliefs that water looks and tastes a certain way, is the predominant substance in the rivers and lakes, quenches thirst, etc. Then, according to Restricted Conceptual Apriorism, given

the information that H<sub>2</sub>O looks and tastes that way, is the predominant substance in the rivers and lakes, and quenches thirst, I can determine, conclusively and a priori, that water is H<sub>2</sub>O.

Why think that Restricted Conceptual Apriorism is true? Well, Restricted Conceptual Apriorism would help explain our ability to pick out the referents of our concepts in various epistemic possibilities. For example, I can imagine discovering that I inhabit Twin Earth, where XYZ looks and tastes a certain way, is the predominant substance in the rivers and lakes, quenches thirst, etc. I know that if this possibility actually obtains, then XYZ is water. Restricted Conceptual Apriorism explains how I acquire such knowledge. I take for granted certain claims about the referents of the concepts which determine the reference of my concept WATER, and this puts me in a position to determine the reference of my concept WATER itself. It's also important to note that Restricted Conceptual Apriorism is compatible with many different stories about how I figure this out. Perhaps I perform deductions, or form mental models, or have a rational intuition, or something else entirely.

Moreover, anyone who is generally sympathetic to the idea that we can know conceptual truths a priori should be sympathetic to Restricted Conceptual Apriorism. For Restricted Conceptual Apriorism says that we can know a certain limited set of conceptual truths – conceptual truths about any concept whose reference is wholly determined by the reference of certain other concepts. Someone who endorses the Argument from Strong Inferences is committed to the broad notion that we can know some conceptual truths a priori and shouldn't find Restricted Conceptual Apriorism hard to swallow.

Together, PQIT-Determination and Restricted Conceptual Apriorism straightforwardly explain PQIT-Inferentialism. According to PQIT-Determination, the reference of our macroscopic concepts is determined by the reference of our PQIT concepts. According to Restricted Conceptual Apriorism, one can figure out the reference of a concept conclusively and a priori once one knows the reference of all the concepts which determine its reference. Combining these claims, we arrive at PQIT-Inferentialism, the claim that there is a strong inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to macroscopic truth M.

For this explanation to help the physicalist, both PQIT-Determination and Restricted Conceptual Apriorism must be compatible with physicalism.<sup>15</sup> And both are compatible with physicalism. PQIT-Determination says something about which concepts are basic and which are non-basic; none of this has any obvious anti-physicalist import. Indeed, the only bearing PQIT-Determination has on physicalism is this: PQIT-Determination supports (though by itself does not entail) the idea that phenomenal concepts are basic; they don't have their reference fixed wholly via the reference of other concepts. But physicalists will almost certainly go in for that anyway; physicalists typically say that one's (pre-conceptual) perceptual faculties play a major role in determining the reference of one's phenomenal concepts.

Restricted Conceptual Apriorism says that we have the ability to determine the reference of some of our concepts under certain conditions. Again, physicalism does not conflict with this claim in any obvious way. Physicalists need not oppose the idea of a priori conceptual knowledge. Still, many philosophers have strong scruples concerning conceptual truths, so let me briefly consider the most prominent of these scruples.

---

<sup>15</sup> They are also each compatible with dualism. Indeed, I suspect that Chalmers would be sympathetic to each, though he might not like the further physicalist account that I used to support PQIT-Determination.

First scruple: one might think that *meaning holism* is true, i.e., that for almost any concept C, the meaning of C depends partly on its relations to almost every other concept one possesses. On one version of such a view, my concept BIRD gets its meaning via all of my inferential dispositions vis-à-vis that concept. Since I associate birds with having wings, laying eggs, eating worms, concepts like WING, EGG, and WORM will help determine the meaning of my concept BIRD. But my concepts WING, EGG, and WORM will themselves get their meanings via their relations to still other concepts, which may in turn get their meanings via their relations to other concepts again. So the reference of my concept BIRD will depend on the meanings of a very large number of my concepts. Someone drawn to this view might worry that there just aren't any conceptual truths involving my concept BIRD.

This worry doesn't follow from the idea that meaning holism is true. There may not then be any *short* conceptual truths about birds, but there may still be quite long conceptual truths about birds. As I suggested above, these long conceptual truths may issue from the fact that all of our macroscopic concepts ultimately derive their reference from our PQIT concepts. This suggestion is perfectly compatible with the thought that these macroscopic concepts penultimately get their reference via their relations to one another. And, of course, my account is also compatible with meaning anti-holism.

Second scruple: There are compelling philosophical reasons for doubting that most of our concepts have definitions. One such reason is that there are very few remotely plausible examples of such definitions, even of concepts (like KNOWLEDGE) which philosophers have attempted to analyze for decades. There are also compelling scientific reasons for doubting that most of our concepts have definitions. Here's one: if the concept BACHELOR is defined in terms of the concepts UNMARRIED and MAN, then



one would expect subjects to take longer to evaluate claims about bachelors than to evaluate claims about men. For, it seems, evaluating claims about bachelors should require deploying one's concepts UNMARRIED and MAN, which ought to take longer than deploying just one's concept MAN. However, such differences in processing speed haven't turned up.<sup>16</sup> But, the scruple goes, one must have definitions to have conceptual truths.

These kinds of objections carry a great deal of force against some ways of thinking about what definitions of concepts would look like. But I think that we can modify our understanding of definitions so that these objections do not apply. For example, consider again the observation that there are very few remotely plausible examples of definitions. If one requires a concept to be defined by a relatively small set of concepts – as one might hope that BACHELOR is defined in terms of UNMARRIED and MAN – then the inability of philosophers to find any good candidate definitions is very worrying. But, as I emphasized a few paragraphs back, macroscopic concepts may typically get their reference determined holistically, via the reference of a very large set of other macroscopic concepts. Thus, it may be that few or no macroscopic concepts have short definitions; they may have only very long definitions. Such definitions might even be so long that they could not easily be stated.<sup>17</sup> If this were the case, then it would be unsurprising if philosophers had so far discovered at best very few definitions.

Now turn to the scientific reason for skepticism about definitions: that the differences in processing speed that one might have predicted – e.g., taking a longer time to deploy one's concept BACHELOR than one's concept MAN – have not been found.

---

<sup>16</sup> Fodor (1998) presses these points vigorously.

<sup>17</sup> One might take it to be a requirement on definitions that they be relatively short. If one prefers to talk this way, then I can instead say that my view is not committed to there being definitions at all, since the conceptual truths invoked by my account are quite often very long.

It is certainly the case that, on some views of definitions, one would expect such differences in processing speed to turn up. For instance, on one view, whenever a concept A is defined in terms of concepts B and C, deploying concept A requires deploying concepts B and C. Someone who held this view (and hypothesized that BACHELOR is defined in terms of UNMARRIED and MAN) might reasonably predict that thinking about bachelors would take longer than thinking about men.

But we need not hold such a view about definitions. The assumption that deploying a concept requires deploying all of the concepts that define it may be dropped. And once we drop this assumption, we need not be committed to any predictions about differences in processing speed between (e.g.) BACHELOR and MAN.

To be sure, definitions had better have *some* kind of psychological reality. Even if facts about definitions don't turn up in processing speed, they had better turn up somewhere. Perhaps, for example, facts about definitions turn up in our ability to strongly infer from (P & Q & I & T<sub>P&Q&I</sub>) to truths like "water is H<sub>2</sub>O"; perhaps we can do this partly in virtue because of our knowledge of definitions. The crucial point for present purposes is that theories that posit definitions may differ widely in exactly what kind of psychological phenomena are explained by definitions.

That is all I will say in response to these scruples about conceptual truth. My discussion of these scruples has been exceedingly brief; I have barely scratched the surface of these topics. Still, I hope that my short remarks have shown that these scruples do not immediately refute the view that there are conceptual truths. So let us continue to explore the hypotheses that PQIT-Determination and Restricted Conceptual Apriorism are true.

In the previous two sections, I have sketched an independently plausible explanation of PQIT-Inferentialism, which I'll call the *Semantic Explanation*. The Semantic Explanation consists of two hypotheses, PQIT-Determination and Restricted Conceptual Apriorism, which are compatible with (but don't entail) physicalism.

Recall that PQIT-Inferentialism is the thesis that there is a strong inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$ , where  $M$  is the conjunction of all ordinary macroscopic truths. The Semantic Explanation explains PQIT Inferentialism via two key ideas. The first idea (embodied in PQIT-Determination) is that fixing the reference of the concepts that we use to entertain  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  suffices to fix the reference of all the concepts in  $M$ ; put another way, it is a conceptual truth that, if  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  holds, then  $M$  holds. The second idea (embodied in Restricted Conceptual Apriorism) is that we are in a position to know such conceptual truths conclusively and a priori. These ideas together entail that there is a conclusive and a priori inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to  $M$ , just as PQIT-Inferentialism says.

## 5. Evaluating the explanations

The Semantic Explanation purports to explain PQIT-Inferentialism; so does Global Inferentialism (which, recall, says that for any global truth  $E$ , there is a strong inference from  $(E \ \& \ I \ \& \ T_{E\&I})$  to  $M$ ). The Semantic Explanation, as the name suggests, explains PQIT-Inferentialism in terms of the semantic distinction between basic and non-basic concepts. Global Inferentialism explains PQIT-Inferentialism metaphysically, by positing strong inferences wherever there are global truths. What is the relationship between these two explanations?

The Global Inferentialist must accept either the Semantic Explanation or something very much like it. For, as I argued in section 2, *anyone* who accepts PQIT-Inferentialism will have to accept both Analyticity and Conceptual Apriorism. The Semantic Explanation is the conjunction of PQIT-Determination and Restricted Conceptual Apriorism, which are more specific versions of Analyticity and Conceptual Apriorism. Even if the Global Inferentialist does not like these specific formulations, she will have to settle for something in the vicinity.

But one can accept the Semantic Explanation while rejecting Global Inferentialism. On this view, PQIT-Inferentialism is true because our PQIT-concepts together determine the reference of our macroscopic concepts. But this semantic point is separate from metaphysics; it indicates nothing about the availability of strong inferences from global truths more generally.

I now offer three reasons for preferring the Semantic Explanation alone to Global Inferentialism (which will include the Semantic Explanation or something in the vicinity).

First, Global Inferentialism will push one towards dualism, as the Argument from Strong Inferences shows. This is a *prima facie* cost. (Even dualists should admit that dualism is a *prima facie* worse theory than physicalism, as it is more ontologically inflationary. They merely think that dualism compensates for this lack of parsimony by better handling certain data about phenomenal experience.)

Second, and more tellingly, the extra commitments of Global Inferentialism are explanatorily idle. The Global Inferentialist and the Semantic Explanation theorist both accept certain semantic theses, but the Global Inferentialist further draws a connection between semantics and metaphysics: she says that for any *global truth* E, there is a

strong inference from (E & I & T<sub>E&I</sub>) to M. (The “global truth” bit is where Global Inferentialism makes a metaphysical commitment.) The crucial point, though, is that the putative semantics-metaphysics connection plays no additional explanatory role (at least in explaining PQIT-Inferentialism). The Semantic Explanation alone does all the heavy lifting; as long as it is correct, PQIT-Inferentialism will hold, *regardless* of whether there is any semantics-metaphysics connection.

But might not Global Inferentialism be the best explanation of the Semantic Explanation itself? No. The Semantic Explanation consists of two claims, PQIT-Determination and Restricted Conceptual Apriorism. PQIT-Determination, the thesis that our PQIT concepts determine the reference of all our macroscopic concepts, is unsurprising. So long as there is a distinction between basic concepts and non-basic concepts, there will be *some* proper subset of our concepts (the basic ones) which fix the reference of all of our concepts. Our PQIT concepts just happen to be among the concepts which fix the reference of our macroscopic concepts. An evolutionary story may well be available for why these particular concepts are basic: perhaps it is that many of them developed first. Sensory concepts, the self concept, and temporal, spatial, and causal concepts were likely among the first concepts (or proto-concepts) that our ancestors possessed.<sup>18</sup>

Meanwhile, Restricted Conceptual Apriorism, which posits a certain sort of a priori conceptual knowledge, wouldn't be explained at all by Global Inferentialism. Global Inferentialism just posits strong inferences, for any total truth E, from (E & I & T<sub>E&I</sub>) to M. The Global Inferentialist would need to *assume* that something like

---

<sup>18</sup> Perhaps the relevant issue is not that such concepts developed *first*, but that they are *innate*. The details of this evolutionary speculation aren't especially important. The key point is that deflationary explanations of PQIT-Determination are readily available.

Restricted Conceptual Apriorism was correct to explain how there could be such strong inferences. So Global Inferentialism is not an especially promising explanation for why either component of the Semantic Explanation is correct.

