11-5-2015 12:00 AM

Representationalism About Sensory Phenomenology

Matthew Ivanowich, *The University of Western Ontario*

Supervisor: Chris Viger, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Philosophy

REPRESENTATIONALISM ABOUT SENSORY PHENOMENOLOGY

(Thesis format: Monograph)

by

Matthew Ivanowich

Graduate Program in Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

This dissertation examines *representationalism about sensory phenomenology*—the claim that for a sensory experience to have a particular phenomenal character is a matter of it having a particular representational content. I focus on a particular issue that is central to representationalism: whether reductive versions of the theory should be *internalist* or *externalist*. My primary goals are (i) to demonstrate that externalist representationalism fails to provide a reductive explanation for phenomenal qualities, and (ii) to present a reductive internalist version of representationalism that utilizes the empirical framework of psychophysics and neuroscience to develop a philosophical theory of content. The bulk of the project is an attempt to provide the outlines of what such an empirically-based representationalist theory would look like.

In chapter 2, I argue that reductive externalist representationalism fails because it is vulnerable to the problem of *bad structural correlation*: the mismatch between the structure of phenomenal qualities and the structure of physical properties "tracked" by our sensory systems with which phenomenal qualities are identified. In chapter 3, I develop the outlines of a reductive internalist version of representationalism that characterizes phenomenal qualities as *modes of presentation* of sensory representations that can be reductively characterized in terms of a location in a psychophysically-defined *quality space* (described in chapter 4), which can in turn be given a neurophysiological interpretation. In addition, I argue that the neural mechanisms responsible for the representation of spatial locations and visual objects serve a referential role in sensory psychosemantics that grounds the intentionality of sensory states. In chapter 5 I provide a detailed account of the representation of spatial locations in sensory experience, and in chapter 6 I examine the "binding problem", demonstrating how object-based sensory representations function to demonstratively pick out sensory individuals without explicitly representing any of their features.

I describe the resulting view as *methodological representationalism*: an attempt to demonstrate how a particular philosophical theory of sensory phenomenology (representationalism) can be integrated into the empirical framework of cognitive science, and thereby provide an explanatory psychosemantic framework for sensory phenomenology that is valuable to both philosophers and cognitive scientists.

Keywords

representationalism, phenomenology, phenomenal content, perception, sensory, quality space, experience, Russellianism, Fregeanism, tracking, spatial representation, binding problem, intentionality, mode of presentation

Acknowledgments

I would like to extend my deepest thanks to my supervisor Chris Viger for his invaluable guidance, support, and patience; without him this project would not have been possible. I would also like to extend my sincere thanks to John Nicholas, who has provided me with countless wonderful discussions and debates on these and other issues, and Angela Mendelovici, who provided indispensible feedback on this project. I am also indebted to Mel Goodale for graciously hosting me in his lab and providing me with wonderful opportunities for interdisciplinary research. In addition, I would like to thank my fellow graduate student friends and colleagues in the philosophy department, whose contributions both to this project and to others have made my time at Western truly remarkable. Finally, I would like to thank my parents for their unwavering support over the years and years. But most of all, I want to thank my wonderful partner and the love of my life, Christina. Her infinite love and encouragement mean more to me than I could ever possibly say.

Table of Contents

Chapter 1

1 Introduction

In this chapter, I introduce the problem of explaining the qualitative phenomenology of sensory experience in naturalistically acceptable terms. I then describe one of the most popular approaches to this problem—*representationalism*—and attempt to explain its ubiquity in both philosophy and psychology/neuroscience. I argue that a representationalist theory has two parts (namely, a theory of sensory content and a theory of sensory intentionality), and focus on examining two particular issues that are central to any representationalist theory: whether it is *reductive*, and whether it is *internalist* or *externalist*. Finally, I lay out the goal of the rest of my dissertation: arguing that any adequate theory of representationalism about sensory phenomenology should be both reductive and internalist, and should utilize the empirical framework developed by psychophysics and neuroscience to flesh out a philosophical theory of content.

1.1 Representationalism about Sensory Phenomenology

Sensory experiences possess certain *phenomenal qualities* such that there is, as Thomas Nagel (1974) puts it, 'something that it's like' to undergo that experience.[1]

Representationalism about sensory phenomenology can be understood as the claim that for a sensory experience to have a particular phenomenal quality is a matter of it having a particular representational content. This view is not only widespread in contemporary philosophy of mind, but moreover, characterizing sensory phenomenology in terms of representational content is a ubiquitous practice in psychology and neuroscience.

Of course, there are a variety of ways to interpret the claim that for a sensory experience to have a particular phenomenal quality is a matter of it having a particular

[1] Traditionally, philosophers have often used the term "qualia" to refer to the qualitative aspect of sensory experience, but that term has become loaded with problematic theoretical baggage. (See, e.g., Dennett, 1988.) A variety of other names such as "phenomenal character", "sensory qualities", and "raw feels" have also been used, but I will temporarily set these aside here for the more neutral "phenomenology", and return to clarify this issue in section 1.3 below.

representational content. Minimally, representationalism about sensory phenomenology entails that if there is a difference in the phenomenal quality of an experience there must be a corresponding difference in the representational content of that experience. Different versions of the theory differ with respect to the exact nature of the relationship between phenomenal quality and representational content – there are *identity* theories and *supervenience* theories.

The supervenience version of the theory—sometimes referred to as "weak" representationalism—simply holds that sensory experiences have representational content, and that the phenomenology of an experience depends on the representational content of the experience in the "minimal" sense described above. The problem with supervenience representationalism is that it is so weak that it is practically uncontroversial: nearly everybody agrees that sensory experience is representational in some sense – even those who argue against representationalism.[2] (e.g., Block) Furthermore, it seems obvious that if the phenomenology of a sensory experience changes, then the way the world seems to the subject—the way that the world is *represented*—has changed. What's more, weak representationalism faces an additional problem: it fails to provide an *explanation* for sensory phenomenology, and indeed is compatible with a variety of different views concerning the nature of phenomenal qualities as well as the nature of their relation to representational contents.

Strong representationalism is the claim that the phenomenology of sensory experience is *identical to* a type of representational content (typically referred to as "phenomenal content"). Put somewhat differently, strong representationalism holds that the phenomenology of a sensory experience is *exhausted by* its representational content. On this view, neural states are the representational vehicles of phenomenal experiences, like ink marks on a page are the representational vehicles of words. As Tye (2000) puts it,

---

[2] It should be noted that it has been argued that not all types of qualitative phenomenology supervene on representational contents. For example, some philosophers argue that emotions and moods have a phenomenal character but lack representational contents, as does the affective nature of pain. However, I am restricting my claims here to representationalism about *sensory* phenomenology; that is, to the phenomenology of perceptual experience. If there are indeed additional types of phenomenology over and above sensory or perceptual phenomenology, my argument does not touch them.

strong representationalism is a metaphysical claim that aims to *explain* phenomenal qualities by telling us what they *are*. (p.45) Hereafter, I will only use the term "representationalism" to mean strong representationalism, unless otherwise noted.

In this dissertation, I will examine the (strong) representationalist approach to sensory phenomenology. Importantly however, this is not to say that I will attempt to argue for the *truth* of representationalism. That is, I will not attempt to establish that phenomenal qualities are necessarily identical with representational contents. Rather, my argument can be understood as taking the form of a conditional: *if* one is going to endorse representationalism, then it must be of the general type that I describe. Thus, I shall implicitly assume the representationalist hypothesis as given in what follows.

Since representationalism about sensory phenomenology—hereafter, simply "representationalism"[3]—purports to explain the phenomenal quality of a sensory state in terms of the representational content of that state, the main issue for representationalism is obviously its *theory of content*. However, a representationalist theory also needs to explain how and why a sensory state comes to represent and have content in the first place. In other words, there are two parts to a representationalist theory: (i) an account of the nature of phenomenal content, and (ii) an account of sensory intentionality that explains how we are able to represent that content. Different accounts of these two parts of the theory will result in different types of representationalism, which in turn will generate different predictions about the phenomenology of a given sensory experience.

But perhaps the most popular version(s) of representationalism attempts to provide *reductive* theories of phenomenal content (described below). Indeed, one of the main motivations for reductive representationalism is that it appears to offer a *naturalistically acceptable* picture of sensory phenomenology; that is, an account that is compatible with both physicalism and our general scientific image of the world. This makes the theory particularly attractive because the qualitative phenomenology of conscious experience is

[3] It is important to distinguish representationalism *about sensory phenomenology* from the sort of "representationalism" about mental content in general, such as that endorsed by, e.g., Fodor (1975, 1987, 1990).

notoriously thought to pose a special challenge to physicalism and/or naturalism, since it is unclear how phenomenal qualities can be integrated into the empirical framework of cognitive sciences such as neuroscience and psychology. This view has been articulated in many different ways, from Nagel's (1974) claim that scientific accounts of conscious experience necessarily 'leave something out', to Levine's (1983) description of the "Explanatory Gap" to Chalmers (1996) identification of the "Hard Problem" of consciousness. And while these various formulations of the problem differ with respect to the exact nature of the challenge, they all point to the need for a theory of sensory phenomenology that is explicable in reductive/naturalistic/scientific terms.[4] And this is precisely what reductive representationalism purports to offer.

Reductive representationalism is motivated by the long-held view that it is possible to naturalize and/or reduce intentionality. In fact, there is a long history in contemporary philosophy of trying to provide a reductive/naturalistic account of intentionality, and of *sensory* intentionality in particular.[5] And by supplementing a reductive theory of sensory intentionality with a reductive theory of sensory *content*, one can provide a fully reductive representationalist account of sensory phenomenology.

Importantly however, by attempting to provide a reductive theory of sensory intentionality and phenomenal content, representationalism traditionally ends up being committed to *phenomenal externalism*.

---

[4] To be clear, not everybody takes the existence of the "Hard Problem" or "Explanatory Gap" to indicate the need for a reductive theory of sensory phenomenology. Indeed, many philosophers have instead taken these arguments to show the *impossibility* of such a theory, and thus have opted to endorse various types of *non-reductive* theories of sensory phenomenology, including non-reductive versions of representationalism. (Indeed, Chalmers famously introduced the "Hard Problem" in 1996 specifically to argue for the unavoidable necessity of non-reductive explanations of phenomenology. Moreover, his 2004 version of representationalism is explicitly non-reductive.)

[5] For example, the computational eliminativism of the Churchlands (1989, 1995), the interpretivism of Dennett's "intentional stance" (1987), the information-theoretic approach of Dretske (1981, 1988), and the asymmetric-dependency theory of Fodor (1987, 1990) are all attempts to provide naturalistic accounts of intentionality. And all use *sensory* intentionality as a starting point. (See Akins 1996 for a critical overview.)

1.2 Externalist vs. Internalist Representationalism

The vast majority of representationalist theories—and those that are my primary target here—are *externalist* about sensory content,[6] holding that the representational content of sensory states is 'wide' in the sense that it is determined by factors that are external to the subject. On this view, molecular duplicates can be in exactly the same physiochemical state yet have different phenomenal experiences if their environments or histories relevantly differ. Externalist representationalism is thus a form of *phenomenal externalism*, the claim that, as Tye (2003) puts it, "phenomenology ain't in the head". (p.12) Phenomenal externalism stands in contrast to what Dretske (1995) calls "the Internalist Intuition"; the idea that the phenomenology of perceptual experience is fixed by internal properties (e.g., brain activity) alone.

Representationalism typically manifests itself in externalist form because of an underlying commitment to two claims about the nature of sensory intentionality and sensory content—i.e., the two parts of a representationalist theory as described above. Specifically, externalist representationalism is committed to: (i) *The Tracking Thesis* (a claim about how sensory states come to represent properties in the external environment), and (ii) *Russellianism* (a claim about the nature of the representational contents of sensory states). In what follows, I will argue that both claims are fundamentally incorrect, in the sense that they are incompatible with representationalism about sensory phenomenology. That is, I will argue that sensory intentionality is not based on 'Tracking' and that the phenomenal content of sensory representations is not Russellian.

Importantly however, The Tracking Thesis and Russellianism by themselves do not automatically entail phenomenal externalism. Indeed, many internalist accounts of representation are also based on Tracking and Russellianism.[7] Rather, as I will

[6] E.g., Byrne (2001), Dretske (1995), Hilbert (2003), Harman (1990), Jackson (2007), Lycan (1996), Tye (1995, 2000, 2003, 2007, 2009), etc.
[7] Examples of internalist representationalist views based on Tracking and Russellianism include Shoemaker (1994, 2000, 2001, 2003), Levine (2003), and Boghossian and Velleman (1989). Examples of internalist representationalist views that reject either Tracking and/or Russellianism include Rey (1998), Chalmers (2004), and Thompson (2009).

demonstrate below, it is Tracking and Russellianism, *together with a commitment to reductionism and/or naturalism* that leads to phenomenal externalism.

This situation has convinced many representationalists who are skeptical of phenomenal externalism that the proper course of action is to give up on reductionism. That is, in order to avoid phenomenal externalism—which many are convinced that we should—philosophers have turned to *nonreductive* forms of representationalism. In other words, the philosophical consensus seems to be that representationalists can't "have their cake and eat it too": representationalism can either be *reductive & externalist* or *non-reductive & internalist*, but not *reductive & internalist*.[8] However, I think that we can have the best of both worlds. We can have our cake and eat it too: a representationalist theory that is both reductive and internalist.

In this dissertation, my goal is to examine the prospects for *reductive* versions of representationalism. Thus, I will not directly address arguments for or against non-reductive versions of the theory.

In chapter 2 of my dissertation, I argue that externalist representationalism is false. Specifically, I argue that both Tracking and Russellianism are incompatible with representationalism about sensory phenomenology because both are vulnerable to two different sides of the same problem: *bad structural correlation* (i.e., the bad correlation between the structure of phenomenal qualities and the structure of physical properties "tracked" by our sensory systems). In part one, I present the central problem for physicalist (i.e., reductive) Russellianism about sensory phenomenology: that the physical properties that are constituitive of the Russellian content of our sensory representations are not sufficiently *response-independent* to provide a satisfactory explanation for the structure of phenomenal qualities. In part two, I present the central problem for the 'Tracking Thesis': that it delivers the wrong verdict about the

[8] Likewise, Chalmers (2004) describes what he calls the "inconsistent triad": (i) internalism about phenomenal qualities, (ii) representationalism, and (iii) externalism about content. Most reprensentationalists reject (i), whereas I reject (iii), at least for phenomenal content.

phenomenology of certain sensory experiences, such that the representational predictions do not match the empirical (behavioral) ones.