Here is the third reason for preferring the Semantic Explanation alone to Global Inferentialism. Consider this claim, which is very similar to PQIT-Inferentialism:

*PQT-Inferentialism:* There is a strong inference from  $(P \ \& \ Q \ \& \ T_{P\&Q})$  to  $M$ .

PQT-Inferentialism is demonstrably false. For, as is well known,<sup>19</sup> there is no way to strongly infer indexical truths (like “I am here”) from truths couched in non-indexical terms. Moreover, it is unlikely that there is a strong inference from  $(P \ \& \ Q \ \& \ T_{P\&Q})$  even to non-indexical truths like “water is  $H_2O$ .” For I conceive of water as the stuff of *my* acquaintance which has certain features, and without indexical information about who I am, I cannot strongly infer which stuff is the right stuff of *my* acquaintance.

The Global Inferentialist has great difficulty explaining why PQT-Inferentialism is false while PQIT-Inferentialism is true. Put crudely, the picture that the Global Inferentialist would like to endorse is that any global truth permits a strong inference to any truth whatsoever. Moreover,  $(P \ \& \ Q \ \& \ T_{P\&Q})$  is a global truth. But there is not a strong inference from  $(P \ \& \ Q \ \& \ T_{P\&Q})$  alone to  $M$ ; indexical truth  $I$  is required as well. The Global Inferentialist has no obvious explanation of why *this* truth in particular about non-fundamental matters should be necessary for strongly inferring

---

<sup>19</sup> See Lewis (1979) and Perry (1979).

M. She endorses PQIT-Inferentialism rather than PQT-Inferentialism because the latter is extensionally inadequate, but she has no further rationale for this shift.<sup>20</sup>

By contrast, the pure Semantic Explanation theorist can easily explain why PQT-Inferentialism is false but PQIT-Inferentialism is true. The key claim is that we possess some basic indexical concepts. Because these concepts (like phenomenal concepts) are basic, we need not be able to strongly infer from  $(P \ \& \ Q \ \& \ T_{P\&Q})$  to claims featuring such indexical concepts. The fact that  $(P \ \& \ Q \ \& \ T_{P\&Q})$  is a global truth is irrelevant.<sup>21</sup>

If indexical concepts like I and NOW are basic, how might they get their reference? Perhaps at least partly via their tendency to trigger action directly. If I believe that the pants of the person in the mirror are on fire, I may not do anything; if I believe that *my* pants are on fire, I will immediately try to extinguish the flames. Similarly, if I believe that the meeting which I wish to attend is at noon, I need not do anything, but if I believe that the meeting which I wish to attend is *now*, I will head to the meeting room.

I conclude that the pure Semantic Explanation, which is compatible with physicalism, is preferable to the Global Inferentialist explanation, which puts pressure on physicalism.

To summarize the central argument of this chapter: *if* PQIT-Inferentialism is true, as I have granted CJ in this chapter, then there is a powerful argument for two

---

<sup>20</sup> Indeed, all that CJ can say in defense of the shift from PQT-Inferentialism to PQIT-Inferentialism is that the addition of I is “sufficiently minor” and “[does] not change much” (351).

<sup>21</sup> I suspect that the Semantic Explanation theorist is also better positioned than the Global Inferentialist to explain why there may be no strong inference from  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  to certain other truths, like normative, psychological, social, and semantic truths. The Semantic Explanation theorist can say that such strong inferences are missing because certain normative, psychological, social, and semantic concepts are basic and are not among our PQIT concepts. It is unclear what the Global Inferentialist should say. However, space does not permit me to fully explore these issues here.

conclusions: first, the conclusion that there is a significant semantic distinction between basic and non-basic concepts; and second, the conclusion that we are in a position to know certain conceptual truths conclusively and a priori. Neither of these conclusions exert any obvious pressure on physicalism. Nor does explaining PQIT-Inferentialism require us to accept any other thesis (like Global Inferentialism) that does exert obvious pressure on physicalism.

## 6. Are CJ committed to Global Inferentialism?

CJ do not explicitly state Global Inferentialism anywhere in their (2001). Still, as I will argue in this section, they are committed to this claim. The first reason for thinking this is that the best anti-physicalist argument I can extract from CJ's paper, the Argument from Strong Inferences, requires Global Inferentialism. I leave an open invitation to CJ (or the reader) to provide a better anti-physicalist argument from the considerations raised in that paper.

More importantly, there is abundant textual evidence that CJ accept Global Inferentialism. CJ say this in the opening lines of their (2001): "If there is no [strong inference] from microphysical truths to phenomenal truths, does reductive explanation of the phenomenal fail? We say yes." Why think this? Presumably because one thinks, quite generally, that something like the following holds:

*Ultrastrong Local Inferentialism:* If truth  $E'$  is fully reductively explainable in terms of truth  $E$ , then there is a strong inference from  $E$  to  $E'$ .



Ultrastrong Local Inferentialism can't be quite what Chalmers accepts (the considerations I raise won't apply to Jackson). For consider the truth "water is liquid at room temperature." This truth is reductively explainable in terms of the physical truths alone – phenomenal truths aren't any part of the reductive explanation of water's liquidity at room temperature. However, there is no strong inference from the physical truths alone to the truth "water is liquid at room temperature." For we pick out water largely via its phenomenal effects on us, and the physical information alone won't inform us of these effects. (Recall that Chalmers thinks that the physical truths alone don't a priori entail the phenomenal truths.)

CJ aren't after a principle about which truths reduce to which other local truths. Rather, they're after a principle about which global truth all other truths are reducible to. So here's a plausible revision of Ultrastrong Local Inferentialism:

*Ultrastrong Global Inferentialism:* There is a strong inference from any global truth E to any truth at all.

But CJ back off this strong claim:

"we have not argued that  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$ <sup>22</sup> implies every truth in every domain (though we are inclined to accept this claim). Given what we have said here, it could be that certain truths in special domains – perhaps concerning mathematics, metaphysics, morality, or mentality? – are not implied by  $P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I}$  ... But we hope we have said enough to make it plausible that ordinary macroscopic truths concerning everyday macroscopic natural phenomena [may be strongly inferred from]  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$ " (2001, 335-6).

This suggests a retreat to this slightly weaker claim:

---

<sup>22</sup> CJ use a slightly different notation in place of " $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$ "; for consistency, I replace their notation with mine even when quoting them.

*Strong Inferentialism:* There is a strong inference from any global truth E to any macroscopic truth M.

And the following passage suggests that CJ would like to accept Strong Inferentialism:

When a concept of some natural phenomenon supports a priori entailments from the microphysical, there is a clear sense in which the phenomenon can be reductively explained. These a priori entailments might not support a *reduction* of the phenomenon in question to a microphysical phenomenon (at least in some senses of this term), perhaps because such entailments are compatible with multiple realizability. But, nevertheless, in showing how any instance of the phenomenon is itself implied by microphysical phenomena, we show that there is a sort of transparent epistemic connection between the microphysical and macroscopic phenomena. Both the microphysical and the macroscopic phenomena are epistemically contingent, in that they involve the actualization of just one of a host of coherent epistemic possibilities. But where this sort of transparent entailment is present, the epistemic contingency in the macroscopic phenomena is reduced to the epistemic contingency in the microphysical phenomena: there is no further epistemic contingency in the connection. (2001, 350-1)

The considerations raised in this passage seem to support something like Strong Inferentialism. Even Strong Inferentialism is too strong, though. For physicalists and anti-physicalists alike can agree that  $(P \ \& \ Q)$  is a global truth (though physicalists also think that P alone is a global truth). However,  $(P \ \& \ Q)$  alone won't support strong inferences to all of macroscopic truths. For, as was demonstrated by the earlier example concerning the truth "water is  $H_2O$ ," many such strong inferences would require knowledge of the indexical truth I and totality truth T, neither of which is strongly inferable from  $(P \ \& \ Q)$ . CJ recognize this problem:

"When a phenomenon is entailed a priori by  $(P \ \& \ Q \ \& \ I \ \& \ T_{P\&Q\&I})$  ... , something similar applies ... T and I are sufficiently minor additions to the reduction base that they do not change much" (2001, 351).

Taking this suggestion into account, we arrive at the thesis which, so far as I can tell, CJ accept without qualification:

*Global Inferentialism.* For any global truth E, there is a strong inference from (E & I & T<sub>E&I</sub>) to any macroscopic truth M.

## 7. Conclusion

I conclude that if PQIT-Inferentialism is true, then there is a powerful argument for the claim that we possess both basic and non-basic concepts. Specifically, hypothesizing such a distinction is a promising explanation of PQIT-Inferentialism. The distinction between these two types of concepts obviously matters for semantics. I've also argued that this distinction matters in metaphysics (by deflecting certain anti-physicalist arguments) and epistemology (by grounding a priori conceptual knowledge). I suspect that the significance of the distinction extends much further, and I hope to explore the topic on another occasion. In any case, there is nothing here that should keep the physicalist up at night.

## Chapter 4.

# How I learned to stop worrying and love phenomenal externalism

This chapter examines another objection to subjective representationalism – an objection that also applies to many forms of objective representationalism. According to the representationalist, phenomenal experiences essentially involve representations of environmental/bodily properties. It is very plausible, however, that that such representation is *wide*, i.e., that exact intrinsic duplicates<sup>1</sup> may differ in which environmental/bodily properties they represent.<sup>2</sup> If this is right, then we arrive at the following view:

*Phenomenal Externalism*: Two individuals who are exact intrinsic duplicates may differ in their phenomenal experiences.

Phenomenal Externalism is extraordinarily counterintuitive; its negation, *Phenomenal Internalism*, is very intuitive. Call our intuition that Phenomenal Internalism is true the *Phenomenal Internalist Intuition*. The Phenomenal Internalist Intuition is

---

<sup>1</sup> It is a familiar and difficult problem to pin down the difference between intrinsic and extrinsic properties. I shan't try to say anything about this problem; intuitively, there is a difference, and our intuitive grasp of this difference suffices for present purposes. Lewis (1983) briefly discusses the issue.

<sup>2</sup> See Putnam (1975) and Burge (1979).

widespread, powerful, and recalcitrant. It is *widespread* in that almost everyone who carefully considers the matter has this intuition. It is *powerful* in that such individuals feel a strong pull, not just a mild inclination, to accept the intuition. And it is *recalcitrant* in that even those committed to phenomenal externalism for theoretical reasons often continue to feel strongly intuitively drawn to Phenomenal Internalism.<sup>3</sup>

Many sophisticated philosophical arguments against Phenomenal Externalism have been developed;<sup>4</sup> I believe that such arguments have been adequately rebutted.<sup>5</sup> Even so, most philosophers accept Phenomenal Internalism, and surely this is at least partly due to its intuitive appeal. In this chapter, I set aside sophisticated philosophical arguments against Phenomenal Externalism. Instead, I aim to undermine the powerful Phenomenal Internalist Intuition. If I succeed, then this relieves considerable pressure on those committed to Phenomenal Externalism – most relevantly for my purposes, the subjective representationalist.

This chapter will proceed in 4 sections. In §1, I will offer a psychological account of why we have the Phenomenal Internalist Intuition. According to this account, we possess a psychological mechanism that produces ineradicable (near-)certainty in specific phenomenal beliefs. Some beliefs of this type give rise to the Phenomenal Internalist Intuition. In §2, I argue that some such ineradicably certain phenomenal beliefs are false. In §3, I suggest that the beliefs which produce the Phenomenal Internalist Intuition are among these false phenomenal beliefs, so the Phenomenal Internalist Intuition is vitiated. Even if we continue to have this intuition, it carries no epistemic weight against Phenomenal Externalism. §4 concludes with some

---

<sup>3</sup> Though phenomenal externalist Christopher Hill (2009, ch. 5) reports that he no longer finds Phenomenal Internalism even intuitively compelling.

<sup>4</sup> For example, Rey (1998) offers a battery of objections to Phenomenal Externalism.

<sup>5</sup> Lycan (2001) surveys arguments for Phenomenal Internalism and provides externalist replies. Dretske (1995b) and Tye (1998) also defend Phenomenal Externalism.

speculations about why a mechanism might evolve that sometimes produces ineradicable certainty in falsehoods.

## 1. The source of the Phenomenal Internalist Intuition

What psychological explanation of the Phenomenal Internalist Intuition might one offer? The philosophical literature contains many arguments for Phenomenal Internalism, such as arguments concerning Inverted Earth, possible swampfolk, and considerations from neuroscience.<sup>6, 7</sup> However, such arguments do not reveal the source of the Phenomenal Internalist Intuition (nor are they intended to); we are attracted to Phenomenal Internalism before considering Inverted Earth or swampfolk scenarios and before learning about the pertinent neuroscientific data.