In chapter 3 of my dissertation, I develop the outlines of a reductive internalist version of representationalism, one that utilizes the empirical framework developed by psychophysics and neuroscience to flesh out a philosophical theory of content. In contrast to the Russellian theory of phenomenal content held by externalist representationalism, this approach characterizes sensory phenomenology as involving a kind of *Fregean* content, and phenomenal qualities as *modes of presentation* of sensory representations. Furthermore, in chapter 4 I demonstrate how this mode of presentation can be empirically characterized in terms of a location in a psychophysically-defined *quality space,* which can in turn be given a neurophysiological interpretation. Next, in contrast to the 'Tracking' view of sensory intentionality held by externalist representationalists, I argue that the neural mechanisms responsible for the representation of spatial locations and visual objects (*feature maps* and *visual indexes*, respectively) serve a referential role in sensory psychosemantics that grounds the intentionality of sensory states. In chapter 5 I provide a detailed analysis of the neural mechanisms that underlie the ability to sensorily discriminate and represent spatial locations, and demonstrate how the spatial content of sensory experience is based on a type of multimodal sensorimotor coordination of different types of spatial information. Finally, in chapter 6 I examine the so-called "binding problem" in order to demonstrate how the modes of presentation described in chapters 3 and 4 must be functionally and semantically linked with the referential pointers to produce sensory representations whose content corresponds to our qualitative phenomenal experiences.

I describe the resulting view as *methodological representationalism*: a theory of sensory phenomenology which takes the (strong) representationalist hypothesis as a background assumption, and attempts to use the general empirical methodology of neuroscience and psychology to provide a philosophical theory of content. Indeed, as noted above, describing sensory phenomenology in representational terms is ubiquitous in psychology and neuroscience. However, despite this implicit commitment to (at least a minimal form of) representationalism, neuroscientists and psychologists typically lack even a cursory psychosematic theory of sensory representation upon which to ground their claims. Thus, methodological representationalism can be understood as attempting to provide a philosophical theory of what cognitive scientists actually (or *should*) mean when they characterize sensory phenomenology in representational or informational terms. Alternatively, it can be understood as an attempt to demonstrate how a particular philosophical theory of sensory phenomenology (representationalism) can be integrated into the empirical framework of cognitive science. In either case, the goal is the same: an explanatory psychosemantic framework for sensory phenomenology that is valuable to both philosophers and cognitive scientists.

In the following chapters, I develop this view in greater detail. However, before I can do so, there are two additional issues that need to be addressed. First, it is important to clarify the precise nature of the explanandum. As described above, representationalism offers the possibility of a reductive explanation of sensory *phenomenology*. However, the meanings of terms like “phenomenology”, “phenomenal”, “experience”, “qualitative”, “what it’s like”, and their ilk can be ambiguous and even confusing due to multiple conflicting interpretations in the literature. Thus, it is extremely important to be clear about the target of explanation, as different versions of representationalism purport to explain different things. Second, as I stated above, my concern in this dissertation is solely with *reductive* forms of representationalism; i.e., those that purport to offer naturalistically acceptable accounts of phenomenology that are compatible with and grounded in empirical science. Accordingly, it is necessary to provide some criterion that any given version of representationalism must meet if it is to be considered reductive. In the following two sections, I address these two issues in turn.

1.3 What is “Phenomenal”?

In this dissertation, I attempt to provide a representationalist theory of phenomenal *content*. However, I do not claim to provide a theory of phenomenal *consciousness*. Yet representationalism is often explicitly presented as a theory of phenomenal consciousness. So what gives?

Following Lycan (2000), one can distinguish between two senses of "phenomenal consciousness". In what Lycan calls the "first-order" sense, the term means something like "Being in a sensory state that has a distinctive qualitative or phenomenal property, such as the color one experiences in having a visual experience, or the timbre of a heard sound". Importantly, the "qualitative or phenomenal property" referred to in this quotation "presents itself as part of the world, not as part of one's mind. It is, e.g., the apparent color of an apparently physical object". (ibid) In other words, these properties are colours, sounds, tastes, smells, and other perceptible properties that we attribute to things in the external world. In chapter 4, I will use the term "appearance property" to denote these types of qualitative features. On the representationalist view, appearance properties are intentional objects—representata. They are the referents of our sensory representations. According to representationalism, sensorily representing an appearance property is sufficient for a qualitative phenomenal experience of that property—in other words, sufficient for the instantiation of a phenomenal quality.[9]

Lycan's second sense of "phenomenal consciousness"—which he calls the "higher-order" sense—describes the phenomenology of the *conscious awareness* of the being in a particular sensory state. Understood this way, the term "phenomenology" refers to a higher-order property; i.e., the property of "what it's like" to consciously experience a first-order sensory state. If one understands conscious awareness in representationalist terms, then the first-order states themselves are the intentional objects of higher-order representational states. This view is known as *higher-order representationalism* ("HOR"), which stands in contrast to *first-order representationalism* ("FOR") of the kind described above.

Thus, a subject could have a phenomenally conscious experience in the first-order sense (such that it possesses a qualitative phenomenology) while nevertheless being

[9]Importantly, the relationship between appearance properties and phenomenal qualities varies according to the theory of phenomenal content on offer. According to Russellian representationalism, appearance properties are constitutive of the phenomenal content of a sensory state, and thus are *identical with* phenomenal qualities. In contrast, according to Fregean representationalism, the phenomenal content of a sensory state is identified with modes of presentation of those appearance properties—which I refer to as phenomenal qualities. I will elaborate on both these views in chapters 2 and 3, respectively.

phenomenally *un*conscious and unaware of that experience in the higher-order sense (such that there is nothing "that it's like" for the subject to have that experience).[10] Bynre (2004) refers to this distinction as being between "experiences" and "non-conscious experiences". Similarily, Carruthers (2000) refers to the distinction as the difference between "worldly" and "experiential" subjectivity, and Rosenthal (2002) describes it a distinction between "thin" and "thick" phenomenality.

Given this proliferation of conflicting interpretations of the relevant terminology, I propose that the clearest, most unambiguous way to refer to this distinction is simply as between "phenomenal" and "conscious". "Phenomenal consciousness", then, is simply consciousness of a phenomenal state, and "phenomenally conscious" simply refers to a phenomenal state that is conscious. The fact that a particular state is phenomenal does not entail that it is conscious, nor does the fact that a particular state is conscious entail that it is phenomenal.

On this terminology, proponents of FOR and HOR disagree about whether or not the instantiation of the first-order representational state is sufficient for phenomenal consciousness (FOR), or merely phenomenality (HOR). However, in this dissertation I will remain neutral on this question. That is, as stated above, my project is to provide a theory of phenomenal *content*, not phenomenal *consciousness*.

Thus, I argue that we should understand FOR as an attempt to explain what it is for a sensory state to be "phenomenal" (or to have a certain "phenomenology" or to possess certain "phenomenal qualities") in the first-order sense of possessing a qualitative character, whereas HOR attempts to explain what it is for a state to be "conscious" in the higher-order sense of experiential awareness of "what it's like". In both cases, the target of explanation—phenomenology and consciousness, respectively—are purportedly explained by positing a special role for representation or intentionality; specifically, in

---

[10] The classic example of this sort of thing comes from Armstrong (1981) in his discussion of the long-distance truck driver who, after driving for long periods of time, suddenly 'comes to' and realizes that for some time he has been driving without being *consciously aware* of his surrounding or of what he has been doing. In such situations, it seems implausible to deny that the driver is undergoing sensory experiences of the road (given his successful driving), despite his lack of conscious awareness.

both cases the *content* of the representation is identified with either the phenomenal character (in the case of FOR) or the conscious experiential awareness of the phenomenal character (in the case of HOR).

In this dissertation, I present a *first-order* representationalist account of sensory phenomenology.[11] Equally importantly however, I do not claim to provide an account of consciousness, nor of what makes a given phenomenal state a phenomenally conscious state. Furthermore, although I argue that phenomenology is explained by intentionality/representation, I do not address the question of whether consciousness is also explained in those terms. In other words, my account is in no way committed to HOR as a theory of consciousness, and is indeed compatible with a non-representationalist view of consciousness such as, e.g., global workspace theory.

1.4 Reductive Representationalism

A useful way to characterize reductive representationalism comes from Chalmers (2004), who argues that representationalism is reductive to the extent that it can explicate the notion of a phenomenal quality in non-phenomenal terms. As described above, there are two parts to a representationalist theory: (i) a theory of sensory intentionality, and (ii) a theory of sensory content. Thus, in order to be reductive, both parts of the theory must be fully explicable in non-phenomenal (i.e., naturalistically acceptable) terms. If, on the other hand, either of them can only be understood in phenomenal terms, then the account is non-reductive.[12] So a version of representationalism is not necessarily reductive simply because it provides a reductive account of sensory intentionality/representation, for it may nevertheless hold that phenomenal content is only specifiable in phenomenal terms.

[11] Hereafter, I will use the term "representationalism" to mean first-order representationalism, unless otherwise stated.
[12] Chalmers (2004) adds a third part of a representationalist theory: a *manner of representation* that is supposed to distinguish phenomenal from non-phenomenal representation. In essence, the idea is that it's plausible that we can represent the very same content both phenomenally and non-phenomenally, and thus we need to provide an account of what distinguishes the two by appealing to a 'manner' of representation. For my purposes here, the 'manner' can be subsumed as a part of the theory of sensory representation, such that if such a manner exists, it must be specifiable in non-phenomenal terms in order for the resulting representationalist theory to be genuinely reductive. (For example, Tye (1995, 2000) appeals to the "poised" criterion, which is essentially a certain functional role.)

Nor is it fully reductive if it provides a reductive account of phenomenal content while nevertheless holding that sensory intentionality is only specifiable in phenomenal terms.

Chalmers' definition of reduction with regards to representationalism helpfully avoids the usual confusion associated with the concept of reduction—theory reduction, ontological reduction, Nagelian bridge law reduction, etc.—and opts for a clear criterion of reducibility; namely, the ability to be fully explicated in non-phenomenal terms. However, although Chalmers' definition provides a good starting point, I argue that being specifiable in non-phenomenal terms is not sufficient for a representationalist theory to be genuinely reductive. In addition, the relevant specification must also be *non-trivial* and *non-circular*. (That is, I will argue that Tye et al.'s theories are not reductive because they fail on just this criterion.)

Finally, it is important to distinguish between *metaphysically* reductive and *epistemically* reductive versions of the theory. To understand the difference, consider whether the account is supposed to be able to close the above-mentioned "Explanatory Gap". If so, then the account is both epistemically and metaphysically reductive, since epistemic reduction entails metaphysical reduction, but not vice-versa. Put differently, the issue of whether or not a representationalist theory is epistemically reductive depends on whether the identification of phenomenal qualities with representational content is supposed to be a *conceptual* (a priori) or *empirical* (a posteriori) truth. The latter is only committed to metaphysical reduction, whereas the former entails the stronger position of epistemic reductivism. The vast majority of representationalist theories—including my own—are only metaphysically (but not epistemically) reductive.

So, with a *prima facie* concept of reduction in hand, I now turn to examine how reductive representationalism actually works, specifically, *externalist* reductive representationalism.

Chapter 2

2 Externalist Representationalism

In this chapter, I argue that both the Tracking Thesis and Russellianism are incompatible with representationalism about phenomenal qualities, because both are vulnerable to two different sides of the same problem: *bad structural correlation* (i.e., the bad correlation between the structure of phenomenal qualities and the structure of physical properties "tracked" by our sensory systems). In section 2.2 I describe Russellianism and the different varieties of Russellian representationalism. I then present the central problem for physicalist (i.e., reductive) Russellianism about phenomenal qualities: that the physical properties that are identified as being constituitive of the Russellian content of our sensory representations are not sufficiently *response-independent* to provide an explanation for the structure of phenomenal qualities. In section 2.3 I describe the 'Tracking Thesis' and its teleological views of sensory intentionality. I then present the central problem for tracking: that it delivers the wrong verdict about the phenomenology of certain sensory experiences, such that the representational predictions don't match the empirical ones. I also consider the externalist representationalist's response (the *Unnatural Relation R** account), and show why it fails.

2.1 Tracking, Russellianism, and Phenomenal Externalism

As defined here, reductive externalist representationalism is committed to the following claims about sensory representation:

1) *"The Tracking Thesis":* The representational contents of sensory states are fixed by certain causal-informational ("tracking") relations that hold between brain states and physical properties in the external world, and the existence of these tracking relations grounds the intentionality of sensory states.

These tracking relationships typically appeal to something like reliable causal covariance under certain conditions. For example, Dretske (1995) holds that a sensory state represents something when it has the biological function of "indicating" that thing, and

Tye (1995, 2000) defines representation in terms of counterfactual-supporting causal covariance under “optimal conditions”.

2) *“Russellianism”:* The representational contents of sensory states are *singular*, in the sense that they include the referent of the representation as a proper constituent.

In other words, the external properties tracked by our sensory systems are literally constituents in the representational content of that state. According to the reductive accounts under consideration, these properties are identified with physical properties of external objects – e.g., surface reflectance properties (colour vision), chemical properties (smell and taste), mean molecular kinetic energy (thermoreception), tissue damage and bodily disturbances (pain), etc. I will refer to this view as *physicalist Russellianism*.

The conjunction of these two claims (The Tracking Thesis & physicalist Russellianism) entail:

3) *“Phenomenal Externalism”:* The phenomenal quality of a sensory experience is identical with the external physical property that is tracked by that experience.[13]

That is, assuming that—*ex hypothesi*—the phenomenal quality of a sensory state is a matter of the representational content of that state, then since (i) the content of a sensory state is fixed by the particular tracking relations that it bears to external physical properties, and (ii) that content is Russellian, such that it is literally composed of the external physical properties tracked by the state, it follows that (iii) those external physical properties themselves fully determine what it is like for us to experience them.

I argue that all three of the above claims are false. In what follows, I present arguments against both physicalist Russellianism and the Tracking Thesis, in that order.

---

[13] It is important to note that, as defined here, phenomenal externalism amounts to more than the basic claim that the phenomenology of sensory experience is determined by factors that are external to the subject. Rather, phenomenal externalism should be understood here in specifically *representationalist* terms; i.e., as the more specific claim that phenomenal qualities are represented external physical properties.

2.2 Russellianism

> *"Russellianism":* The representational contents of sensory states are *singular,* in the sense that they include the referent of the representation as a proper constituent.

One way to characterize the content of mental states is in terms of the object or property which that state is about – its *extension* or *referent*. For example, Barack Obama is the referent of the term "Barack Obama" and the property of being a Democrat (whatever that may be) is the referent of the term "Democrat". The Russellian view of representational content holds that the content of these terms include their referent as a proper constituent. Similarly, the proposition "Barack Obama is a Democrat" contains both Barack Obama himself as well as the property of being a Democrat as proper constituents of its content.

This approach seems natural to apply to sensory states. Sensory states have referents—objects and properties which the states are 'about'—and one way of characterizing the representational contents of such states is in terms of their referents. For example, a sensory state may represent a red cube, and it seems quite plausible to characterize the content of the state as including both the cube itself as well its colour as proper constituents.

Importantly, Russellian representationalism is not merely committed to the (weaker) claim that sensory states have Russellian content; in addition, it is committed to the further (stronger) claim that a state's Russellian content *exhausts* its phenomenal character.

However, Russellian representationalists (such as Tye 1995, 2000, 2003, 2009; Dretske 1993, 1995, 1996; Byrne, 2001) also typically hold that phenomenal content is not object-involving, but is rather merely property-involving. (Put differently, phenomenal content is a type of *abstract* content, as opposed to *concrete* content.) This is meant to account for the possibility of phenomenally identical experiences that are about numerically distinct objects. For example, one might have an experience of a red square at T_1, then a phenomenally identical experience of a different red square at T_2. (Alternatively, the experience at T_2 might be a phenomenally identical hallucination.) Despite the fact that these two experiences involve numerically distinct objects, it nevertheless would seem that both experiences have the same representational content – namely,the content 'RED SQUARE'.[14]

There are two important points to note about this claim. First, this caveat is meant to accommodate a persistent intuition – that phenomenally identical subjects share representational contents based on the sameness of phenomenology alone. It is this intuition that drives the current 'phenomenal intentionality' research program.[15] However, externalist representationalism is typically committed to limiting the scope of this claim (same phenomenology → same content) such that it's only applicable to members of the same species (for reasons discussed below). Second, it seems prima facie plausible that the argument against object-involving contents could perhaps apply to property-involving contents too -- e.g., if the possibility of phenomenally indistinguishable hallucination makes it plausible that phenomenal content is not object-involving, it might also seems to suggest that it's not property-involving either. That is, it seems plausible that I can have two phenomenally indistinguishable experiences involving different properties – i.e., metamers.[16] Of course, the externalist/Russellian will argue that there's only one property that's being represented in both cases (e..g, "redness") that is multiply realizable.

In any case, according to Russellian representationalism, phenomenal qualities are represented properties of external objects. So the question naturally arises: *what kinds of*

---

[14] Note that in the hallucinated case, the experience includes a *non-instantiated* property. Put differently, despite the fact that there is no property that causes the experience, the experiences nevertheless includes the non-instantiated property that *would* normally cause that experience.

[15] The "Phenomenal Intentionality Research Program" (or PIRP) holds that mental states have intentional directedness and possess representational content purely in virtue of possessing a phenomenal character. In addition, most versions of PIRP also make the stronger claim that all non-phenomenal intentionality is ultimately derived from phenomenal intentionality. PIRP thus reverses the direction of explanation posited by representationalism: whereas the latter holds that intentionality explains phenomenology, the former holds that phenomenology explains intentionality.

[16] Metamers—discussed in detail below in section 2.2.1—are pairs of chromatic stimuli that appear to be phenomenally identical in colour despite the fact that they actually reflect very different wavelengths of light.

properties are these? There are a number of answers to this question. Chalmers (2004) distinguishes between what he calls “physicalist”, “projectivist”, “primitivist”, and “dispositional” forms of Russellian representationalism. Importantly, only the physicalist version of the theory is externalist. The projectivist, primitivist, and dispositional forms of Russellian representationalism are all internalist. However, it is the physicalist version of representationalism that is most popular, especially for those attracted to a *reductive* version of representationalism. (It should be noted, however, that Shoemaker’s (1994, 2003) dispositionalist version of representationalism is also often held to be reductive. And while there is a sense in which this is true, his account nevertheless fails to provide an independent explanation for the structure of phenomenal qualities in the way that I will describe below. Thus, I will set it to the side in my examination of reductive representationalism.) Projectivist and primitivist versions of Russellian representationalism are non-reductive.

TYPES OF RUSSELLIAN REPRESENTATIONALISM	REDUCTIVE	NON-REDUCTIVE
INTERNALIST	Dispositionalist [17]	Projectivist Primitivist
EXTERNALIST	Physicalist	n/a

Since my focus in this dissertation is on the prospects of *reductive* representationalism, and my focus in this chapter is on arguing against externalist representationalism, I will limit my comments here to the physicalist version of Russellianism.

2.2.1 Physicalist Russellianism

According to the physicalist version of Russellian representationalism, the properties that figure in the representational content of sensory experiences are identical with physical

[17] As noted above, although Shoemaker claims that his version of dispositionalist representationalism is reductive, Levine (2003) argues—correctly, in my opinion—that Shoemaker’s version of representationalism fails to be reductive on the grounds that the phenomenal quality of an experience is explained in terms of the representation of a particular property, which in turn is individuated by the disposition to cause an experience with a particular phenomenal quality.

properties of external objects – e.g., surface reflectance properties (colour vision), chemical properties (smell and taste), mean molecular kinetic energy (thermoreception), tissue damage and bodily disturbances (pain), patterns of vibration in the air or other surrounding medium (aural properties), etc. This approach provides a straightforward way for representationalism to reduce phenomenal content. By identifying phenomenal qualities with represented physical properties, one can provide an account of the representational content of sensory states that can be fully explicated in non-phenomenal terms. However, physicalist Russellianism faces a powerful objection: namely, it seems to conflict with decades of findings in psychophysics. Specifically, it fails to provide an independent explanation of the *relational structure* of phenomenal qualities (described below), where an "independent explanation" is one that does not make essential reference to perceivers.

In any sensory modality (or sub-modality), phenomenal qualities stand in certain structural relations such as *relative similarity and difference*. (For example, red is more similar to orange than it is to blue.) Assuming physicalist Russellianism, it follows that relations of similarity and difference between phenomenal qualities in any given modality should match (and in fact, be *explained by*) relations of similarity and difference between the external physical properties represented by that sensory modality.

For example, physicalist Russellianism should predict that structure of our temperature sensations roughly matches the structure of temperature states in the world—i.e., the higher the external temperature, the more intense our sensation of warmth. Likewise, since red is more similar to orange than to blue, externalist representationalism should predict that the surface reflectance properties that evoke sensations of red are more similar to those surface reflectance properties that evoke sensations of orange than they are to the surface reflectance properties that evoke sensations of blue.

However, this is precisely the *opposite* of what is found in psychophysics: no physical property has ever been discovered that can account for the fact that red surfaces are more similar to orange ones than blue ones. Moreover, this phenomenon occurs in every sensory modality—the structural relations between phenomenal qualities in a given modality do not match the structure of the domain of physical properties represented by the modality.

This problem was first highlighted (in contemporary philosophy) by Larry Hardin (1988), who described the problem of *metamerism* in colour perception. Metamers are types of stimuli with distinctly different spectral power distributions[18] that nevertheless are indistinguishable for a given observer under a specific illuminant.[19] Metamerism occurs because distinctly different combinations of wavelengths can produce the same receptoral response. The reason for this has to do with the details of the receptors in the retina.

There are three types of retinal receptors sensitive to chromatic stimuli, typically known as the long- (L), medium- (M), and short- (S) –wavelength cone cells. Although each cell type is maximally responsive to a particular wavelength, they are also sensitive to a broad range of wavelengths, with considerable overlap between the sensitivity ranges of each cell type. The cell's response to incoming light depends on the total cumulative energy absorbed across all wavelengths to which it is sensitive. Furthermore, a cell's response also depends on both the wavelength of the light and its intensity. Thus, e.g., a cell might produce an equal response to 1 quanta of monochromatic light at wavelength *x* and 2 quanta of monochromatic light at wavelength *y*. As a result, light composed of very different combinations of wavelengths and intensities can produce identical responses in a cone cell.

Moreover, the same basic problem occurs in nearly every sensory modality: for example, Pautz (2013) shows how "a variety of different combinations of molecularly different odorants, and different levels of concentration, can produce the very same smell experience." Similar considerations apply to taste: substances which are very different in their chemical structure can produce identical taste experiences. Again, in both cases the explanation for the phenomena has to do with the limited number of ways in which

[18] A spectral power distribution is the total amount of light that is reflected or emitted at each wavelength in the visible spectrum

[19] Metamers are always relative to some illuminant; i.e., two objects are metameric matches with respect to some illuminant. (Moreover, they are also relative to a particular perceiver, a viewing angle, and other such factors (see below).)

chemoreceptors found on the tongue and in the nose can respond to stimuli. Likewise, Akins (1996) has demonstrated how and why a kind of "metamerism" can occur in temperature perception (i.e., how different external temperatures can give rise to identical sensations of warmth and/or coolness).[20,21]

However, it is very important to note that the problem of bad structural correlation goes beyond mere metamerism. As Pautz (2013) points out, the externalist representationalist can avoid the problem of metamerism by simply claiming that the experience represents a *disjunction* of different surface reflectance types (e.g., Byrne & Hilbert, 2003). The problem of bad structural correlation is the problem of explaining the relations of similarity and differences holding between those disjunctions.[22]

Of course, the lack of isomorphism between the structure of phenomenal qualities and the structure of physical properties has been recognized for some time by representationalists who are committed to physicalist Russellianism. Faced with this problem, physicalist Russellians attempt to explain the bad structural correlation between the phenomenal and the physical domains by claiming that the external properties stand in the relevant relations of similarity and difference *by virtue of the way that they affect our sensory systems*. In other words, they explain the similarity and difference holding between the disjunctions in terms of the similarity and difference in evoked receptoral response (under optimal conditions). For example, Byrne & Hilbert (2003) and Tye (2000) claim that a

---

[20] Briefly, the explanation for "metamerism" in temperature perception is based on the fact that certain types of thermoreceptors have a response that is *context-dependent*, based on the initial temperature of the skin. Thus, for receptors that are sensitive to warmth, the intensity of the response to warm stimuli is dependent on the starting temperature, such that higher starting temperatures lead to greater receptor response and thus a sensation of greater warmth. Likewise, receptors that are sensitive to cold have the opposite response: the colder the initial conditions of the skin, the more drastic the receptor response to cool stimuli, and thus the stimuli is felt to be colder. This effect can be keenly felt in the "Tepid Water Illusion": let one hand soak in a bowl of warm water, let the other hand soak in a bowl of cold water. After a few minutes, place them both in a bowl of tepid water. The water will feel to be two different temperatures to the different hands.

[21] Interestingly however, Clark (1992) argues that "It is practically impossible to find the auditory equivalent of metamers. Almost any difference in the pitch-loudness spectrogram of two sounds leads to a discriminable difference between them." (p.138)

[22] Put slightly differently, the problem of bad structural correlation is the problem of explaining how and why a particular surface reflectance type is categorized as belonging to one disjunction rather than another one. In what follows below, I argue that Russellian representationalists cannot provide an answer to this question that does not make essential reference to perceivers.

surface is red iff it reflects light that would evoke a specific neural response in a 'normal' human visual system under certain conditions. Similar proposals apply to the other sense modalities as well: for example, given the bad correlation between the structure of taste and smell qualities and the structure of their chemical causes, the physicalist Russellian will individuate the properties by reference to their physical effects on our sensory systems.

However, despite the fact that the disjunctive properties identified by physicalist Russellians are individuated by reference to their effects on our sensory systems, it's important to note that these are *not* supposed to be dispositional properties, nor are they supposed to be response- or perceiver- dependent in any sense. Rather, they are characterized as response-independent physical properties of objects that do not depend in any way on the existence of perceivers.[23] As Byrne & Hilbert put it, "It is just a plain matter of fact that an object has a particular type of reflectance, and this fact need not depend in any interesting way on the existence of creatures with color vision." (2003, p.11)

In other words, the idea is that even if we need to make reference to neural responses in order to pick out the relevant structural relations between properties (for epistemic reasons), the properties themselves nevertheless have that relational structure independently of any such reference.[24] For example, suppose that R is the following relative similarity relation that holds between physical properties x, y, and z: if $R(xyz)$,

---

[23] The problem with specifying the external properties solely in terms of their effects on perceivers, of course, is that those properties would no longer be explicable in non-phenomenal terms. (Indeed, the physical Russellian appeal to evoked receptoral response is only due to those responses being correlated with particular phenomenal qualities.) Thus, one might wonder if *zombies* 'count' as perceivers on this view. For if they do, it seems possible that one could perhaps reductively specify the relevant external properties in non-phenomenal terms by making some reference to their receptoral effects on "zombie perceivers". However, such a move would only push the problem back a step: the problem then would be to explain how and why zombie perceivers' receptoral responses are correlated with *our* phenomenal qualities that are the target of explanation. What's more, even if such an explanation were available, the properties in question would still be relational or dispositional at best (regardless of whether they were specified in terms of zombie perceivers or machines built to replicate human sensory systems) and thus would fail by the physicalist Russellian's own criterion.

[24] So we might say that while there is an *epistemic* sense in which the properties are response-dependent (because we need to make reference to neural responses in order to individuate them), metaphysically speaking they are nevertheless response-independent.

then it is the case that *x* causes a receptoral response that is more similar to the receptoral response caused by *y* than the one caused by *z* (in normal human perceivers under optimal conditions). According to physicalist Russellianism, there exists another, distinct relation ***R**** that orders *x*, *y*, and *z* in exactly the same way but does not make essential reference to perceivers. Of course, science has discovered no such relation. However, if one conceives of the metaphysics of relations in a liberal enough way, than such a relation certainly exists.

Nevertheless, I think that there are good reasons to be skeptical of physicalist Russellianism. For one, on this proposal, the claim that the structure of phenomenal qualities is explained by the structure of the external properties that they represent seems *vacuous* (or at least *trivial*). As Akins (1996) points out, on this view one can no longer meaningfully ask whether sensory systems accurately reflect the structure of the domain of external properties to which they respond, since on this view they in fact *define* that domain.

Furthermore, Pautz (2013) presents two additional arguments against this view (which he calls the "unnatural relation account"): the 'Metasemantic' argument, and the 'Radical Indeterminacy of Judgements' argument. Basically, the idea common to both is that, in order to be true, judgements or statements about the relative similarity of stimuli must refer to the response-independent relations *R**. For example, on this view, the (true) statement or judgment that red is more similar to orange than blue refers to a perceiver-independent relation that holds between types of surface reflectances and that does not make essential reference to perceivers (i.e., it does not include receptoral effects as relata).

The 'Radical Indeterminacy' problem starts from the premise that, given the liberal conception of the metaphysics of relations required to posit the existence of *R**, there are actually an *infinite* number of different similarity relations $R^*_1, R^*_2, R^*_3 \ldots R^*_n \ldots$, that have the same relational structure with respect to the physical properties to which our sensory systems are responsive but whose relational structure differs with respect to those physical properties to which they are not—e.g., electromagnetic radiation outside of the visible spectrum, chemical properties to which our olfactory systems are not sensitive (such as odorless gas), etc.[25] (This is supposed to be roughly analogous to Kripke's (1982) 'plus' and 'quus'.) If we grant the existence of these other relations, then it would seem that our statements about relative similarity are indeterminate with respect to which relation is being referred to.

The 'Metasemantic' problem is about how our statements manage to refer to R^* given the indeterminacy described above. (That is, how and why they are able to refer to R^* rather than R^*_1 or R^*_2 or R^*_n.) However, Pautz argues that even without the radical indeterminacy problem, it is still the case that there are no credible metasemantic accounts of how our statements and judgements could manage to refer to R^*, and furthermore, that the most plausible metasemantic account will have the result that statements about relative similarity refer not to relations between external physical properties, but rather relations between internal neural states.