But here is one natural explanation of the Phenomenal Internalist Intuition (call it the *Correlation Explanation*). As is demonstrated by dreams, hallucinations, etc., I can experience phenomenal redness even if there is nothing red in my environment. Indeed, these cases show that my experience of phenomenal redness does not correlate well with the instantiation of *any* current environmental property. So, one might infer, my experience of phenomenal redness does not constitutively depend on any environmental property. Generalizing the point to all phenomenal properties, we arrive at Phenomenal Internalism.

---

<sup>6</sup> Block (1990) presents the Inverted Earth argument; Tye (1998) discusses (and rebuts) the swampfolk objection; and Egan and John (unpublished manuscript) offer neuroscientific considerations against Phenomenal Externalism.

<sup>7</sup> Moreover, Lycan (2001, p. 28) offers an interesting psychological explanation of the Phenomenal Internalist Intuition that posits a conflation of perceptual experiences with bodily-sensation experiences. While I think that the explanation I offer below is more psychologically plausible, I need not argue with Lycan. If he is right, then the Phenomenal Internalist Intuition has been debunked anyway.

But this is poor reasoning. Even if my experience of phenomenal redness does not correlate well with the instantiation of any *current* environmental property, it may still correlate well with (and constitutively depend on) the instantiation of some *past* environmental property. The analogy with belief is instructive: my belief that there is water in front of me need not correlate well with the instantiation of any current environmental property – I may be prone to hallucinations as of water in front of me. Still, it is plausible that my belief that water (rather than twater) is in front of me constitutively depends on a past causal chain (perhaps mediated by my community or ancestors) linking me to water. Even narrow-content enthusiasts like Chalmers (2006) and Jackson (1998) will concede this point. For they agree that some content is wide; they think additionally that such wide-content properties depend on more explanatorily fundamental narrow-content states of the subject. Similarly, perhaps my experience of phenomenal redness (as opposed to phenomenal blueness) constitutively depends on some past causal chain linking me to red objects in the environment.

Thus, the Correlation Explanation would be a *debunking* explanation of the Phenomenal Internalist Intuition, i.e., an explanation according to which the Phenomenal Internalist Intuition is epistemically inappropriate. For the Correlation Explanation traces the Phenomenal Internalist intuition to an inference that is an egregious overgeneralization. If the correct psychological explanation of the Phenomenal Internalist Intuition is a debunking explanation, then the phenomenal externalist has nothing to fear from this intuition.

In any case, the Correlation Explanation is a mistaken psychological explanation of the Phenomenal Internalist Intuition. For most philosophers at least once found intuitively compelling both Phenomenal Internalism and *Belief Internalism*, the view that

two individuals who are exact intrinsic duplicates cannot differ in their beliefs. But, after exposure to theoretical considerations that favor Belief Externalism, many philosophers are willing to give up Belief Internalism. By contrast, very few philosophers are willing to give up Phenomenal Internalism even after exposure to powerful theoretical considerations for Phenomenal Externalism. This suggests that there is some difference in the psychological mechanisms that give rise to the Belief Internalist Intuition and the Phenomenal Internalist Intuition. But the Correlation Explanation predicts no such difference.

Let's refine the Correlation Explanation. The reasoning featured in the Correlation Explanation failed because of insufficient attention to the way in which *past* environmental conditions might help constitute my current mental state. The problem was that *my experiencing phenomenal redness* might be partially constituted by past environmental conditions, just as *my believing that water is in front of me* is partially constituted by past environmental conditions. Plausibly, however, *phenomenal redness itself* is not a property partially constituted by past environmental conditions. If an entity is phenomenally red, any entity intrinsically identical to it will also be phenomenally red.

The *Hallucination Explanation* (which echoes Ayer's (1963) famous Argument from Illusion) exploits this point. According to the Hallucination Explanation, we have the Phenomenal Internalist Intuition because we accept this claim:

*Hallucination Objectualism:* In certain cases of hallucination (and dreaming, etc.), I am related to a phenomenally red object even when there is no phenomenally red object in my environment.



Suppose Hallucination Objectualism is true. Then that phenomenally red object is not an environmental object. Assuming that the objects of experience are of the same type in all experiences, it follows that the objects of experience quite generally are not environmental objects; they are plausibly mental objects intrinsic to me, and phenomenal properties are plausibly intrinsic properties of such objects. Thus, Phenomenal Internalism is true.

The argument from Hallucination Objectualism to Phenomenal Internalism looks eminently reasonable. But the Hallucination Explanation is incomplete, for it does not explain why we accept Hallucination Objectualism. If it visually appears to me that there is an apple ahead, but I know that I am hallucinating, I will not believe that there is an apple ahead. By contrast, if it visually appears to me that something is phenomenally red, but I know that I am hallucinating, I will still believe that something nearby is phenomenally red. Why the psychological difference?

Fancy arguments may be adduced to support Hallucination Objectualism. But we are seeking a plausible psychological explanation of the Phenomenal Internalist Intuition, so those fancy arguments are irrelevant. As I pointed out in the introduction, acceptance of Phenomenal Internalism is widespread (almost everyone who has considered the matter has the intuition), powerful (the intuition is very compelling), and recalcitrant (even theoretically committed phenomenal externalists often have the intuition). But it is very unlikely that, as a matter of psychological fact, we all happen to accept the same fancy argument for Hallucination Objectualism and therefore accept Phenomenal Internalism. And it is similarly unlikely that such a fancy argument would produce such a powerful and recalcitrant intuition.

The following observation is more helpful. Note that experience is often associated with (or, depending on one's view, constituted by) belief-like mental states. For example, if I attend to an apple before me and experience phenomenal redness, I form a belief-like mental state in which I attribute phenomenal redness to *something* – perhaps to the apple, or to my experience, or to some non-physical entity. This state is *belief*-like in that it has mind-to-world direction of fit; it is *belief-like* in that it may be (some think) “non-conceptual” in some appropriate sense. For brevity, I will refer to these belief-like mental states as *phenomenal beliefs*, but this should be considered mere shorthand.

When I have a conscious phenomenal experience, I form a phenomenal belief of the form [E IS (NOW) Q], where my concept E refers to some entity and my concept Q refers to some phenomenal property. Call beliefs of this form *basic phenomenal beliefs*.

When one forms a basic phenomenal belief, one's confidence in that belief is very high – perhaps approaching certainty. Moreover, this certainty, or near-certainty, cannot be lowered, even in the face of extreme skeptical scenarios. For example, suppose I take seriously the possibility that Descartes' demon is deceiving me. I can lower my confidence that I have hands, or that there is a table in front of me, or that I am a human being. But I still just *can't* lower my confidence in my belief [E IS Q]; this certainty is ineradicable. The same point can be made, *mutatis mutandis*, for any basic phenomenal belief in whatever skeptical scenario you like.

Call a belief *tenacious* just in case one is (nearly) certain of that belief and that certainty is ineradicable. We have arrived at the following thesis:

*Tenacity*: All basic phenomenal beliefs are tenacious.

Perhaps one can *simultaneously* be certain of one's basic phenomenal beliefs while harboring other theoretical beliefs that conflict with them; and certainly basic phenomenal beliefs may be rapidly extinguished and replaced with new basic phenomenal beliefs. But Tenacity is compatible with these claims.

The pieces of the puzzle are now beginning to fall into place, leading us to this expanded version of the Hallucination Explanation, which I'll call the *Tenacity Explanation*. As a matter of psychological fact, we have ineradicable certainty in our basic phenomenal beliefs. Thus, even in the hallucination scenario, I have the tenacious basic phenomenal belief [E IS PHENOMENALLY RED]. Hallucination Objectualism follows from this belief and leads naturally, though perhaps not inevitably, to Phenomenal Internalism.

Because everyone has tenacious basic phenomenal beliefs and is familiar with hallucinations (or at least dreams), the Phenomenal Internalist Intuition is widespread. Because we are certain of our basic phenomenal beliefs, and the rest of the reasoning outlined above is very persuasive, the Phenomenal Internalist Intuition is powerful. Because our certainty in our basic phenomenal beliefs is ineradicable, even theoretically committed phenomenal externalists (who likely think that such beliefs are sometimes false) often continue to have the Phenomenal Internalist Intuition. Thus, the Tenacity Explanation is a very psychologically plausible hypothesis.

## 2. A pessimistic view

If the Hallucination Explanation is correct, where does this leave the phenomenal externalist? Well, according to the Hallucination Explanation, we accept

Hallucination Objectualism because of Tenacity – because we have tenacious basic phenomenal beliefs. But, while Tenacity says something about our *confidence* in our basic phenomenal beliefs, it is silent on the *truth* of those beliefs. This thesis may be true:

*Optimism:* All basic phenomenal beliefs are true.

If Optimism is true, then it may be a fine thing, epistemically speaking, that we cannot lower our certainty in our basic phenomenal beliefs. Perhaps such ineradicable certainty is appropriate for beliefs which are all true.

Alternatively, perhaps this thesis is correct:

*Pessimism:* Some basic phenomenal beliefs are false.

If Pessimism is correct, then we are in a more epistemically unfortunate position: some of our beliefs are false even though we are certain of those beliefs and simply can't bring ourselves to lower this certainty. Pessimism is the negation of Optimism, so one of these two theses must be true and the other false.

In the remainder of this section, I will argue that Pessimism is true – some basic phenomenal beliefs are false. In the next section, I will argue that these false basic phenomenal beliefs include those that lead us to accept Hallucination Objectualism (and thus Phenomenal Internalism), thus vitiating the Phenomenal Internalist Intuition.

Let's begin with my defense of Pessimism. The primary consideration in favor of Pessimism is that it is extraordinarily difficult to say how Optimism could be true. My strategy will be to canvass plausible explanations of the truth of Optimism, including

the most popular explanations suggested in the literature, and show that none of these explanations succeeds. Certainly my catalogue of candidate explanations of Optimism will be far from exhaustive. Moreover, I do not doubt that some suitably complicated tale may be spun to preserve Optimism. But I hope to show that the costs of being an optimist are significantly higher than it might initially appear, thereby making my own pessimistic phenomenal externalist view more attractive. Now let us consider some candidate explanations for Optimism.

*The Higher-Order Explanation:* Perhaps my basic phenomenal beliefs are beliefs about how things seem to me perceptually. While I may be wrong about how things are, I can't be wrong about how things seem to me. For instance, I may believe [SAN DIEGO IS NORTH OF LAS VEGAS]. Though this belief is in fact false, my belief [IT SEEMS TO ME THAT SAN DIEGO IS NORTH OF LAS VEGAS] can't be false – or so the reasoning goes. Basic phenomenal beliefs are analogous to higher-order beliefs about my own beliefs; since they concern how things perceptually seem to me, they are always true. For example, let *Q* be the phenomenal property associated with my experiences of a particular shade of red – say, red<sub>17</sub>. Then, on this view, my basic phenomenal belief [E IS *Q*] is simply the belief [E SEEMS RED<sub>17</sub> TO ME]. The natural way to develop the view is to suppose that *E* is some physical entity or spatial region.

One worry about the Higher-Order Explanation is that it is quite plausible that I can be wrong about how things seem to me.<sup>8</sup> Though I am sympathetic to this worry, I shan't pursue it here.

Instead, I will pursue the worry that the Higher-Order Explanation is incompatible with the conceivability of individuals who are spectrum-inverted relative

---

<sup>8</sup> Williamson (2002, ch. 4) presses this objection.

to us. When we see limes, we experience a certain phenomenal property (call it  $Q_g$ ), and when we see tomatoes, we experience a different phenomenal property (call it  $Q_r$ ). But consider Jill, who instead experiences  $Q_r$  when she sees limes and  $Q_g$  when she sees tomatoes. Moreover, suppose that these experiences are normal for Jill; limes normally cause her to experience  $Q_r$ , and tomatoes normally cause her to experience  $Q_g$ . Jill also uses color words in the standard way: she says that tomatoes are red and limes are green, and even that tomatoes *look* red and that limes *look* green. I claim that individuals like Jill are coherently conceivable, though I say nothing about the metaphysical possibility of Jill's existence.

What happens when Jill sees the lime? What color does it look? Surely it looks green to her - she sincerely says that it looks green, and she acts with respect to green things just like we do. But Jill takes the lime (or the corresponding region of space) to instantiate  $Q_r$ .<sup>9</sup> Now, according to the Higher-Order Explanation, my basic phenomenal belief [E IS  $Q_r$ ] - formed when I see a tomato - just is the belief [E SEEMS RED TO ME]. Further, according to the Higher-Order Explanation, when Jill sees the lime and experiences  $Q_r$ , she forms the belief [E SEEMS RED TO ME]. But E doesn't seem red to Jill; it seems green to her. So Jill is a metaphysical impossibility on the higher-order account. But this is not the real trouble; many plausible views deny the metaphysical possibility of spectrum-inverted individuals like Jill.