Typically, metasemantic problems of this kind can be resolved by appeal to either some kind of causal factors or to speaker intentions. However, neither of these options are open to the physicalist Russellian. A causal account does not provide any grounds to justify the claim that R^* rather than R^*_n is the cause of our judgements of relative similarity, and it's unclear how speakers could possibly intend to refer to one of these relations to the exclusion of all others. Furthermore, physicalist Russellians seemingly cannot respond to these arguments by denying that those properties to which our sensory systems are not responsive can be relata in the R^* relation, for one could then legitimately ask how and why the R^* relation can be restricted to only those properties to which our sensory systems are sensitive without making some kind of essential reference to perceivers.

Thus, it seems that physicalist Russellianism—and thus, externalist representationalism—faces some very difficult (and potentially fatal) problems. What's more, even if the above objections *could* somehow be overcome, externalist representationalism faces an

[25] The objection, in other words, sees the global structure of similarity and difference relations to which our sensory systems are responsive as only a local part of a larger global structure of relations between physical properties.

additional fatal flaw related to the second part of the theory: the tracking account of sensory representation. And such an account is necessary in order to overcome the *other* half of bad structural correlation: the fact that we need to privilege some set of circumstances and observers.

Specifically, it is important to recognize that the problem of bad structural correlation runs deeper than has been so far described. For not only can stimuli that are physically very different appear indistinguishable**,** but furthermore, physically identical stimuli—or even the *very same* stimulus—can produce very different phenomenal experiences in different perceptual conditions (and to different observers). For example, in one experimental trial a colour stimulus might appear blue to a subject, while in the next trial that very same stimulus may appear green (to the same subject). Indeed, the quality that a colour stimulus presents is dependent on at least *six* distinct factors, including: (i) the physiological idiosyncrasies of the subject's visual system, (ii) the state of adaptation of their visual system, (iii) the ambient illumination, (iv) the surround, (v) the viewing angle, and (vi) the angular subtense. If any of these factors are changed, the colour quality that the stimulus presents can change as well. In order to overcome this problem, one needs an account of how and when sensory representations are *veridical*, and when they misrepresent. It is to this problem to which I now turn.

2.3 The Tracking Thesis

> *"The Tracking Thesis":* The representational contents of sensory states are fixed by certain causal-informational ("tracking") relations that hold between brain states and physical properties in the external world, and the existence of these tracking relations grounds the intentionality of sensory states.

It's important to note that the Tracking Thesis has two distinct (but related) parts, and that the tracking relation itself is supposed to do two things: First, it determines the representational contents of the sensory state. It does this by providing a set of *satisfaction conditions* for the representation, which have long been identified with representational contents. Second, it provides the *source of intentionality* for the sensory state: it explains why that sensory state is representational in the first place. (Namely, because it causally co-varies with external properties under optimal conditions.) Here, I am only concerned with the first part of the Tracking Thesis—i.e., the claim that the representational content of a sensory state is determined by the tracking relations that it bears to external properties. I will examine the second part—the claim that tracking grounds sensory intentionality—in chapter 5, in my discussion of neural intentionality (i.e., binding and spatial representation).

These tracking relationships typically invoke something like reliable causal covariance under certain conditions. For example, Tye (1995, 2000, 2009) defines representation as causal covariance under "optimal conditions".

> "for each sensory state, *S*, of a creature *c*, within the relevant set of alternative sensory states of *c*, we can define what *S* represents as follows:
>
> *S* represents that $P =_{df}$ If optimal conditions were to obtain, *S* would be tokened in *c* if and only if *P* were the case; moreover, in these circumstances, S would be tokened in c because *P* is the case." (2000, p.136)

Similarly, Dretske (1995) holds that a sensory state *S* represents some external object or property *P* when *S* carries information about *P* (when *S* "indicates" *P*, to use Dretske's terminology) and *S* has the systemic biological function[26] of providing the organism with information about *P*.

An important feature of the tracking account is that the representational contents of sensory states are determined not by the causal-informational covariance relations that *actually* obtain between sensory states and external properties, but rather, by the covariance relations that *would* obtain under certain counterfactually-specified, "optimal"

---

[26] Dretske (1995) contrasts *systemic* biological functions, which are phylogenetically determined, from *acquired* biological functions, which are ontogenetically determined. According to Dretske, the representational properties of perceptual experiences are a kind of systemic function, whereas the representational properties of thoughts and concepts are acquired.

conditions.[27] This counterfactually-based theory of content is necessary in order to account for the possibility of *misrepresentation*; i.e., the possibility that sensory state *S* might be tokened in creature *c* even in the absence of *P*s.[28] If *c* is perceiving under non-optimal conditions (due to either external interference or internal malfunction), then the co-variance relation between the external physical stimulus which causes a sensory experience and the representational content (and hence, the phenomenology) of that sensory experience no longer holds.

On this view, sensory systems work much like measuring instruments. For example, in a simple mercury thermometer there is a reliable causal co-variance relation between external temperature and the height of the mercury column that holds under certain conditions. This causal co-variance relation determines the representational content of any given state of the thermometer: e.g., the thermometer represents 20°C when the mercury column is at such-and-such a height, and it does so because that particular internal state of the thermometer causally covaries with an external temperature of 20°C when conditions are "optimal". Importantly, when those conditions fail to obtain—due to either "non-optimal" external conditions or to internal malfunction—the thermometer misrepresents. (For example, the reliability of the covariance relation might fail at radically different air pressures.)

From this perspective, the step to externalism about content is not difficult to appreciate. Identical internal states of a measuring instrument can have different representational contents when put to different uses. Consider a simple voltage meter. When connected to a microphone, the internal state of such a device might represent 60db. However, if the very same voltage meter is hooked up to the world differently—say, to an electric scale—then that very same internal state of the instrument might represent 6kg. Similarly, if that same voltage meter-electric scale device is selected for use in a different context—say, in

---

[27] Hereafter, I will use "optimal" as a general term to refer to whatever privileged set of counterfactually-specified conditions the causal co-variance relations are supposed to hold under, regardless of whether or not those conditions are explicitly teleological.

[28] It is important to recognize that providing an account of misrepresentation is absolutely indispensible to a theory or representation. Indeed, the very possibility of representation *at all* requires the possibility of misrepresenting.

an environment with lower gravity—the very same internal state might represent 10kg instead of 6kg. In short, physically identical internal states can have very different representational contents in different contexts and under different conditions.

Given that the representational content of a sensory state is determined by the features that it would causally covary with under a counterfactually-specified set of 'optimal conditions', the crucial question for tracking versions of representationalism concerns the way in which these conditions are determined. Typically, tracking representationalists hold that the 'optimality' of conditions is determined by teleology. Dretske (1995), Tye (1995, 2000, 2003), and Byrne & Hilbert (2003) all hold that the optimal conditions for evolved creatures are determined by their evolutionary history: i.e., some sensory state *S* represents some external property *P* for creature *c* because *S* was selected by evolutionary processes to be tokened in *c*s in the presence of *P*s.

However, consider the well-documented example of interpersonal variation in human colour vision. It's a well known fact that not everybody 'sees the same colours': subjects can vary by a wide margin in their perceptual discriminations of chromatic stimuli. Yet most evolutionary-teleological accounts of sensory representation are committed to the claim that there is a single 'veridical' perception of colour (for human observers), and that all variations from that norm are 'misperceiving'.

For example, consider Tye's (2002, p.451) description of spectrum inversion in conspecifics: Tom's colour spectrum is inverted relative to mine, such that when we are both viewing the same tomato under the same conditions, he has an experience that is qualitatively, phenomenally different than mine—I experience a reddish sensation and Tom experiences a greenish sensation. This leads to a dilemma for tracking representationalism: since (*ex hypothesi*) there is a difference in phenomenal quality of our respective experiences, the representationalist (of any kind!) must claim that there is a corresponding difference in representational content of the experience. However, since our respective sensory states are both tokened in response to the same external property,[29]

[29] On the physicalist Russellian account, a surface reflectance type.

the tracking account would seem to predict that our experiences have the *same* content. How can this be resolved?

The answer, according to Tye (and other externalist representationalists that are committed to a teleological view of sensory representation) is that Tom's experience *misrepresents* the colour of the tomato, because for Tom, optimal conditions do not obtain. Specifically, Tom's visual system is not functioning as it was designed to, since the external properties that Tom's sensory states would track, were they discharging their biological function properly, are not the external properties that they actually track. In other words, Tom has a certain type of evolutionary history, which plays a role in determining the representational content—and thus, the phenomenal quality—of his experiences. Because of this history, the operations of his visual system are held to the norms that apply to normal human visual systems. And in such systems, the internal state S was selected to covary with property P, whereas state S* covaries with P in Tom. Hence, optimal conditions do not obtain for Tom.

Thus, externalist representationalism is committed to denying the possibility of *phenomenal variation without misrepresentation*: i.e., the possibility that two conspecifics could be veridically representing the same property while having phenomenally different experiences.

The externalist representationalist claim about misrepresentation in conspecifics has two interesting consequences that should be noted: First, the externalist representationalist position entails that Tom's visual experience in fact represents—and on the Russellian view, literally *includes*—a property that is completely distinct from the one that is actually causing the experience. Put differently, on the physicalist Russellian view, despite the fact that surface reflectance property P causes S* to be tokened in Tom, Tom's experience does not include the surface reflectance property P as a proper constituent, but rather includes the *non-instantiated* surface reflectance property P*. Second, given the aforementioned variation in human colour vision, this result immediately raises the question of who perceives correctly. (Is it me, or is it Tom?) After all, externalist representationalism is committed to the claim that there is a single veridical perception of the colour of the stimulus, such that all other subjects are misperceiving. Importantly however, it is also committed to the claim that there is no empirical way of deciding who's visual system perceives "correctly" (nor which viewing conditions are 'optimal'[30]), and thus that it is *in principle* unknowable whether or not a stimulus is "really", e.g., unique green.[31] For example, Byrne & Hilbert are explicitly willing to accept the existence of "unknowable colour facts". (see e.g., Byrne & Hilbert, 2003, p.16-17n50)

However, I will set these two issues to the side for the moment in order to focus on constructing a counterexample to externalist representationalism. Specifically, as the 'Tom' example demonstrates, externalist representationalism is committed to denying phenomenal variation without misrepresentation, and therefore can be falsified by the existence of even a single case in which two qualitatively different experiences track—and veridically represent—the same external property under optimal conditions. Happily, such a counterexample is available, thanks to an argument from Adam Pautz (2006a, 2006b, 2013).

2.3.1 Pautz: Maxwell & Twin-Maxwell

Pautz (2006a, 2006b, 2013) has convincingly argued that externalist representationalism delivers the wrong verdict about phenomenal content in certain cases. Specifically, he shows it's possible for two individuals to be veridically 'tracking' the same external

---

[30] Consider the difficulty of specifying the correct circumstances for visual perception: as noted above in section 2.1, the "verdical perception" relation between a perceiver and a colour patch is at least an *six*-place one, holding between the stimulus, it's surround, the observer, the state of adaptation in the observer's visual system, the ambient illumination, the viewing angle, and the angular subtense. Change any of these relata and the phenomenal quality of the sensory experience can change too. The problem then, simply put, is that there is no non-arbitrary answer to the question, e.g., "do we perceive correctly under viewing angle *x* or viewing angle *y*?"

[31] Of course, the defender of externalist representationalism would dismiss this critique on the grounds that it is merely a form of *verificationism*—and that may well be true. However, I argue we are justified in being suspicious of any theory which makes predictions that are unknowable in principle. Second, and more importantly, I argue that a particular sort of verificationism is necessary when attempting to explain phenomenal qualities. Specifically, I argue that an explanation of phenomenal qualities requires that one adopt what Dennett (1991, 2005) calls a *heterophenomenological* approach—a third-person, scientific, empiricist approach to the study of phenomenology. This 'moderate' version of verificationism is not the overly strong kind endorsed by, say, the logical positivists, but rather is merely the kind required by a basic commitment to the empiricist foundations of scientific inquiry.

property (via sameness of sensory receptors), yet have different behaviour due to differences in their respective post-receptoral neural wiring. The correct verdict seems to be that they have *different* experiences (given that psychophysical experiments would show that, e.g., their *sorting* behaviour will be different) yet externalist representationalism should predict that they have *identical* experiences (because it seems to be committed to the claim that both experiences are veridically representing the same external property, and thus have the same content, and thus have the same phenomenal quality).

Pautz's argument is based on the *opponent process theory* of colour vision. According to (a very oversimplified version of) this theory, the outputs of the three different kinds of cone cells in the retina are summed and differenced to create three different post-receptor neural channels. The cells in these channels can assume different firing rates, either above or below their 'base' firing rate, and this difference can be referred to as the channel taking a 'positive' or 'negative' value. Two of the three opponent process channels encode chromatic information: one channel corresponds to a red-green opponency relation, whereas the other corresponds to a yellow-blue one. The third channel encodes achromatic information about lightness.

Opponent process theory explains a number of phenomena related to perception of colour, including the structure of perceptual discriminations (such as judgements of relative similarity, discriminability and matching, sorting behavior, colour categorization etc.), the unitary-binary distinction,[32] data from hue-cancellation experiments[33], facts about coloured afterimages, the Bezold–Brücke shift[34] (and the related Purkinje effect[35]),

[32] There are 4 "unitary" colours; that is, colours that are not a mixture of any more basic colours. These are red, green, yellow, and blue. All other colours are "binary", in that they are composed of a combination of two primary colours. (e.g., purple is a binary colour because it is a mixture of red and blue, which themselves are unitary colours that do not admit of further decomposition)

[33] Hue cancellation experiments use coloured lights to measure how much of an "opponent" colour is needed to remove any trace of one of the starting colour's components. For example, one might be presented with an orange coloured light, with the task of progressively increasing the amount of blue light until the target light is red (since the addition of blue light would 'cancel out' the yellowish hues in the original orange light.)

[34] The Bezold–Brücke shift is the phenomena whereby hue perception changes as a function of the intensity of light. Specifically, as the intensity of light increases, the perception of hue is shifted towards either blue

and many other phenomena.[36] In short, opponent process theory explains why subjects make the perceptual discriminations that they do. Thus, if one accepts what Pautz calls "The Experience-Behaviour Link", (which holds that identical phenomenal experiences give rise to identical behavioural dispositions), then it seems that opponent processing activity plays a direct role in determining the phenomenal quality of a colour experience.

However, As Byrne and Tye (2006) point out, there are 'weak' and 'strong' interpretations of the claim that opponent processing channels play a "direct" role in determining the phenomenal quality of a sensory experience. According to the strong version, opponent process states "metaphysically or nomologically necessitate" (p.4) the character of colour experience. However, it can be objected that this strong reading simply begs the question against externalist representationalism, since that is precisely what is being denied. The 'weak' version on the other hand, construes colour experience as *counterfactually dependent* upon opponent process channel states. This is closer to what Pautz has in mind, for as he construes it, this claim

> "says only that internal factors **play a role** in determining color experience. For this reason, it does not entail **Internalism** about color experience: the strong thesis that internal factors *completely* determine color experience, so that neurobiological duplicates living under the same laws have the same color experiences." (2006a, p.212, original emphasis)

So, for example, this weaker version is compatible with the claim that a creature needs to have *the right kind of history* in order to have phenomenal experiences (i.e., it's

---

or yellow, depending on the wavelength in question. The explanation for this phenomenon is that at lower light intensities, the red-green channel is dominant, and the yellow-blue channel is largely inactive.

[35] Under low illumination, the luminance sensitivity of the human visual system peaks in the low end of the colour spectrum, thereby causing blue and green stimuli to appear brighter (and red stimuli to appear darker) relative to their appearance under brighter illuminants.

[36] Furthermore, the theory has received some direct neurophysiological support: post-retinal opponent processing cells have been found in the LGN that roughly correspond to red-green and yellow-blue channels (De Valois & De Valois, 1993). Similarly, colour-sensitive opponent processing cells have been discovered in various areas of the visual cortex, including V1 and V4 (Zeki, 1993). And while no direct neural correlates of colour experience have been discovered, there is nevertheless good evidence that the brain uses opponent processing mechanisms to encode chromatic information.

compatible with denying that Swampman[37] has phenomenal experiences). What's more, nearly all externalist representationalists—including Hilbert, Byrne, and Tye—accept the weak version of the claim, and in fact rely upon it in their own accounts. Nevertheless, Pautz demonstrates that even the weakest version of the claim that opponent processing channels play a "direct" role in determining the phenomenal quality of a sensory experience is enough to present a fatal problem for externalism.

In Pautz's thought experiment, Maxwell and Twin-Maxwell are near-duplicates in different possible worlds. (To make the situation neutral and symmetrical, assume that neither one is in the *actual* world.) Both are products of evolution by natural selection, however, being from different possible worlds, they are products of wholly different evolutionary histories. Furthermore, both have identical configurations of receptors in their retinas, such that identical stimuli will cause identical activation patterns across their cone cells. However, the *post-receptoral* wiring in their retinas is different, such that identical receptoral activation patterns will nevertheless cause them to be put into different opponent process channel states.

To be clear: by virtue of the identical configurations of their cone receptors, Maxwell and Twin-Maxwell are responding to the same external physical property *P* (presumably, a surface reflectance type), which causes the same pattern of activation *A* in the retinal cone cells of both subjects. However, *A* leads to differences in processing downstream: Maxwell is caused to go into opponent channel state *M*, whereas Twin-Maxwell is caused to go into opponent channel state *TM*.

Furthermore, given that (ex hypothesi) both individuals are products of evolution by natural selection, and (ex hypothesi) there is nothing internal or external that is interfering

---

[37] "Swampman" is a hypothetical creature that is the subject of a philosophical thought-experiment presented by Davidson (1987), wherein a freak lightning accident causes the spontaneous generation of a molecule-for-molecule duplicate of a human being—*Swampman*. Although Swampman looks, acts, and talks like a genuine human being, he has no evolutionary history, and therefore is not actually a member of the species. Furthermore, if one holds that sensory representation is determined by evolutionary history, it follows that Swampman has no sensory representations, and therefore (assuming representationalism) no phenomenal experiences. Dretske (1995) explicitly endorses this view of Swampman, although Tye (1998, 2000) rejects it, and holds that Swampman does in fact have phenomenal experiences despite his lack of evolutionary history.

with their respective perceptual processes, it seems clear that parallel conclusions should apply in both cases: each individual is in "optimal conditions" relative to their respective evolutionary histories. Therefore, it follows that both experiences are veridical.

The problem for externalist representationalism is that since (1) optimal conditions obtain for both, and (2) both experiences are caused by the same external property *P*, externalist representationalism should predict that state *M* in Maxwell and state *TM* in Twin-Maxwell both represent *P*, and that their respective sensory experiences are therefore phenomenally identical (since they have the same representational content).[38] However, if even the *weak* version of the claim that opponent processing channels play a "direct" role in determining the phenomenal quality of a sensory experience is true, then it seems to follow that they have phenomenally different experiences.

For example, given the difference in opponent processing, an external stimulus that Maxwell perceives as being a unitary colour might be perceived as a binary colour by Twin-Maxwell. As a result, their behavioural dispositions will be different—namely, psychophysical experiments would show they will differ with regards to their sorting behaviour, matching behavior, judgements of relative similarity, etc. Therefore, it seems that one is inevitably led to the conclusion that Maxwell and Twin Maxwell have different colour experiences.

2.3.2 The Response from Byrne & Tye

The core of B&T's response to Pautz is that either Maxwell or Twin Maxwell must be misperceiving, and thus externalist representationalism in fact delivers the correct verdict of *different experience*. But it is not clear what justifies this claim. One cannot simply stipulate that one experience is veridical and the other misrepresents, because any attempt to do so leads to a circular argument. That is, the claim that one is misrepresenting entails that optimal conditions do not obtain for that individual. However, on the tracking account under consideration, the notion of verdicality is explained in terms of causal co-

[38] Again, this is very similar to how measuring instruments represent: an electric thermometer and a mercury thermometer can both track the same external property (and thus both have the same representational content) despite the physical differences in their respective internal states.

variation under optimal conditions. As such, one cannot then define “optimality” in terms of whether or not the representations produced in response to stimuli under those conditions are veridical.[39] In short, optimal conditions cannot be defined as simply those conditions in which our experiences represent things *as they are*; rather, in order to show that non-optimal conditions obtain, one would have to specify some non-representational fact which justifies the assessment of non-optimality.

However, it’s not clear what sort of non-representational fact could justify the claim that optimal conditions fail to obtain for one of the individuals. For example, one cannot solve the problem by treating the case of Twin-Maxwell just like the case of Tom. That is, one cannot apply the “optimal conditions” for colour perception that hold for members of Maxwell’s species to Twin-Maxwell (nor vice versa).[40] As should be clear from the above description, Twin-Maxwell is *not* the same species as Maxwell—any more than Swampman is a human being. Rather, both Maxwell and Twin-Maxwell are a result of *wholly different* causal-selection histories that just happened to produce something

---

[39] This is precisely the strategy that B&T adopt in their argument against Pautz. They claim that Twin Maxwell is “in general…seriously in error about the colors of things.” (p.13) Their reasoning behind this claim has the following structure:

1. Twin-Maxwell perceives object x to have colour y
2. Object x ‘really’ has colour z
3. Therefore, Twin-Maxwell's experience is nonveridical (1,2)
4. Therefore, it must be the case that “optimal conditions” do not obtain for Twin-Maxwell (3)

Using this argument, Byrne & Tye claim that Tye’s theory (correctly) delivers the verdict of different experience, and Pautz’s argument against externalist representationalism therefore fails. However, the above argument is clearly fallacious. As David Chalmers (2005) puts it, “It seems illegitimate to appeal to nonveridicality in explaining why conditions are not optimal. For an experience to be nonveridical is for it to have a false content; and on the Dretske/Tye account, the content of an experience is to be explained partly in terms of the notion of optimality. If optimality is then explained partly in terms of veridicality, this account will be circular.”
This precisely mirrors Pautz’s response. As he puts it, Byrne & Tye’s argument violates what he calls the *Circularity Constraint*: “if a psychosemantics is to reductively explain sensory representation in terms of causal-covariation under optimal conditions, then the relevant notion of ‘optimal conditions’ must be non-representational. Otherwise the account will be circular.” (p.1)

[40] Byrne & Tye state their argument in terms of possible worlds, arguing that the closest possible world in which Maxwell’s opponent channels are in state TM is one in which evolution is kept fixed, and thus is one in which Twin-Maxwell would be misperceiving, because optimal conditions fail to hold for him. However, this response depends on the assumption that Twin-Maxwell is in the closest possible world in which *Maxwell himself* would be in state TM. However, this response radically mischaracterizes Pautz’s argument. We’re not talking about the closest possible world in which Maxwell’s opponent channels are in state TM. Rather, we are talking about a wholly different individual; one with a completely different evolutionary history.

extraordinarily similar. So given that M and TM are members of different species, nothing seems to preclude us from stipulating that they are both in 'optimal conditions' relative to their respective evolutionary history. (This is especially so given that even in principle we have no way to decide which one—if either!—is correctly perceiving.)

In response to this, B&T simply say:

> "Allegedly, Twin Maxwell is a product of natural selection, someone operating under the same laws as Maxwell with a similar kind of visual system, whose experiences represent the same range of colors as Maxwell's, and who not only has no abnormalities whatsoever in his visual system but also is subject to significant color illusions. Pautz simply stipulates that all these conditions can be met together. A defender of Tye's theory may reasonably deny it. Each condition is indeed metaphysically possible, but they are not all possible together." (p.13)

But why aren't these conditions co-possible? Although Byrne & Tye are not explicit about this, their view seems to be that natural selection simply could not produce an organism that could be perceiving veridically under the specified set of conditions. However, it's again unclear as to what justifies this claim. Indeed, as Pautz points out, natural selection does not—and indeed, *cannot*—discriminate between opponent processing systems that are equally adaptive. In other words, natural selection *underdetermines* the structure of our opponent processing system. There are many possible variations of opponent processing mechanisms that provide the same degree of evolutionary fitness.[41] (Indeed, the classic conception of the "undetectable inverted spectrum" is one such example of an equally adaptive opponent processing structure,

[41] Of course, this argument would seem to apply to members of the same species as well. That is, if natural selection underdetermines the exact structure of opponent processing, then there is no *adaptive* story to be told that explains why Tom's experiences aren't veridical. Nevertheless, one can simply stipulate that there is in fact a single 'correct' opponent processing structure for members of the same species, even if there are equally adaptive variations in the population. In other words, the teleological story that justifies the privileging of one particular opponent processing structure over others as the 'correct' one for members of that species typically does not claim that evolutionary fitness is the overriding criterion for 'correctness'.

since by stipulation the undetectable inverted spectrum allows different perceivers to make identical perceptual discriminations.)

Given these problems, I argue that Byrne & Tye's official response to Pautz's argument fails to secure the desired conclusion of *different experiences*. However, despite the failure of their official response, I believe that there is a better solution open to the defender of externalist representationalism: namely, to hold that Maxwell and Twin-Maxwell are in fact tracking *different* properties.

In order to accommodate this claim, the externalist representationalist could hold that the tracked properties should be identified not by reference to their *receptoral* effects, but rather, to their 'downstream' effects. This would allow the defender of externalist representationalism to hold that Maxwell and Twin-Maxwell are tracking different properties (thanks to their differing post-receptoral wiring), despite the fact that they are both responding to precisely the same physical stimulus (thanks to their identical receptors).

At first glance, this would seem to be at odds with the way that externalist representationalism is typically characterized. Specifically, externalist representationalism tends to be explicit about individuating external properties by reference to their receptoral effects. For example, both Tye (2000) and Byrne & Hilbert (2003) individuate colour properties in terms of their disposition to reflect light that would affect the three types of retinal cone cells in specific ways; specifically, in terms of their relative levels of activation. So, for example, both Tye and B&H claim that a surface is unique red iff it has a reflectance profile type that (under "optimal conditions") would reflect light in such a way as to cause: (i) a greater degree of activation in the L-cones than the M-cones, and (ii) a degree of activation in the S-cones which is equal to the sum of the activation in the L- and M- cones.

Importantly though, these receptorally-inclined proposals are specifically designed with downstream effects in mind. Indeed, both Tye (2000, p.160) and B&H (2003, p.15) are explicit that it is the fact that the cone activation patterns described above lead to certain opponent processing states (under 'optimal conditions') that makes them relevant to the individuation of colour properties.

Furthermore, claiming that Maxwell and Twin-Maxwell represent different properties seems like the more natural response for the teleologically-inclined representationalist. After all, M and TM are supposed to be members of different species, and externalist representationalism is only committed to sameness of phenomenology when members of the *same species* are veridically representing the same physical property. In other words, to adopt this view would be to claim that (regardless of Pautz's claim to the contrary) sameness in configuration of sensory receptor does not necessarily guarantee sameness in the properties tracked (when two individuals are viewing the same stimulus). Indeed, as externalist representationalists often point out, the physical properties that are tracked by our sensory systems "will be quite uninteresting from the point of view of physics or any other branch of science unconcerned with the reactions of human perceivers" (Byrne & Hilbert, p.11). Indeed, these properties are explicitly recognized to be fully anthropocentric properties, although it is denied that this makes them "subjective" properties in any sense (Hilbert (1987) calls this position "anthropocentric realism").

However, the problem with claiming that Maxwell and Twin Maxwell are tracking different properties is that it seems to totally abandon any sense in which the tracked properties are *response-independent*. That is, given the set of physical properties to which a sensory system is responsive, the physicalist Russellian account posits that there is some perceiver-independent relation (*R**) that orders the set and which explains the structure of perceptual discriminations and/or phenomenal qualities in that sensory modality. However, the claim that this kind of perceiver-independent relation exists seems unlikely if one holds that Maxwell and Twin Maxwell are tracking different properties, for it requires that there is some relation which orders the set of physical properties in such a way that guarantees identity of receptoral effects but nevertheless has different ‘downstream’ effects. Clearly, any such relation would seem to require *essential* reference to the internal organization of different perceivers.[42]

In short, it would seem that the price to be paid for salvaging the externalist account of representationalism is giving up on the physicalist Russellian conception of reductionism. Strikingly, Byrne & Tye (p.10) explicitly argue that this would be the preferred way out of such a dilemma: they claim that it’s more likely that reductivism is false than content externalism is false. However, I argue that such a move is both unnecessary and unmotivated. Rather than give up on reductivism, we can preserve it in the framework of an *internalist* version of representationalism that characterizes sensory content as Fregean, not Russellian. I explore this idea further in the next chapter.

[42] Shoemaker (2003, p.265) provides an excellent illustration of this point with the following example: “…suppose I am presented with a coin, and told that it has two different intrinsic powers. It gets you a Coke if inserted into the slot of an Alpha machine, and it gets you a Pepsi if inserted into the slot of a Beta machine. It might turn out that Alpha and Beta machines are mechanically identical, the only difference between them being that Alpha machines are stocked with Coke and Beta machines are stocked with Pepsi. So the proximate effect of inserting the coin is the same whether it is inserted in an Alpha machine or in a Beta machine. Here it seems clear that in no sense are different intrinsic properties of the coin involved in producing the Coke output than in producing the Pepsi output.”