What is troubling is that the Higher-Order Explanation predicts that we will not be able to imaginatively put ourselves in Jill's position. For if the higher-order account is correct, then when I try to imagine Jill's experience of a lime, I form the basic

---

<sup>9</sup> Some think that tomatoes are not what seem to instantiate phenomenal properties like  $Q_r$  or  $Q_g$ . Instead, some mental entities like sense-data appear to instantiate such properties. That's fine for my purposes, for sense-data don't instantiate redness or greenness. So the coherence of the belief [that object is  $Q_r$  but not red] is guaranteed, since the object in question may be a sense-datum.

phenomenal belief [ $E$  IS  $Q_r$ ], which is simply the belief [ $E$  SEEMS RED]. But to imaginatively put myself in Jill's position, I must imagine that the lime looks *green*, not red. Of course, I can easily imagine myself in Jill's situation, and so can you. Since the Higher-Order Explanation entails that we cannot do this, the Higher-Order Explanation must be false.<sup>10</sup>

*The Character Explanation:* One might think that utterances of the following types are always true: "I am here now," "this is there," or "'Holmes' refers to Holmes." Perhaps the truth of these sorts of utterances may be explained by the peculiar way that certain linguistic devices – in particular, indexicals, demonstratives, and quotations – get their reference. (I will call these three types of devices *character devices*, for reasons which will become clear shortly.) This suggests a strategy for defending Optimism. If there are mental analogues to indexicality, demonstration, and quotation – if there are mental character devices – then they may secure their reference in a way that guarantees the truth of basic phenomenal beliefs. To see how this might work, let us examine ordinary-language character devices.

Examples of ordinary-language character devices include "here," "now," "I," "this," "that," and inscriptions in quotation marks (like "'Paul Auster'"). On Kaplan's (1989a and 1989b) analysis, character devices do not have a stable content. Rather, they have *characters* – functions from context to content. For example, in the standard case, a token of "here" uttered by a speaker refers, roughly, to the salient place where the speaker is located. (In the non-standard case, some other contextually relevant place may be referred to.)

---

<sup>10</sup> Note that I am using a point often made by *opponents* of representationalism (like Block (1990)) in service of my defense of phenomenal externalism.

Consider three different types of character devices. *Indexicals* (like “here,” “now,” and “I”) have their character determined solely by context. For instance, “now” standardly refers to the time of the utterance, and “I” standardly refers to the utterer. *Demonstratives* also have their character determined by context, but to secure a referent demonstratives require a *demonstration* (e.g., pointing, gazing in a particular direction, or attending). For example, “this” standardly refers to the object demonstrated. Finally, consider *quotation*. (Here, I consider written inscriptions rather than utterances, for these are the canonical instances of quotation.) “Paul Auster” refers to the name between the quotation marks because that name is between the quotation marks; generally, quotation marks refer to whatever is between them. One may think that quotation is a type of demonstration: the quotation marks demonstrate the object inside them just as a pointing finger demonstrates the object in the appropriate spatial relation to it. (I will not assume the truth of this view; I mention it for expository purposes only.)

The fact that each of these devices has a character rather than a stable content seems to explain the guaranteed truth of certain types of utterances. Consider an utterance of this type: “I am here now.” In any ordinary context, “I” will refer to the speaker of the utterance, “here” will refer to the place of the utterance, and “now” will refer to the time of the utterance. Since (one might think) a speaker must be at the place of the utterance at the time that it is uttered, such utterances are guaranteed to be true. This guarantee is secured by the special way that reference is determined for the terms “I,” “here,” and “now” – by the fact that these terms have characters which operate serendipitously in utterances of the type “I am here now.”

There may be mental analogues of these character-devices (indexicals, demonstratives, and quotations). Let’s consider what these might look like. A *mental*



*indexical* would be, roughly, a mental representation with a character-like reference that did not involve demonstration. Such a representation might refer to whatever place one is currently in, or to the current time. A *mental demonstration*, by contrast, would refer to a demonstrated entity. These (presumably mental) demonstrations might include acts of attention or mental focusing. Explaining a mental analogue of quotation is a bit trickier, but here is a possibility. Perhaps one can possess a *mental representation template* (which is not yet a full-fledged mental representation) which contains an empty “slot”. This mental representation template is analogous to unfilled quotation marks. But once the slot in the template is filled with a mental entity, the template becomes a genuine mental representation that refers to that entity or to a salient property instantiated by that entity.

Now we have an apparently promising model for the truth of basic phenomenal beliefs – beliefs of the form  $\lceil E \text{ IS } Q \rceil$ . Perhaps our mental representations of E and Q are both mental character devices. Perhaps they secure their reference in a way that guarantees the truth of any belief of the form  $\lceil E \text{ IS } Q \rceil$ , just as ordinary-language character devices apparently secure the truth of any utterance of the form “I am here now.” Call this approach the *Character Explanation*, and call the defender of this approach the *character theorist*.

There are many variations on the Character Explanation. For we are considering beliefs of the form  $\lceil E \text{ IS } Q \rceil$ ; one may hold that one’s representation of E is indexical, demonstrative, or quotational, and similarly one may hold that one’s representation of Q is indexical, demonstrative, or quotational.<sup>11</sup> But two arguments suffice to show that all of those permutations are unpromising. The first argument will

---

<sup>11</sup> But note this salient difference between the mental representations of E and Q: a mental representation of E refers to an entity, whereas a mental representation of Q refers to a property.

show that the Character Explanation cannot support Optimism *if* one's mental representations of E and Q are *both* quotational. The second argument will show that all other permutations of the Character Explanation fail.

Here is the first argument. According to the *Double Quotation View*, my mental representations of E and Q are both quotational.<sup>12</sup> Corresponding to each is a mental representation template (I'll refer to these templates as the E-template and the Q-template). The E-template does not refer at all until an entity fills the slot; then it refers to that entity. Similarly, the Q-template does not refer until the slot is filled by an entity; then it refers to the appropriate phenomenal property filling the slot. This view may seem congenial to Optimism, for as long as the same entity fills the slots of my E-template and my Q-template, my belief [E IS Q] must be true. For my E-template picks out an entity, and my E-template refers to a phenomenal property that is then instantiated by that entity.

But Optimism is not so easily protected. For the Double Quotation theorist must concede that my (allegedly quotational) E-template and Q-template are *distinct* – they must be distinct, since one refers to entities and the other to properties. And quotational representations in general refer in virtue of what fills the appropriate “slot” in the mental representation template. So it should sometimes happen that (via some deviant causal process) one entity fills the slot in my E-template while some different entity fills the slot in my Q-template. When this happens, my belief of the form [E IS Q] will likely be false. For example, the slot in my E-template may be filled by an entity that instantiates  $Q_r$ , while the slot in my Q-template is simultaneously filled by an entity that

---

<sup>12</sup> Chalmers (2003) defends the Double Quotation view and suggests that it supports Optimism.

instantiates  $Q_g$ . Then my belief [E IS Q] says of a  $Q_r$  entity that it is  $Q_g$ , and I have a false basic phenomenal belief.

Perhaps the Double Quotation theorist will respond that my E-template and my Q-template are somehow fused, so that the same entity must fill the slot in both templates. But since the E-template and Q-template are distinct – as they must be, since one refers to an entity and the other refers to a property – these templates can at best be connected causally. And if their connection is merely causal, then there will be *some* way of breaking the causal link between the two to produce a false basic phenomenal belief.

Now for my second argument, which applies to all other variations of the Character Explanation. The following point will be crucial for my argument: in almost all cases, beliefs are not instantaneous. If I glance out the window and think [THE LEAVES ON THAT TREE ARE A LOVELY SHADE OF RED], I may run through this belief rapidly, but there is plausibly a moment at which I have only partially completed this belief. To dramatize the point: if I had been vaporized at the right moment, I would have had an incomplete belief like [LOVELY SHADE OF RED]. The same should be true of basic phenomenal beliefs. One does not (or at least *need* not) think, all in an instant, [E IS Q]. Rather, one does something like the following: one focuses (perhaps on a region of physical or mental space or on an entity) and thinks about E; *then* one attends to E's instantiation of Q.<sup>13</sup>

Bearing this point in mind, consider a case like the following. I stare at a red light and begin forming the basic phenomenal belief [E IS  $Q_r$ ]. To make the point vivid, we may suppose that there is an entity instantiating  $Q_r$  literally in my head.<sup>14</sup> But before

---

<sup>13</sup> Or maybe the order of one's belief is reversed: one notes the phenomenal property and then realizes that entity E instantiates the property. For ease of exposition, I'll stick to the order in the text.

<sup>14</sup> The argument goes through just as well, *mutatis mutandis*, if the  $Q_r$  entity is in the world, or in dualistic ether, or elsewhere.

I complete my belief, the light turns blue. My visual system processes this, and a different  $Q_g$  entity is now in my head in place of the  $Q_r$  entity. I then complete my belief. Call this the Switching Case.<sup>15</sup> The Switching Case poses serious problems for the Character Explanation of Optimism.

The character theorist says that my mental representation of E gets its reference via its relation to the  $Q_r$  entity in my head (or wherever). She further says that my mental representation of Q refers to whatever appropriate phenomenal property is instantiated by the entity in the right location in my head (or wherever). What will she say about my belief [E IS Q] in the Switching Case, though? Clearly, she should say that my representation of E refers to a  $Q_r$  entity. That is the only relevantly located entity in my head at the time; whether my representation of E is indexical, demonstrative, or quotational, the  $Q_r$  entity is the most obvious candidate as its referent. Now for the crucial question: what should she say my representation of Q refers to? Surely the natural response is that it refers via the  $Q_g$  entity in my head. For that is the entity in my head at the time I token my representation of Q. Again, whether my representation of Q is indexical, demonstrative, or quotational, it seems that its reference should be fixed by the entity *currently* in the right position in my head. If this is correct, then I think [E IS  $Q_g$ ]. But this belief is false, for my representation of E in fact picks out a  $Q_r$  entity!

Perhaps the character theorist will reply that, whenever I form a phenomenal belief, my mental representations of E and Q get their reference simultaneously. It is never the case that one of these representations gets its reference fixed before the other, so the Switching Case is ruled out. This reply is ad hoc, however, for there is no

---

<sup>15</sup> The Switching Case is much like Hawthorne's (2007, pp. 197-8) "dancing qualia" case, though the details are slightly different.

independent reason for thinking that one *cannot* represent a phenomenal property without simultaneously representing the entity that instantiates that property.<sup>16</sup>

Thus the Character Explanation has great difficulty supporting Optimism. But suppose that the character theorist concedes that Optimism is false. She may be cheered by the fact that, for all I have shown, she may retain this weaker thesis:

*Near-Optimism:* Basic phenomenal beliefs are true as long as they are formed in *stable situations*, i.e., situations in which the context does not change in ways relevant to the reference of the appropriate mental representations.

Near-Optimism may appear to contain the damage. For suppose that unstable situations – situations like the Switching Case – are especially rare. Then, for all I have shown, *almost* all basic phenomenal beliefs are false.

But Near-Optimism does not contain the damage as well as it may seem. For notice that many basic phenomenal beliefs are not formed in stable situations. When I am in motion for any reason – driving a car, walking down the sidewalk, etc. – my visual field is changing very rapidly. Similarly for situations in which I change the direction of my gaze. Suppose that, in some situation of this type, I am rational and attending to the phenomenal properties that I experience. Even so, it may well be that (because of how rapidly my experience is changing) I form false basic phenomenal beliefs, for such cases may be relevantly similar to the Switching Case.

---

<sup>16</sup> This argument, and the Switching Case which drives it, turn on a possible *temporal* separation between applications of a concept of a phenomenal property and a concept of an entity. A similar argument can be made by exploiting a possible *spatial* separation between those applications of concepts instead. I do not have room to explore such an argument here, though I think this argument is also promising.

I have examined many plausible explanations for the truth of Optimism. These explanations all have serious defects. By contrast, we can see quite easily how Pessimism could be true, for there is no general difficulty in explaining how some belief-forming mechanism occasionally goes astray. This suggests (though certainly it does not prove) that Optimism is false.

I will close this section by noting an epistemologically significant feature of Pessimism. It has often been thought that basic phenomenal beliefs provide a perfectly secure foundation for empirical beliefs, for we may hold these beliefs with justifiable certainty. But if Pessimism is true, then while we may be certain of our basic phenomenal beliefs, this certainty is unjustified, and even this seeming bedrock of epistemology has been shaken.

### 3. From pessimism to radical pessimism

Consider three views about basic phenomenal beliefs. According to the *optimist*, all basic phenomenal beliefs are true. According to the *near-optimist*, the only false basic phenomenal beliefs are some subset of those formed in unstable situations, i.e., situations in which the context changes in ways relevant to the reference of the appropriate mental representations. And according to the *radical pessimist*, this thesis is true:

*Radical Pessimism:* Basic phenomenal beliefs formed in some unstable situations and in cases of hallucination, dreaming, etc. are false.

I argued against Optimism in the previous section; now I'll press the case for Radical Pessimism against Near-Optimism.

Here's one consideration in favor of Radical Pessimism. Consider Pam, who has ingested a psychedelic drug and is now hallucinating purple frogs. We know that Pam is in an epistemically defective mental state. If she takes her experience at face value, she will form radically false perceptually-based beliefs. *Given* that Pessimism is true – given that some basic phenomenal beliefs are false – it is natural to trace the epistemic defect in Pam's mental state to her basic phenomenal beliefs. These basic phenomenal beliefs falsely represent something as being phenomenally purple and give rise to Pam's epistemically defective inclination to think that there are purple frogs around. One is not forced to locate the epistemic defect in Pam's basic phenomenal beliefs, but this radically pessimistic hypothesis is an obvious one.

A second reason for favoring Radical Pessimism over Near-Optimism turns on the spatiality thesis defended in Chapter 1. There I argued that some, and arguably all, instantiated phenomenal properties are instantiated external to the subject. But I may hallucinate phenomenal redness even if nothing in my environment or body is phenomenally red. So it is again natural to think that my basic phenomenal belief [E IS PHENOMENALLY RED] is false.

Neither reason for preferring Radical Pessimism to Near-Optimism is decisive, but they are persuasive considerations. Moreover, there's much less to be said for Near-Optimism than Optimism. The main advantage of Optimism is its great intuitiveness. But Near-Optimism is not nearly as intuitive as Optimism. As I pointed out at the end of the previous section, it leaves room for the falsity of a large range of actual phenomenal beliefs (including, for example, all visually-based phenomenal beliefs formed by a typical subject who is moving or looking around).

Thus I recommend the following radically pessimistic phenomenal externalist view. We are in an unfortunate epistemic position – the position of having certain tenacious false beliefs – and we are therefore saddled with wildly unreliable intuitions about phenomenal properties. We think that we are related to phenomenal objects instantiating phenomenal properties even when we hallucinate, dream, imagine, etc., so we accept Hallucination Objectualism and the Phenomenal Internalist Intuition. In fact, there simply is no phenomenal object, and no property instantiated, in such cases: the reasoning that leads us to the Phenomenal Internalist Intuition is vitiating. Perhaps this debunking explanation of the Phenomenal Internalist Intuition will sap our resistance to phenomenal externalist views like subjective representationalism.

#### 4. Some further speculations

If Pessimism is true, then we have a psychological mechanism that produces tenacious but sometimes false beliefs. Why would such an apparently perverse mechanism evolve? In this section, I provide a highly speculative answer to this question.

It is well-known that we have a holistic belief-formation system. If I discover a tension or incompatibility among some of my beliefs, I typically have a choice about which belief(s) to discard and which to retain.

Consider a cognitively sophisticated creature with such a holistic belief-formation system. Such a creature (call her Eve) believes, perhaps with very good



reason, that there are no predators around. She then apparently perceives something shaped and colored like a predator. She now entertains three beliefs:

- (1) [THERE ARE NO PREDATORS AROUND]
- (2) [THAT HAS SHAPE S AND COLORATION C]
- (3) [IF SOMETHING HAS SHAPE S AND COLORATION C, THEN IT IS A PREDATOR]

These beliefs are mutually incompatible. Which will she reject? (1) and (3) may be very well-supported theoretical beliefs. So, *if* Eve's perceptual belief (2) is not designed to resist such holistic pressures, then she will abandon that belief. Obvious and unfortunate consequences will result.<sup>17</sup>

A better-designed creature might instead have some perceptual beliefs that are tenacious, i.e., ineradicably certain. Such beliefs cannot be rejected no matter how poorly they sit with one's other beliefs and thus will not be subject to the trouble just pointed out.<sup>18</sup> If Eve were equipped with such a psychological mechanism, then she could not reject (2). Instead, she would likely reject (1); she would then have a much better chance of avoiding the predator.

*Very* roughly, having a psychological mechanism that produces tenacious perceptual beliefs will be beneficial for a creature that has a holistic belief-formation system and modular perceptual faculties, where the latter is markedly more reliable than the former (at least with respect to opportunities for gaining food, mating opportunities,

---

<sup>17</sup> Fodor (1984) makes a related point.

<sup>18</sup> Why should Eve be *ineradicably certain* in such beliefs rather than merely *very confident*, in a way that is *very hard to eradicate*? Perhaps because speed of action is crucial in such high-stakes situations, and ineradicable certainty eliminates any deliberation whatsoever about whether or not to retain one's basic phenomenal beliefs.

etc.). I speculate that we are such creatures and that evolution has thus outfitted us with tenacious perceptual belief-formation mechanisms.

Much more might be said about this hypothesis, and I hope to explore it further on another occasion.

## Chapter 5.

# Is there a phenomenological argument for higher-order representationalism?

I have been defending a specific form of representationalism, namely, subjective representationalism. Broadly speaking, representationalists analyze experiencing in terms of representing. However, I have not yet taken sides in a major internecine dispute among representationalists, which concerns *how many layers* of representation are required for a state to be an experience.

According to *first-order representationalists*, experiences merely involve ground-level representations of certain properties – say, environmental or bodily properties. First-order representationalists usually also require that such representations play a certain functional role – for instance, that they be available for practical reasoning, verbal report, and/or motor-control. So *first-order representationalism* is the view that a state is a phenomenal experience in virtue of appropriately representing certain properties, perhaps in conjunction with having a certain functional role.<sup>1</sup>

According to *higher-order representationalists*, a mere ground-level representation of this sort is insufficient for a phenomenal experience; experiences also require certain

---

<sup>1</sup> For influential and extensive defenses of first-order representationalism, see Dretske (1995) and Tye (1995) and (2000).

types of higher-level representations targeting these ground-level states.<sup>2</sup> More precisely, *higher-order representationalism* is the view that a state is a phenomenal experience in virtue of both (i) appropriately representing certain properties, perhaps in conjunction with having a certain functional role; and (ii) being the target of, or being disposed to be the target of, the appropriate type of higher-order representation.<sup>3</sup>

Finally, *self-representationalism* is a brand of higher-order representationalism according to which the ground-level representations and the higher-order representations of those ground-level representations are part of a single mental state. Thus, self-representationalism has it that experiences all involve representations of at least certain parts of themselves.<sup>4</sup>

I am tentatively sympathetic to first-order representationalism, as it is the most parsimonious view. However, many arguments have been proposed for adopting a fancier higher-order view. Such arguments tend to have little in common and must thus be examined singly. Since I lack time to discuss the myriad such arguments for higher-order representationalism, this chapter will merely focus on rebutting one recent argument that I find especially intriguing.

In his 2009 article “Self-Representationalism and Phenomenology,” Uriah Kriegel argues for self-representationalism primarily on phenomenological grounds. Kriegel’s argument can naturally be cast more broadly as an argument for higher-order representationalism. In this paper, I will examine this broadened version of Kriegel’s

---

<sup>2</sup> On some varieties of the view, dispositions to form such higher-order representations are enough; actual higher-order representations are not required.

<sup>3</sup> See Carruthers (2000) for a comprehensive defense of this view. Lycan (1996) and Rosenthal (2005) also defend a view they call higher-order representationalism. However, so far as I can tell, the higher-order element in their views is meant to explain only *access* consciousness, not *phenomenal* consciousness (see Block (1995) for discussion of the distinction). So their views may be compatible with first-order representationalism about phenomenal experience.

<sup>4</sup> For a variety of defenses of self-representationalism, see Kriegel and Williford (2006). For an especially complete defense of the view, see Kriegel (2009b).

argument in detail and show that it is unsuccessful for at least two reasons. First, it relies on an inference to the best explanation that is insufficiently defended. Second, phenomenological investigation does not adequately support the key phenomenological premise of the argument.

## 1. The phenomenological case for higher-order representationalism

Let us begin by examining Kriegel's argument. To introduce his argument, Kriegel contrasts *peripheral* awareness with *focal* awareness. As I type these words, I am focally visually aware of my computer screen. But I am simultaneously peripherally visually aware of my keyboard, the wall behind my computer screen, and the hum of the overhead lights. This isn't a precise theoretical characterization of the distinction between focal and peripheral awareness, but Kriegel presumably thinks (and I'm happy to grant) that such examples provide us with a good enough handle on this distinction.

Quite generally, phenomenological investigation suggests that experience is associated with a great deal of peripheral awareness. Now, Kriegel proposes, phenomenological investigation further suggests that experience is associated with peripheral awareness *of the experience itself*. To return to my example in the previous paragraph: when I experience my computer screen, I am also aware, albeit only peripherally, of my *computer-screen experience*.<sup>5</sup> This awareness can *become* focal, as when I

---

<sup>5</sup> See thesis (1) in Kriegel (2009a, p. 361). Kriegel talks of "conscious states" instead of phenomenal experiences, but he states in the abstract of this paper that he is speaking of *phenomenal* consciousness (rather than what Block (1995) calls "access consciousness").

introspect on that experience. But, Kriegel thinks, even before I engage in any such introspection, I am still peripherally aware of my computer-screen experience.

There is nothing special, on Kriegel's view, about my computer-screen experience; such peripheral awareness of one's experiences can be found in any experience one likes. This leads Kriegel to advance the following claim:

*Universal Awareness Thesis:* I have peripheral awareness of all of my phenomenal experiences.<sup>6</sup>

Kriegel says quite explicitly that the Universal Awareness Thesis is a purely phenomenological claim:

"That a certain feature is sometimes *or always* present in consciousness may be phenomenologically manifest to us, say via introspection ... to the extent that there is a genuinely phenomenological argument for self-representationalism, it must take [the Universal Awareness Thesis] as its starting point" (2009a, p. 364, emphasis mine).

From here, Kriegel supposes, plausibly enough, that awareness requires representation: if I am aware of something, then I represent it.<sup>7</sup> When we apply this supposition to the Universal Awareness Thesis, we arrive at the following claim:

*Egocentric Higher-Order Representationalism:* For any phenomenal experience E that I have, I represent E.<sup>8</sup>

---

<sup>6</sup> See thesis (2) in (ibid, p. 361).

<sup>7</sup> See thesis (a) in (ibid, p. 364).

<sup>8</sup> See thesis (2a) in (ibid, p. 365).

Finally, Kriegel takes this to support the following claim via “something like an inference to the best explanation” (2009a, p. 364):

*Necessary Higher-Order Representationalism*: Necessarily, for any phenomenal experience E of subject S, S represents E.<sup>9</sup>

Necessary Higher-Order Representationalism is not quite higher-order representationalism simpliciter. Higher-order representationalism is usually thought of as a thesis about what it is *in virtue of which* a state is a phenomenal experience, while Necessary Higher-Order Representationalism says nothing about this. Still, we may suppose that Necessary Higher-Order Representationalism strongly militates in favor of higher-order representationalism simpliciter.<sup>10</sup> So we have an argument for higher-order representationalism with a purely phenomenological starting point.

Indeed, one may strengthen this argument by inserting an intermediate step between Egocentric Higher-Order Representationalism and Necessary Higher-Order Representationalism. By discussing the topic with others, I can learn that they, too, have peripheral awareness of all of their phenomenal experiences. This, combined with the Awareness-Representation Thesis, supports a broadened version of Egocentric Higher-Order Representationalism. We need not pin down exactly which broadened

---

<sup>9</sup>Ibid. Kriegel’s argument is slightly different than the argument in the text. After putting forward Egocentric Higher-Order Representationalism, Kriegel argues that self-representationalism is the most plausible form of higher-order representationalism. So the conclusion of Kriegel’s inference that is “something like an inference to the best explanation” is that self-representationalism is true. I will ignore this wrinkle, as my criticisms of the argument in the text apply without modification to Kriegel’s original argument.

<sup>10</sup> One might argue, for example, that if higher-order representations did not at least partly constitute experiences, then there ought to be some possible experiences that aren’t accompanied by such higher-order representations. So, if every possible experience is accompanied by a higher-order representation, that must be because higher-order representations do at least partly constitute experiences – that is, if Necessary Higher-Order Representationalism is true, then higher-order representationalism simpliciter must be true.

version of Egocentric Higher-Order Representationalism is best, but to have a concrete thesis in mind I will work with the following claim:

*Restricted Higher-Order Representationalism*: for any phenomenal experience E of any normal adult human subject S, S represents E.

Presumably, Restricted Higher-Order Representationalism (or some nearby thesis) provides even more support for Necessary Higher-Order Representationalism than Egocentric Higher-Order Representationalism provides. Thus, my discussion will focus on this strengthened version of the argument. Call this the *Higher-Order Representationalist Phenomenological Argument*, or, more simply, the *HOR Phenomenological Argument*.