Chapter 3

3 Internalist Representationalism

This chapter lays out my positive account of representationalism, which is (i) internalist, (ii) based on the notion of Fregean content, and (iii) reductive. Sections 3.1 to 3.3 explain the idea of Fregean content, and introduces the idea of *phenomenal modes of presentation*. I describe how such modes of presentation (MoPs) can be understood as placing "conditions on extension", and I provide an overview of the veridicality conditions for Fregean phenomenal contents, which I argue can be understood in terms of a two-dimensional semantic framework. In section 3.3.3, I provide a reductive account of phenomenal MoPs in terms of *quality space theory* (QST), which I argue can be used to reductively explicate the representational content of phenomenal MoPs using a psychosemantic framework known as *functional role semantics* (FRS). Finally, in sections 3.4 and 3.5, I argue that FRS needs to be supplemented by a theory of sensory intentionality in order to solve the *symbol-grounding problem*, and thus a two-factor theory of sensory content is required. In section 3.6 I provide a theory of sensory intentionality (which I call *sensory ostension*) that: (i) supplements the QST account of phenomenal MoPs for properties with an account of phenomenal MoPs for objects and locations, and (ii) is empirically cashed out in terms of mechanisms such as "visual indexes" and "feature maps".

3.1 Fregean Representationalism

In this chapter, I will argue that phenomenal content should be understood in Fregean (as opposed to Russellian) terms. Specifically, I will argue that the phenomenal content of a sensory experience can be understood as a *mode of presentation* (MoP). Put differently, the phenomenology of a sensory experience is *identical with* or *exhausted by* its MoP/Fregean content. In section 3.2, I describe how phenomenal MoPs have been typically characterized in other versions of Fregean representationalism; specifically, those of Chalmers (2004) and Thompson (2009). In section 3.3 I will build on this basic account by providing a theory of Fregean phenomenal content that differs from that of

Chalmers or Thompson. Specifically, I will attempt to provide the outlines of a reductive version of Fregean representationalism that—*contra* Chalmers and Thompson—does not identify the phenomenal content of a sensory representation with its conditions on extension.

On the Fregean view of representationalism about sensory phenomenology, the phenomenal content of an experience does not include the referent as a proper constituent as it does on the Russellian view. Rather, it involves a *mode of presentation* of the referent. Of course, that is still compatible with the claim that the experience also has Russellian content, of which the referent is a proper constituent.[43] However, the Russellian content of the experience does not affect its phenomenology. (Put differently, the Russellian content of an experience is not part of its *phenomenal* content.)

3.2 Fregean Content

Frege (1892) famously argued for a distinction between sense and reference. Senses are supposed to be *modes of presentation* of the reference. To take a classic example, the concepts Hesperus and Phosphorus have the same reference (the planet Venus), but different senses/MoPs. The same goes for Clark Kent/Superman, H_2O/water, Mark Twain/Samuel Clemens, and Cicero/Tully. Modes of presentation for concepts are typically understood as *ways of thinking about* an object or property. Fregeans argue that they are necessary to explain the non-substitutivity of co-referring expressions in intentional contexts; e.g., one can—without contradiction—believe that Superman can fly while simultaneously believing that Clark Kent can't fly. Senses (or MoPs) are supposed to capture something like the *cognitive significance* of a concept or proposition. For example, one might hold that MoPs can be understood as a particular location in a Quinean 'web of belief'. Alternatively (and very similarly), we might argue that MoPs capture something like the *functional role* of the term/expression.

[43] As Thompson (2009) puts it, "According to Fregean theories of phenomenal content, the phenomenal content that is shared by any two phenomenally identical experiences is a matter of *how* the world is represented, and need not involve sameness in *what* is represented."

Both Chalmers (2004) and Thompson (2009) argue that MoPs should be understood as something like *conditions on extension* or *conditions on reference*; that is, a kind of condition which an object or property must satisfy if it is to be 'counted' as being in the extension of the concept. To take Chalmers' example, the condition on extension/reference for a concept like "Hesperus" might be something like "*the object usually visible at such-and-such a point in the evening sky*". Similarly, the condition on extension for a thought or a proposition is a kind of truth condition; it specifies the way that the world must be for the thought to be veridical. For example, the thought "Hesperus is Phosporus" would translate to something like "*the object usually visible at such-and-such a point in the evening sky is the same object that is usually visible at such-and-such a point in the morning sky*".

This approach can be extended rather naturally to the content of sensory experiences. That is, one can characterize the MoP for a given type of sensory experience—what I will refer to as a *phenomenal MoP*—as a kind of condition that an object or property must satisfy in order for it to be in the extension of the experience. Likewise, the *combined* phenomenal MoPs for a sensory experience provide a set of satisfaction conditions that must be met in order for the experience to be veridical. For example, as a rough, preliminary characterization, the conditions on extension ("CoE") imposed by the phenomenal MoP associated with experiences of greenness might be something like "*the property that causes phenomenally green experiences under certain conditions*." Similarly, a phenomenal MoP for a sensory representation of an object might be something like "*the object that is causing this experience in the appropriate way*". Putting both together, an experience of a green object would have a phenomenal MoP that imposed satisfaction conditions something like "*the object causing this experience has the property that usually causes experiences of phenomenal greenness under certain conditions*." (see, e.g., Chalmers, 2004) Of course, this characterization of phenomenal MoPs is clearly non-reductive, since it refers to (among other things) "phenomenal greenness" and "this experience".[44] However, this preliminary, non-reductive characterization will serve for the moment, and I will return to the issue of reduction below in section 3.3.3.

It's important to note that the structure of Fregean content is different for sensory representations of properties than it is for sensory representations of objects. For example, with regard to the notion of *causation*, the CoE for properties makes reference to a *potential* cause—the property that would cause the experience under certain conditions (which I will elaborate on in more detail below). This is necessary to account for the possibility of misrepresentation—e.g., the possibility that a property that typically causes red experiences is now causing a green experience because viewing conditions are abnormal. In contrast, the CoE for objects makes reference to the *actual* cause of the experience. This is because it's plausible that in order for a particular object to be the referent of a sensory representation it must cause the experience *in the right sort of way*. This is meant to account for the possibility of what's known as "veridical hallucination". For example, suppose that a subject hallucinates that there is a red square in front of them. Furthermore, that subject coincidentally happens to be in a situation in which there *is* in fact a red square in front of them that would cause an experience phenomenally indistinguishable from the hallucinatory one.[45] The causal clause in the CoEs for sensory representations of objects justifies the claim that the hallucinatory experience is non-veridical, despite the fact that the non-causal veridicality conditions of the experience would seem to be satisfied.

An alternative approach would be to reject the causal criteria in favor of a purely *existential* type of Fregean content. This kind of CoE could be characterized as something

[44] Indeed, recall that reduction was defined above (section 1.4) as being explicable in non-phenomenal terms, which excludes expressions such as "this experience". (After all, the experience is precisely what we are trying to reduce.)

[45] We can distinguish cases of genuine perception from veridical hallucination by holding that the former and not the latter are counterfactual-supporting in some appropriate way: e.g., the claim "iff the real red square was not there, the subject would not have an experience of a red square" is true of genuine perception of but not hallucinations. Of course, there are well-known problems with attempting to clearly differentiate veridical hallucination from perception in the same way that Gettier-type problems complicate the distinction between genuine knowledge and mere true belief. However, that's simply *everybody's* problem, and is not unique to my representationalist approach.

like "*there exists an object at such-and-such location with the property that causes experiences of phenomenal redness under certain conditions*". On this view, the hallucinatory experience described above would count as veridical. However, it seems implausible that an object must actually be located at the location that it is represented as being at in order for it to count as a referent of a sensory experience. For example, when one views a scene while wearing goggles with lenses that distort the apparent location of objects, it seems plausible that the objects causing the experience are indeed the referents of the experience, despite the fact that the experience misrepresents their locations.[46] Importantly, this example also suggests that there are not only MoPs for properties and MoPs for objects, but MoPs for *locations* as well. And indeed, not only do our sensory experiences attribute properties to objects, but to locations as well (e.g., "*the sound coming from above and to the right*"). CoEs for locations are somewhat more complicated than those of properties and objects, as they inherently involve the coordination of different types of spatial information, including motor information. For example, the CoE for a particular spatial location might be roughly described as something like "*the visual location graspable by reaching thusly*" or "*the location in auditory space that can be made visible by moving your eyes and head thusly*". I will elaborate on this issue in more detail in chapter 5.

In any case, this general view of sensory content fits nicely with our intuitions about sensory experience and veridicality. Specifically, it fits with our epistemic judgments about the veridicality of sensory experience based on what we know about the causes of those experiences. For example, we will tend to judge a colour experience as being non-veridical when a property that we know to typically cause red experiences is currently causing a blue experience. Similarly, we tend to judge an experience as being non-

[46] Likewise, to take the famous example from Grice (1961), suppose that you are looking down a hallway, and there appears to be a red ball at the end of it. However, unbeknownst to you, there is in fact a mirror part-way down the hall that is angled at 45 to reflect a perpendicular hallway containing the red ball that is actually causing your experience. However, there is also an identical red ball behind the mirror that is located relative to your viewing position in such a way that you would have phenomenally identical experiences regardless of whether or not the mirror was there. The existential view of content for phenomenal MoPs for objects holds that your experience is veridical, made so by the ball behind the mirror. The causal view of content for phenomenal MoPs claims that your experience misrepresents the location of the ball.

veridical when we know that an object is not at the location at which it is perceived, or when we know that there is no object that is causing the experience (i.e., hallucination). As Chalmers (2004) puts it, the Fregean content of an experience directly reflects the *inferential role* of the experience – the pattern of judgments that one makes about the veridicality of one's experience based on other information about the state of the external world.[47]

3.3 Response-Dependent MoPs

Consider again the preliminary characterization of the CoE corresponding to phenomenal MoPs for colour experiences given above: "*the property that causes phenomenally green experiences under certain conditions*." Thompson (2009) calls MoPs of this type "response-dependent". According to him,

> "a mode of presentation is response-dependent if its conditions on reference are properly characterized as picking out at a world only the properties or objects that are disposed to produce a response *r* in subjects *s* under conditions *k*" (p.103)

The three variables in the above definition correspond to three important issues about phenomenal MoPs that still need to be clarified. The first issue (the response type *r*) has to do with *reduction*; namely, whether or not the relevant response type can be explicated in non-phenomenal terms. The second issue (the subjects *s* for whom the characterization holds) has to do with the *scope* of the characterization. The third issue (the conditions *k*) is relevant to determining the *veridicality conditions* of the representation. In essence then, this definition attempts to establish that a particular response *r* 'counts' as a representation of some object, property, or spatial location for some particular subject(s) *s* by providing veridicality conditions (*k*) for (*r*).

---

[47] This is not to say, of course, that when we perceive a property or an object we consciously conceive of it as being the cause of our experience. That is, one should avoid "over-intellectualizing" (Chalmers, 2004) the Fregean content of an experience, as such content is likely non-conceptual. Characterizing such content in linguistic terms should not mislead one into thinking that the content has anything like a propositional structure.

Applying these three issues to the preliminary characterization of phenomenal MoPs for colour experience provided above: (i) the relevant response type *r* is characterized in purely phenomenal terms ("…*the property that causes phenomenally green experiences…"*) and thus is non-reductive; (ii) the definition makes no reference at all to what sorts of subjects the characterization holds for; and (iii) the conditions *k* under which the object or property must be disposed to produce that response type are not described ("certain conditions").

Thus, in what follows I will attempt to provide an account that fleshes out these details. I will focus primarily on the phenomenal MoPs for *properties*, leaving objects and locations to the side for the moment. Specifically, I will argue that (i) the response type *r* can be characterized in reductive terms (using the empirical framework of neuroscience and psychology), that (ii) the scope of the characterization is limited to individual subjects (because of the indexicality of the Fregean content), and that (iii) the conditions for veridical perception of properties can be understood as those that are statistically 'typical' or 'normal' for the subject in question. I will address the issues of subjects *s* and conditions *k* first, and return to the issues of the response type *r* later, leaving the preliminary, non-reductive characterization to stand for now.

3.3.1 Subjects

What is the scope of the above characterization of phenomenal MoPs? Another way to pose this question is to ask for what subjects is the following claim true: "property *p* is in the extension of a given phenomenal MoP iff *p* produces response *r* under conditions *k*".

One approach would be to adopt the tracking representationalist's teleologically-based view, and hold that the scope of this claim extends over all members of the same species. On this view, a property *p* would count as being in the extension of a given phenomenal MoP if response *r* was *selected* (by evolution) to be tokened in creatures of that species in response to property *p* under conditions *k*. However, this route leads to troubles, as was shown in chapter 2. Alternatively, one could drop the reference to teleology, and hold that *p* is in the extension of the MoP when it is disposed to produce a certain response *in the majority of perceivers in one's community*. However, this view also leads to some counterintuitive consequences: for example, it would entail that a colour-blind person living within a community composed mainly of individuals with normal colour vision would be subject to systematic misperceptions of colour, yet that same colour blind person would be perceived veridically if their community consisted instead of other similarly colour-blind individuals.

Beyond the basic strangeness of the claim that the veridicality of a subject's sensory experience is determined by their neighbours, the reason that the 'majority' view doesn't work is because it fails to capture *how the subject themselves* judges the veridicality of their experience. That is, if we understand MoPs as grounded in the inferential role of an experience, then the veridicality conditions must reflect that role. (Indeed, MoPs are posited precisely to capture something like cognitive significance!)[48]

In contrast to the 'teleological' and 'majority' views, the account proposed here holds that the scope of the claim is limited to *individual subjects*. On this view, a property *p* is in the extension of a phenomenal MoP for a given subject only if it produces response *r* in that particular subject under conditions *k*. On this view, phenomenally identical experiences in different subjects will have identical Fregean contents (MoPs). However, they may have different Russellian content (extensions), since the same MoP may pick out different objects and properties in different individuals. Furthermore, phenomenally identical experiences in the *same subject* may pick out different Russellian content under different perceptual conditions. (I discuss this possibility further in the next section.)

We can explain this feature of phenomenal content by holding that phenomenal MoPs necessarily contain an inherently *indexical* element that can be understood in terms of a two-dimensional semantic framework. This view is defended by Chalmers (2004), Egan (2006), Thompson (2009), and Brogaard (2010, 2011). In such a theory, one considers different ways the world might be, and then determines what the extension of the term/concept or the truth value of the thought/proposition would be for a given individual

[48] Of course, a colour-blind individual could come to adopt their community's standards of veridical perception, such that their epistemic judgments about the veridicality of their own experiences is based on those standards, and thereby reflected in the inferential role of the experience.

at a given time under those conditions. More formally, the primary intension of a concept is a function mapping from *centered possible worlds* (a possible world centered on an individual at a time) to extensions, and the primary intension of an expression or proposition is a function mapping from centered possible worlds to truth values.

This sort of semantic theory is used to make sense of indexicals like “I” and “now”. The extension of “I” is the individual upon which the possible world is centered; likewise, the truth value of an expression like “I am Stephen Harper” is true when uttered by him, but not when uttered by me.