In the next section, I will raise a concern about the inference from Restricted Higher-Order Representationalism to Necessary Higher-Order Representationalism. Finally, in section 3, I will argue that the Universal Awareness Thesis is inadequately supported by phenomenological investigation.

## 2. Worries about the inference to the best explanation

Let us provisionally suppose that the HOR Phenomenological Argument is successful all the way to Restricted Higher-Order Representationalism – the thesis that, for any phenomenal experience E of a normal adult human subject S, S represents E. (I am granting this just for the sake of argument; in the next section I will criticize an earlier step in the argument.) For concreteness, I am focusing just on Restricted



Higher-Order Representationalism, but the criticisms to come will apply to any argument that uses a nearby thesis.

I now want to scrutinize the transition from this premise to Necessary Higher-Order Representationalism – the claim that, necessarily, for any phenomenal experience *E* of *any* subject *S*, *S* represents *E*. My goal in this section is modest: while I will not show that this inference is mistaken, I will show that much more support is required for the inference than Kriegel provides.

Restricted Higher-Order Representationalism is a thesis about a certain limited sample of experiences –the experiences of all normal adult human subjects. Necessary Higher-Order Representationalism is a thesis about all experiences, actual or merely possible. How does Kriegel defend the inference from the former to the latter? In his 2009 article, he says virtually nothing beyond what I quoted in the previous section: that this inference is “something like an inference to the best explanation.”<sup>11</sup>

At first glance, one might think that little more needs to be said: it might appear that Necessary Higher-Order Representationalism is obviously the best explanation of Restricted Higher-Order Representationalism. Why else would all normal adult human subjects represent all of their experiences, if not because this is a necessary requirement for any subject to have an experience? Surely it’s not just a colossal *accident* that any normal adult subject represents all of his or her experiences; surely there is some modal robustness to this fact.

---

<sup>11</sup> Kriegel does make some very brief remarks about the “subjective” or “for-me” character of experience (2009a, pp. 361-2); perhaps he means for these remarks to shed light on the inference to the best explanation, though how to interpret this passage is unclear to me. As far as I can tell, though, Kriegel’s claims about the “subjective” character of experience are grounded in phenomenological investigation, much like the Universal Awareness Thesis. If this interpretation is correct, then I think that my remarks in section 3 about the Universal Awareness Thesis will apply equally well to any claim that experience is subjective.

Fair enough. But to see why this inference is not plain sailing, consider these facts: for any experience *E* of a normal adult human subject *S*, (a) *E* includes a tactile component,<sup>12</sup> (b) *S* has a digestive tract, and (c) *S* has a rich self-concept. Indeed, there is at least some modal robustness to all of these facts. For any given experience *E* of a normal adult human subject *S*, it is plausible that, *in all nearby possible worlds where S has E*, *E* includes a tactile component, *S* has a digestive tract, and *S* has a rich self-concept.

Now, one could perform “something like an inference to the best explanation” on these facts, arriving at claims like this one: “For any experience *E* of a normal adult human subject *S*, *E* includes a tactile component” (or “*S* has a digestive tract,” or “*S* has a rich self-concept”). But any such inference is highly suspect. It is clearly false that any experience must include a tactile component; there are certainly possible experiences (e.g., experiences of creatures that lack any sense of touch) that involve no such component. It is also clearly false that any subject of an experience must have a digestive tract.

It is less clear whether any subject of experience must have a rich self-concept. What *is* clear is that establishing this would take a great deal of work, for it is far from obvious that this is the best explanation of the fact that all actual normal human subjects with experiences have a rich self-concept. Another quite plausible explanation of this fact is simply that human beings have evolved to have a rich self-concept, though having a rich self-concept has nothing to do with the conditions for having an experience.

The moral of such examples is that it is often slippery to infer from the fact that all actual experiences that we’re familiar with have feature *F* to the thesis that,

---

<sup>12</sup> This takes *E* to be the subject’s total experience. For those averse to this approach, I can say instead that *E* occurs at the same time as some experience of *S* with a tactile component.

necessarily, all experiences have feature F. Bearing this moral in mind, let us revisit the inference from Restricted Higher-Order Representationalism to Necessary Higher-Order Representationalism. The latter is supposed to be the best explanation of the former. But is it? Let us consider another initially plausible explanation of Restricted Higher-Order Representationalism.

Consider the explanation that, as a result of natural selection, human beings possess a very sophisticated set of cognitive faculties. In addition to our first-order sensory faculties which monitor our bodies and our environment, we have various faculties which monitor our mental states. One of these faculties produces representations of our experiences. On this view, just as all normal human beings constantly represent at least some features of their environment, they also constantly represent at least some features of all of their experiences. However, these higher-order representations are not required for a state to count as an experience. (Except for this last claim, the higher-order representationalist might happily accept this explanation.)

On this evolutionary explanation, the fact that higher-order representations accompany all experiences of normal adult human subjects may be a very robust matter of fact: there may have been very strong evolutionary pressures against subjects who lacked faculties to monitor their mental states. But, on this explanation, higher-order representations accompany our experiences only as a contingent matter of fact. Moreover, for all this explanation says, there are actual experiences whose subjects do not represent those experiences. Perhaps animals and children have such experiences. The evolutionary explanation of Restricted Higher-Order Representationalism is a plausible rival to the explanation via Necessary Higher-Order Representationalism. But,

at least in his 2009 article, Kriegel gives no reason at all for dismissing the evolutionary explanation.

Here is another way of putting the point. There are at least two types of first-order representationalist views worth contrasting with higher-order representationalism. On one view, many actual paradigmatic experiences are not accompanied by higher-order representations of the relevant sort. But on another view, all actual paradigmatic experiences are accompanied by higher-order representations of the relevant sort. However, what makes these states *experiences* has nothing to do with the presence of these higher-order representations. Similar mental states that were not accompanied by higher-order representations would also count as experiences.

If one overlooks the second view and takes the first view as the only serious rival, then it may seem reasonable to infer from Restricted Higher-Order Representationalism to Necessary Higher-Order Representationalism. For Restricted Higher-Order Representationalism rules out the first view. But it does not rule out the second view, so the inference to Necessary Higher-Order Representationalism requires further support.

The vital question, then, is not whether actual experiences are accompanied by higher-order representations, but whether such higher-order representations are part of what *make* these states experiences. No mere survey of actual experiences will settle this question.

This point is worth lingering on. In general, a theory of experience aims to say what *constitutes* experience; actual experiences may have many non-constitutive features that are nevertheless quite modally robust. Once we recognize this point, we can see that almost any theory of experience is compatible with the claim that all actual

experiences are accompanied by higher-order representations, just as almost any theory of experience is compatible with the claim that all actual experiences include a tactile component. Those who deny higher-order representationalism simply need to deny that these higher-order representations help constitute experiences, ubiquitous though such higher-order representations may be.

In sum, *even if we grant that some thesis like Restricted Higher-Order Representationalism is true*, much work remains to make the case that higher-order representationalism is the best explanation of this thesis. Perhaps this case can be made, but doing so is a substantial task.

### 3. Scrutinizing the Universal Awareness Thesis

In the previous section, I granted my opponent, just for the sake of argument, the truth of Restricted Higher-Order Representationalism. Kriegel's argument for Restricted Higher-Order Representationalism, recall, rested on the Universal Awareness Thesis: the thesis that I have peripheral awareness of all of my phenomenal experiences. Kriegel claims that the Universal Awareness Thesis is phenomenologically manifest; a direct appeal to phenomenology is the only support that Kriegel provides for this thesis. In this section, I will examine this thesis. I will not argue that the Universal Awareness Thesis is false, but I will argue that a straight appeal to phenomenological investigation provides little support for it.

First, a clarification of the thesis. Note that I have good phenomenological access only to my current, and perhaps very recent, experiences. I have virtually no phenomenological access to the nature of my childhood experiences. I can't remember most of those experiences, and even the experiences that I seem to remember may be

subject to memory distortion or fabrication. If the Universal Awareness Thesis is understood as applying to every experience that I have ever had, there is no plausibility to the idea that phenomenological investigation alone could establish the Universal Awareness Thesis. It must be that Kriegel intends the Universal Awareness Thesis to be restricted to my current or very recent phenomenal experiences, and that is how I shall understand this thesis from now on.

Now I will lead up to the objection. First, let me present a helpful example. My remarks about this example are intended to clarify the objection to come, but they will bear no argumentative weight; the objection stands alone.

Suppose that, at time  $t$ , I am visually attending to my computer screen. At a slightly later time  $t^*$ , I shift my attention (without moving my eyes) to the periphery of my visual field. There is a small brown box there, near the boundary of what I can visually attend to without moving my eyes; I'm now focally aware of the box. But was I peripherally aware of that box at time  $t$ , while I was attending to the screen?

Having attempted serious phenomenological investigation of this matter, I state that I have *absolutely no idea*. I simply can't tell either way. I could have had some dim awareness of the box at time  $t$  (in a way that I could not have had even a dim visual awareness of the lamp behind me). But it could also have been that I had no awareness of it at all. And no matter how hard I try to investigate the issue phenomenologically, I report making no progress. Doing some psychological experiments might help, but I don't see how to make progress just by thinking harder about my experience.

Notice that I'm not denying that I have peripheral visual awareness of some objects (like the keyboard right below my computer screen). One could deny this –

indeed, there is a heated debate in the philosophical literature about just such issues.<sup>13</sup> But I'm granting Kriegel that I have peripheral awareness of some things, and that I have a very clear phenomenological sense of such peripheral awareness in a broad range of cases. My point is merely that there are *some* objects such that I have no phenomenological sense whether I was peripherally aware of them just a moment ago.

Perhaps some readers will think that careful phenomenological investigation reveals *no* peripheral awareness of the brown box. Further reflection on the case does not change my own phenomenological verdict of uncertainty, but it's no problem for my argument if phenomenological verdicts differ amongst careful subjects after reflection. To reiterate, I present the brown box case solely to clarify my objection to Kriegel. The role of this case is merely to provide an example where phenomenological investigation delivers no clear verdict about whether or not one has a certain sort of peripheral awareness. Readers with diverging intuitions may substitute some other case of this sort; surely there are such cases.<sup>14</sup>

Here is my main objection. I report that, for me, trying to find peripheral awareness of my experiences themselves is much like trying to find peripheral awareness of the brown box. When I conduct phenomenological investigation into any peripheral awareness of my experiences, attempting to set aside any theoretical prejudices, I *haven't a clue* whether I've ever had such peripheral awareness. I can certainly become focally aware of my experiences; I have no trouble attending to my experiences. But take a moment when I am caught up in the bustle of ordinary life, paying no heed to my experiences. Just after this moment, I have no clear

---

<sup>13</sup> See, for example, discussion of the so-called "refrigerator-light illusion" and related matters in Block (2007), Dennett (1991), Dretske (2007), Schwitzgebel (2007), and Tye (2009).

<sup>14</sup> I thank an anonymous referee from *Philosophical Studies* for helpful discussion on this point.

phenomenological sense whether I was then peripherally aware of my experiences. In stark contrast, I have a very sharp phenomenological sense that I was peripherally visually aware of many external objects.

I want to separate this objection from a different one. One might worry that Kriegel's phenomenological investigations into his awareness of his own experiences are influenced by some background theory (perhaps an innocuous one). If this is so, then the Universal Awareness Thesis (the thesis that one has peripheral awareness of all of one's phenomenal experiences) is not a *purely* phenomenological premise. It is supported by phenomenology *plus* some (perhaps innocuous) background theory.

This is not my objection. I am not worrying about whether the Universal Awareness Thesis is supported purely via phenomenology or not. Presumably, if it is properly supported in any way, then it's fair game for Kriegel to use it in his argument. My objection is that when I conduct such (pure or impure) phenomenological investigation myself, I simply don't reach Kriegel's result.

So far, I've just flatly denied what Kriegel says: he finds it phenomenologically evident that he has peripheral awareness of his own experiences, and I don't. Is there any way of making progress in this standoff?

Well, my discussions of the issue with others suggests that I am not idiosyncratic here. And Kriegel himself admits that peripheral awareness of one's experiences is "singularly elusive among mental phenomena" (2009a, p. 379). Given these facts, I predict with some confidence that if we tabulated the results of philosophers' careful phenomenological investigations, the best result Kriegel could realistically hope for is *recalcitrant disagreement*, with a hefty percentage of philosophers on either side of the issue.



If we did arrive at this result – which, as I’ve suggested, is the best case scenario for Kriegel – what bearing would this have on Kriegel’s argument? I suggest that it would make the argument very easy to resist. As long as prolonged, careful phenomenological investigation delivers no clear verdict, it is perfectly reasonable to deny that we are always peripherally aware of our experiences. For precisely this reason, I am skeptical quite generally of attempts to build theories of experience around highly controversial phenomenological “data.” Given that there are quite a lot of shared intuitions about experience (phenomenologically based and otherwise), won’t *those* data provide a better starting point for theories of experience?

Even if one denies that we have a phenomenological sense of peripheral awareness of our experiences, how could Kriegel and those like him have gotten their own phenomenology wrong? I offer the following explanation, which relies on a *very* diluted construal of the famous *transparency of experience*. Recall Tye’s statement of the transparency point (2002, p. 139):

When you introspect your visual experience, the only particulars of which you are aware are the external ones making up the scene before your eyes. You are not aware of those objects *and* a further inner object or episode.

Here, Tye makes a negative claim – that one is not introspectively aware of any internal particulars – as well as a positive claim – that one is introspectively aware of external particulars.

As I noted earlier in this dissertation, the transparency thesis is enormously controversial. Fortunately, I do not need to appeal to it, but only to this diluted claim which should be much less controversial:

*Minimal Transparency Thesis:* Normally, when one becomes aware of one's experience in the direct way associated with introspection, one does so by first becoming aware of the properties presented by that experience.

The Minimal Transparency Thesis doesn't claim that *merely* becoming aware of the properties presented by an experience makes one aware of that experience. It's natural to think that one must do something more to acquire such awareness. The thesis claims only that awareness of the properties presented by an experience is normally required for awareness of that experience.

The Minimal Transparency Thesis is much weaker than the transparency thesis simpliciter. For one thing, it does not involve any negative claim to the effect that introspection involves no awareness of any internal particulars. Further, its scope is sharply restricted: it concerns how we normally become aware of our experiences, not what nomologically or metaphysically *must* happen for us to become aware of our experiences. For these reasons, it should be far less controversial than the transparency thesis.

Examples support the Minimal Transparency Thesis. For instance: suppose that, while I am looking closely at a peach, I consider the nature of my visual experience. To become aware of my peach-experience, surely I must first become aware of the visually apparent properties of the peach. I might become aware of the particular (apparent) color and shape of the peach, and thereby become aware of what my experience of the peach is like. Rephrasing the point: I cannot bypass the (apparent) properties of the peach when I try to introspect my experience of the peach. To tell what my *experience* of the peach is like, I must first know what the *peach* is apparently like. This example is

representative. Since the Minimal Transparency Thesis concerns only what *normally* happens, unrepresentative counterexamples would not threaten it.

To forestall one possible source of resistance to the Minimal Transparency Thesis, I emphasize that it is compatible with a *very* wide range of views about the nature of introspection. It should be immediately clear that this thesis says nothing about whether introspection is inferential or non-inferential, or whether it provides object-awareness or merely fact-awareness of our experiences.<sup>15</sup> To further demonstrate the relative innocuousness of the Minimal Transparency Thesis, let me show how it pairs with several plausible accounts of introspection. It is unsurprising that the Minimal Transparency Thesis pairs well with the accounts of introspection offered by proponents of stronger transparency theorists, like Tye<sup>16</sup> and Dretske.<sup>17</sup> Thus I will use my limited space to focus on accounts of introspection less obviously compatible with the Minimal Transparency Thesis.

Consider a *syntactic* theory of introspection, according to which introspection of one's experiences involves merely simple syntactic transformation. On a paradigmatic version of this kind of view, introspection works as follows. Introspection (of experiences) takes experiences as input and yields a mental representation as output. Specifically, when I have an experience that presents certain properties  $F_1 \dots F_n$ ,

---

<sup>15</sup> Dretske has consistently underscored the distinction between object- and fact-awareness; see his (1999) for an illustrative and extended discussion.

<sup>16</sup> See his (2000) and (2009) for accounts of introspection compatible with his transparency thesis.

<sup>17</sup> Dretske endorses this strong transparency thesis: "One is made aware of what a pumpkin experience is like ... not by an awareness of the experience, but by an awareness of the pumpkin and awareness of its (the pumpkin's) properties" (1999, p. 112). Dretske offers different accounts of introspection in his (1995) and (1999), both compatible with the Minimal Transparency Thesis.

introspection can take that state as input and deliver as output a state that represents *that I am now experiencing*  $F_1 \dots F_n$ .<sup>18</sup>

The syntactic theorist about introspection can gladly accept the Minimal Transparency Thesis. The idea might be that activation of my introspective faculty on an experience that presents certain properties normally requires me to focus on those properties; normally, only by focusing on the properties presented by an experience can I trigger introspection of that experience. This thought has some independent plausibility: there must be *something* that directs introspection to operate on (say) my visual experience rather than my auditory experience, and focusing on visually presented properties is a reasonable candidate for part of what normally does this.

Now consider *inner-sense* theories of introspection, which hold that introspection is a faculty of inner sense importantly analogous to familiar faculties of outer sense (like vision, hearing, etc.). Inner-sense theorists tend to lack enthusiasm for transparency theses like Tye's (see Lycan (2002), for example). Surprisingly, though, the Minimal Transparency Thesis plays quite nicely with inner-sense theories. A paradigmatic version of inner-sense theory is Lycan's view, where introspection involves the activation of "banks of ... second-order internal monitors" (1996, p. 22) that scan our first-order sensory faculties like vision, taste, etc.

Now, minimal transparency is preserved as long as the exercise of inner sense – these second-order internal monitors – on a first-order sensory faculty normally requires first becoming aware of the properties of the external world presented by that first-order faculty. It's plausible that this is required for the same reasons as on the syntactic view: inner sense needs some cue about which outer-sense faculties to monitor,

---

<sup>18</sup> Nichols and Stich (2003) propose such a theory of the introspection of beliefs and desires. Such an account could easily be extended to the introspection of experiences.

and awareness of the properties presented by outer-sense is a good candidate for a typical part of that cue.<sup>19</sup>

Finally, and most relevantly, let me connect the Minimal Transparency Thesis to Kriegel's theory of introspection. One would expect Kriegel's theory to be compatible with the Minimal Transparency Thesis, since Kriegel (2009a, pp. 371-376) argues that his theory is compatible with the stronger full-fledged transparency thesis. And, as expected, Kriegel's theory of introspection is compatible with the Minimal Transparency Thesis.

On Kriegel's theory, any phenomenally conscious state consists of a first-order representation (typically of some environmental/bodily property) and a higher-order representation of the phenomenally conscious state itself. Such representations provide awareness, so one is typically experientially aware of both (i) some environmental/bodily property and (ii) of the experience itself. Further, normally one's awareness of one's environment/body is focal, and one's awareness of the experience itself is peripheral. In introspection, however, one "reorganiz[es] the center/periphery structure of one's overall experience, by transforming one's peripheral inner awareness of one's current experience into a focal one" (2009a, p. 372). In other words, in introspection one normally promotes one's awareness of the experience to focal awareness, while demoting one's awareness of one's environment/body to peripheral awareness.

---

<sup>19</sup> As an anonymous referee pointed out, some inner-sense theorists will balk at this suggestion. For some inner-sense theorists hold that conscious awareness of any external-world property F involves both a first-order representation of F and a higher-order representation of the first-order representation of F. The idea is that the higher-order representation is what makes the first-order representation of F conscious. So genuine conscious awareness of F cannot precede the higher-order representation of the first-order representation of F.

Still, one may hold an inner-sense theory of introspection without taking on this additional commitment about conscious awareness. So the Minimal Transparency Thesis remains compatible with an inner-sense theory of introspection.

This view is perfectly compatible with the Minimal Transparency Thesis. This thesis, recall, says that normally, when one becomes aware of one's experience in the direct way associated with introspection, one does so by first becoming aware of the properties presented by that experience. On Kriegel's view, "awareness of one's experience in the direct way associated with introspection" amounts to a certain kind of *focal* awareness. And Kriegel thinks that this focal awareness of the experience is normally preceded by focal awareness of the environmental/bodily properties presented by that experience. This does not quite entail the Minimal Transparency Thesis. For Kriegel does not say that one normally becomes focally aware of one's experience *by* first becoming aware of these bodily/environmental properties. The Minimal Transparency Thesis does say this. But there is no obstacle to adding this proviso to Kriegel's theory.

So Kriegel's theory of introspection is compatible with the Minimal Transparency Thesis. What bearing does this have on my response to Kriegel? Well, it means that any remarks Kriegel has made in favor of his theory of introspection do not undermine the Minimal Transparency Thesis; that's good news for my response. Nor do I know of anything else Kriegel has said that would undermine the Minimal Transparency Thesis.

That concludes my discussion relating the Minimal Transparency Thesis to extant theories of introspection. Even so, this thesis may be controversial, and I lack room here to adequately defend it.<sup>20</sup> To those who reject the Minimal Transparency Thesis, I note that my primary point about the dialectic stands without it: Kriegel's central phenomenological claim is just too disputable to be especially dialectically

---

<sup>20</sup> Am I violating my own stricture against resting too much weight on controversial phenomenological points? No, for I think that the Minimal Transparency Thesis should be widely accepted, given a proper defense. Only unnecessarily bold transparency theses are controversial.

effective. But those who accept the Minimal Transparency Thesis can also (I will argue shortly) see why one might mistakenly think that the Universal Awareness Thesis is phenomenologically manifest. Indeed, if my argument succeeds, the fact that the Minimal Transparency Thesis can explain Kriegel's tempting mistake may provide some measure of support for the Minimal Transparency Thesis itself.

Consider this analogy: I may become aware of my cat by becoming aware of her paw prints on my kitchen tiles. But I can be aware of the paw prints without being aware of the cat. Similarly, one might think, I normally become aware of my experience by first becoming aware of the properties that my experience presents to me; but I can be aware of those properties without being aware of my experience.

Now here is how the Minimal Transparency Thesis can explain Kriegel's phenomenological mistake. When I am looking at the peach, I am aware of certain properties apparently of the peach. According to the Minimal Transparency Thesis, becoming aware of my experience of the peach normally requires awareness of just the same properties. So it is easy to become confused about whether, when I am looking at the peach, I am focally aware of these properties of the peach *and* peripherally aware of my peach-experience, or merely focally aware of these properties of the peach – and thereby focally aware of the properties which *are required for* awareness of my peach-experience – *while still lacking* any awareness of my experience. Put another way: it is easy to become confused between *being aware of the properties which put me in a position to become aware of my experience*, and *actually being aware of my experience*.

I conclude that the HOR Phenomenological Argument fails. The inference to the best explanation from Restricted Higher-Order Representationalism (or some nearby thesis) to Necessary Higher-Order Representationalism requires substantial further

support. Further, the pivotal Universal Awareness Thesis cannot be adequately supported solely via phenomenological investigation. Finally, the Minimal Transparency Thesis (which is compatible with many theories of introspection, including Kriegel's) explains the misleading temptation to think otherwise. Thus, the HOR Phenomenological Argument ought not move opponents of higher-order representationalism.



# Bibliography

- Ayer, A. J. (1963). *The Foundations of Empirical Knowledge*. London: Macmillan Press.
- Block, N. (1990). "Inverted earth." *Philosophical Perspectives* 4: 53-79.
- Block, N. (1995). "On a confusion about a function of consciousness." *Behavioral and Brain Sciences* 18, 227-287.
- Block, N. (1996). "Mental paint and mental latex." *Philosophical Issues* 7: 19-49.
- Block, N. (2007). "Consciousness, accessibility, and the mesh between psychology and neuroscience." *Behavioral and Brain Sciences*, 30, 481-499.
- Block, N. and Stalnaker, R. (1999). "Conceptual analysis, dualism, and the explanatory gap." *Philosophical Review* 108: 1-46. Cornell University Press. Reprinted in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David Chalmers (New York: Oxford University Press, 2002), 371-394.
- Boghossian, P. and Velleman, D. (1989). "Colour as a secondary quality." *Mind* 98: 81-103.
- Bourget, D. (2010). "Consciousness is underived intentionality." *Noûs* 44 :32-58.
- Brewer, B. (1995). "Bodily awareness and the self." In *The Body and the Self*, eds. Jose Luis Bermudez, Anthony J. Marcel, and Naomi M. Eilan (Cambridge, Massachusetts: MIT Press).