Applying this two-dimensional semantic framework to a Fregean representationalist theory of phenomenal experience, it follows that the Fregean content of a phenomenal experience is a mapping from centered possible worlds to extensions. For example, on this view the MoP for phenomenal redness is a function which picks out the property (or properties) that cause red experiences in a particular individual under certain conditions. Thus, using this criterion to revise/refine the preliminary characterization of phenomenal MoPs given above, the CoE imposed by the phenomenal MoP for redness might be something like “*the property that causes phenomenally red experiences* in me *under certain conditions”*. Importantly, it is this inherently indexical element of phenomenal MoPs that allows for the possibility of *phenomenal variation without misrepresentation* described above in section 2.3. Since the response *r* that a given property *p* produces is relativized to the individual perceiver, different individuals can veridically represent the same property with phenomenally different MoPs. Likewise, phenomenally identical experiences can pick out different properties in different individuals without misrepresentation.

3.3.2 “Certain Conditions”

A vitally important part of specifying the content of a representation is specifying the conditions under which the representation is veridical. This is necessary to account for the possibility of misrepresentation. Thus, phenomenal MoPs impose conditions on extension such that a property counts as being the referent of an experience only if it is disposed to produce a certain response in subjects *under certain conditions*. Likewise, an experience of an object or location having a particular phenomenal quality (or of an object at a location) only counts as veridical if the state of affairs that the MoP "describes" (i.e., its veridicality conditions) would cause one to have that experience *under certain conditions*. But how should one understand this notion?

As was mentioned in the discussion of "subjects" above, one possible way to characterize these conditions would be to adopt the tracking representationalist view of the conditions for veridical perception as those that are "optimal", where optimality is defined in teleological terms. (i.e., "optimal" conditions are those that our sensory systems were "designed" (by evolution) to operate under). However, this route again leads to troubles, as was shown in chapter 2. Or again, one could drop the reference to teleology, and hold that the conditions under which a sensory representation of a property *p* is veridical are those under which a *p* will produce a certain response *r* in the majority of other perceivers in one's community. But this leads to the same counterintuitive results about colour-blind individuals that was discussed above.[49]

Instead, the indexical element to Fregean content described above suggests that the conditions under which one perceives veridically are those that are *statistically normal* or *typical* for the particular individual in question. One way to illustrate this view is in terms of Block-style "inverted Earth" cases.[50] On the view argued for here, after we're switched with our doppelganger on inverted Earth our experiences *misrepresent* the colours of things (because conditions are not statistically normal), despite the fact that, subjectively

[49] As described above, accounts of this sort fail to capture how the subject themselves judges the veridicality of their experience. (Although again, a colour blind individual could of course adopt their community standards regarding the conditions for veridical perception, which would thereby be reflected in the inferential role.)

[50] Inverted Earth is a thought experiment from Ned Block (1990) which imagines a possible world unique in two respects: First, the colour of everything is inverted relative to the way it is on Earth—that is, grass is red, the sky is yellow, and blood is green. Second, colour words are also inverted. When a resident of Inverted Earth is asked what colour grass is, he correctly answers 'green' (in Inverted Earth English, of course), even though the grass is what we would call 'red'. The situation then, is that while you sleep a team of mad scientists insert colour-inverting lenses in your eyes (as well as appropriately changing the colour of your hair and skin) and switches you with your doppelganger on Inverted Earth. When you awake, you do not notice the change, since the inverting lenses neutralize the inverted colours, and it seems to you that public-language colour words are being applied correctly, since Inverted Earthlings say 'green' when talking about the colour of grass and 'blue' when talking about the colour of sky.

speaking, nothing seems to have changed. However, after spending enough time on inverted Earth, our experiences no longer misrepresent the colours of objects. That is, because phenomenal MoPs contain an inherently indexical element that refer to the ‘normal’ cause of an experience ‘in me’, after sufficient time has passed, the conditions on extension imposed by our phenomenal MoPs of colour are changed so that they now refer to the properties that typically cause our experiences on inverted Earth.

Of course, this result immediately raises the question of how much time is “sufficient” for this shift in content to occur. This is a very difficult question to answer, and indeed, *it may have no definitive answer*. In fact, Thompson (2009) argues that the vagueness here may in fact be a *virtue* for the view, and can also help to make sense of the “in-between” stages:

> “the vagueness of “typicality” might be seen as a virtue in this context. For our intuitions about various scenarios are fuzzy in roughly the same places as are our intuitions about what counts as typical. The Inverted Earth thought experiment discussed above provides one example. Insofar as it is difficult to say what the typical cause of Fred’s red experiences is, we have difficulty deciding whether his experience is veridical. After a long period of time, after which it becomes clear that Fred’s experience has as its typical cause the statistically normal cause on Inverted Earth, we find it easier to decide that Fred’s red experiences are veridical. If there are scenarios of which we have great difficulty deciding whether an experience would be veridical, we should not expect a characterization of the experience’s phenomenal content to deliver a clear verdict about those cases.” (p.105)

To see exactly what this view entails, focus on the extension of MoPs instead of the veridicality of experiences. On the ‘typicality’ view, the property that is picked out by my sensory MoP for phenomenal redness changes after “sufficient” time on inverted Earth; whereas it once picked out one property (or set of properties), it now picks out a completely different one. The problem, however, is how to answer the question of what property is picked out by a phenomenal MoP of redness when it is unclear what conditions are “typical”. Is it the property that caused red experiences in me in the past under “typical” conditions on Earth, or is it the one that now causes red experiences in me under “typical” conditions on inverted Earth? Or perhaps both? Or neither? On the “statistically normal” view of the proper conditions for veridical perception, there doesn’t seem to be a principled way to provide non-arbitrary answers to these questions.

Furthermore, this problem arises in more than just hypothetical “inverted Earth” scenarios. For example, as one ages there is a gradual accumulation of yellow pigments in the lens of the eyes, causing an eventual decrease in the transmission of blue light (~410 nm) to the retina. The result is a constant (albeit gradual) change in colour perception as one ages. Thus the notion of ‘typicality’ for an individual becomes even more problematic.

Similarly, another potential objection to the “typicality” account which is related to the one above is that the account would seem to be vulnerable to the objection that it cannot account for *first* experiences. (Thompson, 2009) That is, *normality* and *typicality* are inherently backward-looking notions, and first experiences have no normal or typical cause. In short, when there is no history, it’s unclear what we should say about the extension and/or veridicality conditions of the experience. If one accepts that infants and Swampmen have phenomenal experiences, and that the phenomenology of those experiences are a kind of representational content, then what are we to say about *what* those experiences represent?

I argue that what these apparent problems actually demonstrate is that *there is no such thing as the “definitive” reference/extension of a phenomenal MoP.* That is, what I’m suggesting is that the notion of ‘typicality’—and thus, the extension (or veridicality) of our sensory representations—is *relative to our explanatory interests*. In short, I’m advocating a kind of instrumentalism or interpretationalism about the Russellian content of phenomenal MoPs. This position is of course fully compatible with Fregean representationalism, since the Russellian content of a sensory experience does not affect the phenomenology of that experience. (In other words, this proposal is not vulnerable to the objection that the phenomenology of a sensory experience is not indeterminate or interest-relative, as a Russellian version of representationalism would be.)

As described above, a phenomenal MoP is a function that maps from centered possible worlds to extensions. It accomplishes this mapping by imposing conditions that a property must satisfy if it is to be in the extension of the MoP. One of these conditions is that the property would cause a certain response in the subject under the conditions for perception that are typical for that subject. However, holding that the relevant notion of typicality can be interest-relative need not be taken to imply that phenomenal MoPs somehow lack clear conditions of satisfaction.[51] Rather, as Thompson puts it, the "conditions of satisfaction involve a presupposition that is not satisfied—that there is a typical cause of experiences of that phenomenal type. This makes them infelicitous, but not false." (p.106)

Finally, it is important to note that the conditions for veridical perception of properties are much more complex than has been described thus far. For example, in colour vision, the quality that a colour stimulus presents is dependent on at least six distinct factors, including: (i) the subject, (ii) the state of adaptation of their visual system, (iii) the ambient illumination, (iv) the surround, (v) the viewing angle, and (vi) the angular subtense. If any of these factors are changed, the colour quality that the stimulus presents can change as well. Thus, for example, a stimulus viewed under artificial indoor lighting can appear to be different in colour than if it is viewed in outdoor sunlight; similarly, the same stimulus can appear to be different in colour in noon daylight as opposed to morning or evening light, or in Northern hemisphere daylight as opposed to daylight at the equator. Thus, at a minimum, the conditions for perception of colour must be relativized to each of these six factors.

In other words, the CoE associated with colour experiences should be revised to be understood as something like "*the property that typically causes experiences of this type in me* under these conditions". Under this revised characterization, the reference to

[51] Indeed, as Thompson correctly points out, the only reason for thinking that MoPs are a type of intentional content *at all* is "because they determine reference and truth conditions."

"typically causes" captures the fact that the conditions for veridical perception must be relativized to what is statistically normal for the perceiver in question, as described above. More importantly however, is that this revised characterization also contains a reference to "these conditions", which is intended to capture the fact that the same external property can look different under, e.g., different lighting conditions. This revision is necessary because this fact—that a stimulus will look different when it is in a dark room versus when it is in direct sunlight—is reflected in the inferential role of the experience. That is, subjects are typically fully aware that although the phenomenology of the experience will vary across those different lighting conditions, it is nevertheless the case that the same external property is being picked out. And indeed, subjects' perceptual capacity to pick out instances of specific colours is stable and robust across changes in ambient lighting conditions, such that one can pick out the 'same' shade of red in daylight, at dusk, under florescent lighting, etc. (This is known as "colour constancy".) Thus, the conditions for veridical perception specified in the response-dependent CoEs must reflect this capacity.

3.3.3 Response Type & Reduction

An important question for the Chalmers/Thompson version of Fregean representationalism is how to understand the response type *r* in the above definition. According to Thompson, we can understand the response type for colour experiences to be "one in which the vehicle of content is an instance of the relevant response type" (p.103). Similarly, in Chalmers' representationalist theory, the response type in the MoP for properties is characterized in entirely phenomenal terms: i.e., "*the property that causes* phenomenally green *experiences...*" His characterization is thus intended to be inherently non-reductive, as it makes essential reference to phenomenal notions in characterizing the content. Indeed, Chalmers explicitly claims that the response type *r* cannot be specified in non-phenomenal terms.[52]

[52] Presumably, the same would apply to sensory representations of objects and locations to: i.e., "the object that..."; "the visual location that..."; etc. The terms would be specified in a phenomenal type way (i.e., not in a way that is directly reducible to neuroscience or psychophysics, etc.) However, I will argue that those MoPs can be so reduced.

However, I will argue that neuroscience and psychology can provide an empirical framework in which one can reductively explain the phenomenal content (i.e., MoPs) of sensory representations. For example, the phenomenal MoP for a sensory representation of a property can be characterized in terms of a psychophysically-defined "quality space" (described below), which can then be given a neurophysiological interpretation. Likewise, this account can provide a reductive explanation of the MoPs associated with locations and objects as well – in terms of neural mechanisms feature-maps and visual indexes, respectively (described in section 3.6 below and in chapter 6).

The most obvious way to specify the response type *r* in entirely non-phenomenal (i.e., reductive*)* terms is simply in terms of a *neural* response type: roughly, the idea would be that the conditions on extension for a phenomenal MoP for a property would be something like "*the property that causes **neural response type X** in me under normal conditions"*.[53] (Indeed, this is precisely the strategy of tracking/externalist representationalism to individuate the tracked properties, as described above.)

However, the problem with this proposal is making the necessary connection between *neural response types* and *types of phenomenal experiences*. It has become widely believed that this is impossible. (Hence the motivation for non-reductive representationalism!) The argument for this claim rests on the possibility of, e.g., *philosophical zombies* and *undetectable spectrum inversion*—in short, the idea that there could be a duplicate that is physically or functionally indistinguishable from me, yet whose experiences are radically different than mine (or even absent altogether). This view is held not only by those who defend a non-reductive view of phenomenal qualities, but also by reductive externalist representationalists. Indeed, externalism is committed to the claim that identical internal brain states could give rise to phenomenally different

---

[53] Similar proposals would apply to objects and locations (e.g., "*the object that is causing neural response x in me")*. There are distinct types of neural responses that underlie our ability to track objects and external locations in our environment, which I will describe in more detail below.

experiences by being related to different external properties (in much the same way that identical internal states of a voltage meter could 'mean' either 6dB or 6lbs).[54]

What is required then is a reductive way of describing the phenomenology of sensory experience that is not vulnerable to the objection that the correlation between response type and type of phenomenal experience is entirely contingent. In other words, we need what Nagel (1974) calls an "objective phenomenology"; that is, a way to "describe…the subjective character of experiences in a form comprehensible to beings incapable of having those experiences." (p.449) Such a theory would allow us to characterize the response type *r* in non-phenomenal terms. Happily, such a theory is in fact available, and is known as *quality space theory*. At essence, quality space theory ("QST") is primarily intended to be a radically *empirical* theory of sensory phenomenology. It is an inherently third-person approach that provides a way in which the phenomenal qualities present in sensory experience (colours, tastes, sounds, smells, etc.) can be both investigated/described by the empirical sciences and explained in objective terms. In what follows, I will briefly describe the quality space approach to sensory phenomenology.

3.3.3.1 Quality Space Theory

In any sensory modality (or sub-modality), phenomenal qualities stand in certain structural relations such as *relative similarity and difference* (e.g., red is more similar to orange than it is to blue). A *quality space* is a spatial ordering of the global structure of similarity and dissimilarity relations that hold between phenomenal qualities in a given sensory modality in which relative similarities between qualities are represented by their relative distance. The similarity relations in question typically vary across multiple dimensions within a single sensory modality: for example, colour qualities vary along three dimensions: hue, saturation, and lightness. By representing each of these dimensions spatially such that the relative similarity of colour qualities in a given

[54] Indeed, this is why externalists must hold a Russellian view of the nature of phenomenal content if they wish to defend a reductive account of sensory phenomenology: i.e., they can't provide a reductive account of phenomenal qualities in terms of brain states, so they provide a reductive account in terms of external properties.

dimension corresponds to their relative distance across that dimension, each colour quality can be characterized as a location in a three-dimensional space, the totality of which constitutes the so-called “colour solid”.

The structure of a quality space for any given sensory modality (or sub-modality) is initially determined by psychophysical experiment; e.g., systematically altering the properties of test stimuli along some physical dimensions and observing the resulting effect on a subject's discriminative capacities. Specifically, by using psychophysical measures such as *discriminability (matching)*, *relative similarity,* and *just-noticeable-differences,* and applying statistical techniques such as *multidimensional scaling* (described below) to this kind of psychophysical data, one can construct a unique multidimensional ordering of the phenomenal qualities in a given sensory modality. Next, this ordering (“quality space”) can be given a neurophysiological interpretation which *explains* the structure that is discovered via psychophysics.[55] (Indeed, while psychophysics can discover the structure of the quality space for a given sensory modality, it is merely an *expanadum* that requires an *explanans*. For the latter, we must turn to neuroscience.)