- Brewer, B. (2008.) "How to account for illusion." In *Disjunctivism: Perception, Action, Knowledge*, eds. A. Haddock and F. Macpherson (Oxford: Oxford University Press).
- Briscoe, R. (2009). "Egocentric spatial representation in action and perception." *Philosophy and Phenomenological Research* 79: 423-460.
- Brogaard, B. (2010). "Strong representationalism and centered content." *Philosophical Studies* 151: 373-392.
- Burge, T. (1979). "Individualism and the mental." *Midwest Studies in Philosophy* 4: 73-121.  
Reprinted in *Externalism and Self-Knowledge*, eds. P. Ludlow and N. Martin (Stanford: CSLI Publications, 1998), 21-83.
- Burge, T. (2010). *Origins of Objectivity*. Oxford: Clarendon Press.
- Byrne, A. (1999) "Cosmic hermeneutics," *Philosophical Perspectives* 13: 347-383.
- Byrne, A. (2001). "Intentionalism defended." *Philosophical Review* 110 (2): 199-240.
- Byrne, A. (2006). "Color and the mind-body problem." *Dialectica* 60 (3): 223-244.
- Byrne, A. (2009). "Experience and content." *Philosophical Quarterly* 59 (236): 429-451.
- Byrne, A. and Hilbert, D. (2003). "Color realism and color science." *Behavioral and Brain Sciences* 26: 3-64.
- Byrne, A. and Logue, H. (eds.). (2009). *Disjunctivism: Contemporary Readings*. Cambridge, MA: MIT Press.
- Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.
- Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. New York: Cambridge University Press.
- Carruthers, P. (2004). "Reductive explanation and the 'explanatory gap.'" *Canadian Journal of Philosophy* 34: 153-173.

- Carruthers, P. (2005). *Consciousness: Essays From a Higher-Order Perspective*. Oxford University Press.
- Chalmers, D. (1996). *The Conscious Mind*. New York: Oxford University Press.
- Chalmers, D. (2003). "The content and epistemology of phenomenal belief." In eds. Q. Smith and A. Jokic, *Consciousness: New Philosophical Perspectives*. Oxford University Press.
- Chalmers, D. (2006a). "Perception and the fall from Eden." In *Perceptual Experience*, eds. T. Gendler & J. Hawthorne (Oxford University Press, 2006).
- Chalmers, D. (2006b). "The foundations of two-dimensional semantics." In eds. M. Garcia-Carpintero and J. Macia, *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press.
- Chalmers, D. (2009) "The two-dimensional argument against materialism," *The Oxford Handbook of the Philosophy of Mind*, ed. Brian McLaughlin (Oxford).
- Chalmers, D. and Jackson, F. (2001). "Conceptual analysis and reductive explanation." *Philosophical Review* 110: 315-361.
- Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown, and Co.
- Dretske, F. (1995a). *Naturalizing the Mind*. Cambridge, Massachusetts: MIT Press.
- Dretske, F. (1995b). "Phenomenal externalism, or, if meanings ain't in the head, where are qualia?" *Philosophical Issues* 7, 143-158.
- Dretske, F. (1999). "The mind's awareness of itself." *Philosophical Studies* 95: 103-124.
- Dretske, F. (2003). "Experience as representation." *Philosophical Issues* 13: 67-82.
- Dretske, F. (2007). "What change blindness teaches about consciousness." *Philosophical Perspectives*, 21, 215 – 230.
- Egan, A. (2006). "Appearance properties?" *Noûs* 40 (3): 495-521.

- Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Fish, W. (2009). *Perception, Hallucination, and Illusion*. Oxford: Oxford University Press.
- Fodor, J. (1984). "Observation reconsidered." *Philosophy of Science* 51, 23-43.
- Fodor, Jerry. (1990) *A Theory of Content and Other Essays*. Cambridge: MIT Press.
- Fodor, Jerry. (1998) *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.
- Gertler, B. (2001). "Introspecting phenomenal states." *Philosophy and Phenomenological Research* 63 (2), 305-328.
- Harman, G. (1990). "The intrinsic quality of experience." *Philosophical Perspectives* 4: 31-52.
- Harman, G. (2001). "General foundations versus rational insight." *Philosophy and Phenomenological Research* 63: 657-663.
- Hawthorne, J. (2007). "Direct reference and dancing qualia." In eds. T. Alter and S. Walter, *Phenomenal Concepts and Phenomenal Knowledge*, pp. 195-209. New York: Oxford University Press.
- Hill, C. (1991). *Sensations: A Defense of Type Materialism*. Cambridge University Press.
- Hill, C. (2009). *Consciousness*. New York: Cambridge University Press.
- Horgan, T. and Tienson, J. (2002). "The intentionality of phenomenology and the phenomenology of intentionality." In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers (Oxford University Press).
- Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge University Press.
- Jackson, F. (1982). "Epiphenomenal qualia." *Philosophical Quarterly* 32: 127-36.
- Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

- Jackson, F. (2007). "The knowledge argument, diaphanousness, representationalism." In *Phenomenal Concepts and Phenomenal Knowledge*, eds. T. Alter and S. Walter (New York: Oxford University Press), 52-64.
- Johnston, M. (2004). "The obscure object of hallucination." *Philosophical Studies* 120: 113-83.
- Kaplan, D. (1989a). "Demonstratives." In *Themes from Kaplan*, eds. J. Almog, J. Perry, and H. Wettstein (Oxford: Oxford University Press).
- Kaplan, D. (1989b). "Afterthoughts." In *Themes from Kaplan*, eds. J. Almog, J. Perry, and H. Wettstein (Oxford: Oxford University Press).
- Kennedy, M. (2009). "Heirs of nothing: the implications of transparency." *Philosophy and Phenomenological Research* 79 (3):574-604.
- Kind, A. (2003). "What's so transparent about transparency?" *Philosophical Studies* 115: 225-244.
- Kriegel, U. (2009a). "Self-representationalism and phenomenology." *Philosophical Studies*, 143, 357-381.
- Kriegel, U. (2009b). *Subjective consciousness*. New York: Oxford University Press.
- Kriegel, U. (2011). *The Sources of Intentionality*. Oxford University Press.
- Kriegel, U. and Williford, K. (eds.). (2006). *Self-Representational Approaches to Consciousness*. MIT Press.
- Levine, J. (1983). "Materialism and qualia: the explanatory gap." *Pacific Philosophical Quarterly* 64: 354-361.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.

- Levine, J. (2010) "The Q factor: modal rationalism vs. modal autonomism," *Philosophical Review* 119 (3): 365-380.
- Lewis, D. (1966). "An argument for the identity theory." *Journal of Philosophy* 63: 17-25.
- Lewis, D. (1979). "Attitudes *de dicto* and *de se*." *Philosophical Review* 88: 513-543.
- Lewis, D. (1983). "Extrinsic properties." *Philosophical Studies* 44, 197-200.
- Loar, B. (2003.) "Phenomenal intentionality as the basis for mental content." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, eds. M. Hahn and B. Ramberg (Cambridge: MIT Press).
- Lycan, W. (1987). "Phenomenal objects: a backhanded defense." *Philosophical Perspectives* 3: 513-26.
- Lycan, W. (1996a). "Layered perceptual representation." *Philosophical Issues* 7: 81-100.
- Lycan, W. (1996b). *Consciousness and Experience*. Cambridge, Massachusetts: MIT Press.
- Lycan, W. (2001). "The case for phenomenal externalism." *Philosophical Perspectives* 15, Metaphysics.
- Lycan, W. (2002). "Dretske's ways of introspecting." In *Privileged Access: Philosophical Accounts of Self-Knowledge*, ed. Brie Gertler (Ashgate Publishing Co.), pp. 15-30.
- Macpherson, F. and Haddock, A. (eds.) (2008). *Disjunctivism: Perception, Action, Knowledge*. Oxford: Oxford University Press.
- Martin, M. (2002). "The transparency of experience." *Mind and Language* 17 (4): 376-425.
- Martin, M. (2006). "On being alienated." In *Perceptual Experience*, eds. T. Gendler and J. Hawthorne (Oxford University Press).

- McDowell, J. (2008). "The disjunctive conception of experience as material for a transcendental argument." *Disjunctivism: Perception, Action, Knowledge*, ed. A. Haddock and F. Macpherson (Oxford: Oxford University Press).
- Millar, B. (2010). "Peacocke's trees." *Synthese* 174: 445-461.
- Millikan, R. (1989). "Biosemantics." *Journal of Philosophy* 86: 281-297.
- Moore, G. E. (1903). "The refutation of idealism." *Mind* 12, 433-53.
- Nagel, T. (1974). "What is it like to be a bat?" *The Philosophical Review*, Vol. 83, No. 4: 435-450.
- Nichols, S. and Stich, S. (2003). *Mindreading*. Oxford: Oxford University Press.
- Nida-Rümelin, M. (2007). "Grasping phenomenal properties." In *Phenomenal Concepts and Phenomenal Knowledge*, eds. T. Alter and S. Walter (New York: Oxford University Press), pp. 307-338.
- Pautz, A. (2010). "A simple view of consciousness." In *The Waning of Materialism*, eds. Bealer and Koons (Oxford University Press).
- Peacocke, C. (1983). *Sense and Content*. Oxford: Oxford University Press.
- Perry, J. (1979). "The problem of the essential indexical." *Noûs* 13: 3-21.
- Prosser, S. (forthcoming). "Affordances and phenomenal character in spatial perception." *Philosophical Review*.
- Putnam, H. (1975). "The meaning of 'meaning.'" *Minnesota Studies in the Philosophy of Science* 7: 131-193.
- Rey, G. (1998). "A narrow representationalist account of qualitative experience." *Philosophical Perspectives* 12 (Language, Mind, and Ontology), 435-57.
- Rosenthal, D. (2005). *Consciousness and mind*. New York: Oxford University Press.

- Schroer, R. (2007). "The reticence of visual phenomenal character: a spatial interpretation of transparency." *Australasian Journal of Philosophy* 85 (3): 393-414.
- Schwenkler, J. (forthcoming). "Does visual spatial awareness require the visual awareness of space?" *Mind and Language*.
- Schwitzgebel, E. (2007). "Do you have constant tactile experience of your feet in your shoes? Or is experience limited to what's in attention?" *Journal of Consciousness Studies*, 14, 5-35.
- Shoemaker, S. (1982). "The inverted spectrum." *Journal of Philosophy* 79: 357-381.
- Shoemaker, S. (1994a). "Phenomenal character." *Noûs*, Vol. 28, No. 1: 21-38.
- Shoemaker, S. (1994b). "Self-knowledge and inner sense, lecture III: The phenomenal character of experience." *Philosophy and Phenomenological Research* 54 (2): 291-314.
- Shoemaker, S. (2000). "Phenomenal character revisited." *Philosophy and Phenomenological Research* 60 (2): 465-467.
- Shoemaker, S. (2001). "Introspection and phenomenal character." *Philosophical Topics* 28 (2): 247-73.
- Shoemaker, S. (2003). "Content, character, and color." *Philosophical Issues* 13 (1): 253-78.
- Shoemaker, S. (2006). "On the way things appear." In *Perceptual Experience*, eds. T. Gendler & J. Hawthorne, Oxford University Press.
- Siewert, C. (2004). "Is experience transparent?" *Philosophical Studies* 117: 15-41.
- Smith, A. D. (2008). "Translucent experiences." *Philosophical Studies* 140: 197-212.
- Stoljar, D. (2004). "The argument from diaphanousness." In M. Escurdia, Robert J. Stainton and Christopher D. Viger (eds.), *Language, mind and world: special issue of the Canadian Journal of Philosophy*. University of Alberta Press.



- Sundström, P. (2007). "Colour and consciousness: untying the metaphysical knot." *Philosophical Studies* 136 (2): 123-165.
- Thompson, B. (2007). "Shoemaker on phenomenal content." *Philosophical Studies* 135: 307-334.
- Thompson, B. (2009). "Senses for senses." *Australasian Journal of Philosophy* 87 (1): 99-117.
- Thompson, B. (2010). "The spatial content of experience." *Philosophy and Phenomenological Research* 81 (1): 146-184.
- Tye, M. (1995). *Ten Problems of Consciousness: a Representational Theory of the Phenomenal Mind*. Cambridge, Massachusetts: MIT Press.
- Tye, M. (1996a). "Orgasms again." *Philosophical Issues* 7:51-54.
- Tye, M. (1996b). "Perceptual experience is a many-layered thing." *Philosophical Issues* 7: 117-126.
- Tye, M. (1998). "Inverted earth, swampman, and representationism." *Philosophical Perspectives* 12, 459-477.
- Tye, M. (2000a). "Shoemaker's *The First-Person Perspective and Other Essays*." *Philosophy and Phenomenological Research* 60 (2): 461-464.
- Tye, M. (2000b). *Consciousness, Color, and Content*. Cambridge, Massachusetts: MIT Press.
- Tye, M. (2002). "Representationalism and the transparency of experience." *Noûs* 36 (1): 137-51.
- Tye, M. (2009). *Consciousness Revisited*. Cambridge, Massachusetts: MIT Press.
- White, S. (2007). "Property dualism, phenomenal concepts, and the semantic premise." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness*

*and Physicalism*, eds. T. Alter and S. Walter (New York: Oxford University Press), pp. 210-248.

Williamson, T. (2002). *Knowledge and Its Limits*. Oxford: Oxford University Press.