For example, the structure of colour appearance can be explained by the *opponent process theory* of colour vision, which was briefly described in section 2.3.1. The outputs of the three different kinds of cone cells in the retina are summed and differenced to create three different post-receptor neural channels. Two of the three opponent process channels code chromatic information (one channel corresponds to a red-green opponent relation, whereas the other corresponds to a yellow-blue one), and the third channel is achromatic and codes for lightness. By representing each of the three channels of the opponent processing system as a spatial dimension, every possible state of our opponent processing channels can be understood as corresponding to a point in three-dimensional colour quality space that is discovered by psychophysics. That is, the ‘value’ of each

[55] Essentially, this amounts to finding the realizers of the representational vehicles of phenomenal qualities; those information-carrying states that have both the right structural relations and causal role to underlie phenomenal qualities.

channel (as determined by its level of activation) can be regarded as a Cartesian coordinate along a particular dimension of qualitative variation within the space.

It is important to understand exactly how the quality space approach to sensory phenomenology applies to the current representationalist proposal. The *wrong* way to understand this version of representationalism is to claim that the relevant response type *r* for a phenomenal MoP for an experience of colour should be characterized in terms of brain states--namely, *opponent processing states*. So, for example, according to this (incorrect) characterization, a particular surface reflectance property is in the extension of the phenomenal MoP for redness if it causes one's opponent processing channels to go into the neural state which codes for redness under (statistically) 'normal' conditions. However, this version of representationalism would face exactly the same sort of problems connecting neural states and phenomenal states that were described above. The *right* way to understand this version of representationalism—which I will refer to as "quality space representationalism"—is to recognize that what is relevant to the response type *r* is not that *r* is a particular type of *neural state*, but rather that *r* has a particular *relational structure*. In essence, quality space theory (hereafter, "QST") offers a purely structural description of the phenomenology of sensory experience. This abstract structure can then be given a neurophysiological interpretation that explains both the topology of the structure and the discriminative capacities of the subject, but it is the *structural* details that are relevant to individuating the response type, not the neurophysiological ones.

As described thus far, the quality space approach to sensory phenomenology plausibly allows for a reductive account of Fregean representationalism by providing an empirical, reductive way of characterizing the response type *r* in Thompson's "response-dependent" MoPs. However, I now want to propose a type of Fregean representationalist theory based on QST that is very different from the standard Chalmers/Thompson account.

Specifically, recall that (as was described in the chapter 1) there are two parts to a representationalist theory: (i) a theory of the nature of phenomenal content, and (ii) a theory of sensory intentionality that explains how and why we are able to represent that content. According to the Chalmers/Thompson version of Fregean representationalism, both of these issues are solved by appeal to the above-described *conditions on extension*. That is, according to the Chalmers/Thompson account, (i) Fregean content of a sensory representation is exhausted by its CoE such that the conditions themselves are identical with its phenomenal MoPs, and (ii) these conditions determine the reference/extension of a sensory state, thereby explaining the intentional directedness of sensory states.

However, in what follows, I will argue that—*contra* Chalmers and Thompson—the CoE imposed by MoPs for properties are *not* identical with the phenomenal content of a sensory state. Rather, I will argue that the MoP of a sensory representation should be identified with the response type *r* itself, whose representational content can analyzed in terms of a psychosemantic theory known as *functional role semantics*.

3.4 Tracking Semantics vs. Functional Role Semantics

Broadly speaking, there are two general theories of how the content of mental representations is determined: *tracking* semantics (also known as "causal-informational" semantics) and *functional role* semantics (sometimes called "conceptual role" or "inferential role" semantics[56]). Roughly, tracking semantics holds that the content of a mental representation is fixed by the causal-informational relations that hold between brain states and external objects and properties, whereas functional role semantics holds that the content of a mental representation is determined by the logical or functional relations it bears to other mental representations. Thus, on the tracking view, a brain state represents trees if there is a reliable causal or informational co-variance relation between that state and trees in the external environment. In contrast, according to functional role semantics (hereafter, "FRS"), a brain state represents trees if it bears the rights relations to other brain states that represent leaves, branches, forests, plants, wood, and so on.

[56] As should be somewhat obvious, I use the term "functional" instead of "conceptual" because I am not concerned with *concepts* as such here. Indeed, phenomenal content is often thought to be specifically *non*-conceptual (although I am neutral on that issue in this dissertation). Moreover, even if functional role semantics is the correct theory of phenomenal content, I do not claim that it also is the best theory of content for *concepts*.

These two approaches roughly correspond to the difference between adopting a *referential* theory of content as opposed to a *use* theory of content. The referential view is by far the more dominant account of mental content, especially in regards to perceptual or sensory content. In contrast, FRS derives from the “use” theory of meaning (Grice, 1957; Wittgenstein, 1953), according to which the content of a representation (e.g., a word) is determined by the way it is used in communication.

Transposed into the domain of representationalism about sensory phenomenology, functional role semantics holds that a sensory state has a certain phenomenal content by virtue of the functional relations it bears to other sensory representations. This proposal fits very naturally with the quality space approach to sensory phenomenology: each location in quality space is defined entirely by the relations that it bears to other locations in the space. Likewise, each state of our opponent processing system is individuated by the well-defined functional relations that obtain between it and every other possible opponent process channel state. Thus, the response type *r* itself can be understood as possessing FRS content, which in turn can be identified with the phenomenal content of the state. On this view, a quality space is a *representational* space, with individual locations corresponding to particular contents. Paul Churchland (1989, 1995, 2005) has made similar proposals about sensory representation with a theory known as *state space semantics* (SSS). As he puts it,

> “The basic idea...is that the brain represents various aspects of reality by a *position* in a suitable *state space*, and the brain performs computations on such representations by means of general *coordinate transformations* from one state space to another." (1989, p.78-79)

Notice that SSS/FRS is a theory about *how* the brain represents, not *what* the brain represents. In other words, the theory captures the *intensional structure* of a representational system, as opposed to the set of objects or properties that are in the extension of its representations. In short, FRS/SSS is concerned with content that is roughly analogous to *sense*, not reference.[57]

However, there is a problem with adopting a version of FRS/SSS that only recognizes internal relations to other representations in the system as being relevant to determining content. This view is known as "short-armed" FRS, and it seems vulnerable to the objection that some aspect of content is constituted at least in part by external factors. (Indeed, most philosophers have come around to some version of the basic Putnam/Burge view on externalism about meaning.) The more immediate problem, however, is the fact that a state is not a *representation* simply because it has certain relations to other states within a common system. In other words, mere relational structure does not provide extensionality or veridicality conditions. A location in quality space is not itself intrinsically representational, in the sense of being intrinsically 'about' external properties. This is essentially the "symbol-grounding problem"; that is, the problem of how a "semantic interpretation of a formal symbol system be made intrinsic to the system" (Harnad, 1990). (I discuss this problem in more detail below in section 6.6.) Similarly, Fodor & LePore (1996) present a closely related objection when discussing Churchland's state-space approach to sensory representation:

> "…Churchland has confused himself by taking the labels on the semantic dimensions for granted. The label on a particular dimension says how positions along the dimension are to be interpreted; for example, it says that they're to be interpreted as expressing degrees of F-ness. To label a dimension as the F-ness dimension is thus to invite the questions "In virtue of what do the values of this dimension express degrees of F-ness rather than, say, degrees of G-ness?"." (1996, p.153)

[57] As an example from Churchland (1989) illustrates, there can be clear divergence between the content attributed by tracking semantics and the content attributed by functional role semantics: some utterance in a foreign language might track thunder (and thus be attributed the extensional/referential content "There is thunder") yet, by virtue of the role that it plays in the language, it could actually *mean* something like "God is shouting".

In short, although FRS provides the framework for understanding a particular aspect of the representational content of sensory states, by itself it is only *half* of a theory of representation. That is, although it provides a theory of the nature of *sense* (in terms of phenomenal MoPs), it still requires the addition of a theory of sensory *reference*. Thus, many proponents of FRS have supplemented the "short-armed" approach to FRS by adopting one of two options: (i) a 'two-factor' theory of content in which the internal functional relations determine a type of "narrow" aspect of content roughly analogous to sense or MoP and external relations (such as causal-informational tracking relations) determine the 'wide', referential aspect of meaning; or, (ii) a 'long-armed' (or "non-solipsistic"; Harman, 1987) version of the theory which includes relations to external properties and objects as part of the content-determining functional relations, "but without any commitment to a separable narrow aspect of meaning." (Block, 1998) In what follows I will argue for a two-factor theory of sensory content, and against the "long-armed" version of FRS.

A "long-armed" version of functional role representationalism would hold that relations to external states play a role in determining phenomenal content. This is compatible with the claim that my microphysical duplicate and I could be in identical internal states but have different phenomenal experiences, since we could be systematically related to different things in the world (e.g., Block's "inverted earth" scenario). This result is obviously incompatible with internalism. However, long-armed FRS doesn't *necessarily* entail externalism about content; a version that was compatible with internalism about phenomenal character would be possible, but it would have the consequence that *content underdetermines phenomenology*. That is, representationalism is only committed to the claim that if there is a difference in phenomenal quality of an experience there must be a corresponding difference in the representational content of that experience. A version of internalist representationalism which holds that different contents could have the same phenomenology is thus technically compatible with representationalism; however, it seems to give up on the *spirit* of it.

In any case, the reason that a long-armed FRS for representationalism about sensory phenomenology doesn't work is because it fails to make a principled distinction between those function relations which determine phenomenal character and those that determine the referential content. That's why a *two-factor* theory of sensory content is necessary. As described above, a two-factor theory of phenomenal content posits two distinct mechanisms of representation, one determining representational properties analogous to sense, the other determining representational properties analogous to reference (e.g., McGinn 1982). The joint contribution of these two mechanisms determines the representational content of the sensory state.

3.5 Two-Factor Sensory Representation

According to the proposal here, the Fregean content of MoPs for properties should be understood in terms of a two-factor FRS framework. As already described above, the sense/MoP can be identified with the response type *r* itself, and understood in terms of an FRS interpretation of QST. On the other hand, I argue that the CoE associated with phenomenal MoPs for properties play the appropriate referential role.

Indeed, the CoEs associated with a phenomenal MoP are the way in which *sense determines reference*: i.e., the way in which the Fregean content of an experience determines the Russellian content of that experience. Specifically, as described above in section 3.2, the conditions on extension are grounded in the inferential role of the experience. This inferential role, in turn, can be understood in broadly *functionalist* terms. Thus, the extension of a phenomenal MoP is determined by the specific functional role that that sensory state plays in inference. In other words, sensory states—specifically, the response-type *r*—can be said to represent specific external properties, objects, and locations because those states (i.e., *r*s) are *used* by the subject (and by 'downstream' consumer systems) as indicators (in Dretske's 1981 sense) of those things. In short, the CoEs reflect how a subject uses *r* as a source of information about the external world.

Clearly, this interpretation of CoEs fits neatly with FRS: CoEs are grounded in the inferential role of the experience, which itself is a kind of functional role. The Russellian content of a sensory representation *r* is determined by the way in which *r* is used by the subject, which in turn depends on the relations that it bears to other representations in the system. Moreover, although CoEs seems to describe a tracking relation, it's important to understand that according to FRS, *the content determines the tracking relation,* whereas according to tracking semantics, *the tracking relation determines the content*.

Thus, in addition to providing a representational interpretation of the response type *r* in terms of QST, FRS also provides a way to understand the CoE imposed by the phenomenal MoPs. Moreover, it explains how the Fregean content of a sensory representation manages to determine the Russellian content of a sensory representation. As described above, Fregean content can be understood as a function that maps from centered worlds to extensions. The CoE associated with the MoP performs this mapping function. Importantly however, the CoE is still part of the Fregean content of the experience, although it is not part of the *phenomenal* content of the experience, which is exhausted by the FRS/QST content of *r* itself.

However, there is still a problem with this two-factor version of Fregean representationalism: namely, it does not solve the "symbol-grounding" problem described above. Although the CoEs are supposed to determine the extensions of phenomenal MoPs for properties, they are still vulnerable to the objection that they fail to *ground the intentionality* of sensory states, because they are entirely determined by relations among internal states. In order for those internal states to be genuinely representational, their meaning must be grounded by some direct connection to the external world. (One that does not rely on a "description" – see section 6.6 below.) For this, sensory representations of objects and locations are required.

As in the case of properties, sensory representation of objects and locations should also be understood in terms of a two-factor theory of sensory content, with the response type *r* playing the role of the MoP, and CoEs playing a referential role. However, there are a number of very important differences between these kinds of representations: First, they contribute to the psychosemantics of sensory representation in different ways: MoPs of properties are *predicative,* and function like general terms. In contrast, MoPs of objects and locations are *demonstrative*, and function like singular terms. Second, these two different types of MoPs contribute to the phenomenology of sensory experience in different ways—one is *qualitative* and represents sensory features, the other is *spatial* and *object-based*, and represents sensory individuals. (These differences are explored in further detail in section 6.7) Most importantly however is that the response types identified with the MoPs for representations of locations and objects act as a kind of *referential pointer* that grounds the intentionality of sensory states.

Sensory representations of locations pick out their referents via a type of *sensorimotor directedness* (described in chapter 5), generated by the process of coordinating different type of spatial information from sensory and motor sources.[58] The response type *r* for spatial representations can be identified with the attentional selection of a location on a *feature map* (described in the next section and in chapter 6). On the other hand, sensory representations of object pick out their referents *directly*, like demonstratives. MoPs for objects refer via direct causal contact between external objects and internal states. On this view, the response type *r* can be understood in terms of being indexed by a "FINST" (described in the next section and in chapter 6). I briefly examine both of these types of referential intentionality in the next section.

3.6 Mechanisms of Sensory Reference

One of the most basic types of sensory reference—one that is found in many different sensory modalities—is what Clark (2000, 2004) calls *feature placing*. Roughly speaking, the claim is that sensory reference proceeds by picking out *spatial locations* and characterizing those locations in terms of the phenomenal qualities that appear there. In essence then, one of the most primitive varieties of sensory reference reduces to our capacity for spatial discrimination.[59]

---

[58] This sensorimotor aspect of sensory intentionality is quite important. Indeed, in his discussion of the symbol-grounding problem, Harnad (2005) explicitly claims that "ultimately [symbol-] grounding has to be *sensorimotor*, to avoid infinite regress."
[59] That is, sensory experiences not only have phenomenal qualities (like colour or pitch or taste), but moreover they also have a *spatial character*.

The neural mechanism(s) that underlies feature placing is what Triesman (1980, 1998) calls a *feature map*: a cortical structure containing neurons that respond to some particular dimension (or dimensions) of variation in phenomenal appearance, such as the red-green and yellow-blue opponent processing cells that underlie colour vision. Moreover, these neural structures are called “maps” because the cells are arranged roughly topographically, so that adjacency relations in the brain mirror adjacency relations in space. For example, V1 famously contains (among other things) a spatial point-to-point mapping of the retina. Indeed, the visual cortex contains a variety of different feature maps that respond to different visual qualities—hue, lightness, shape, motion, etc.

Furthermore, feature placing is not the only referential mechanism possessed by sensory systems. For example, Pylyshyn (2001, 2003) provides evidence for the existence of “visual indexes”—neural mechanisms which keep track of up to four or five individual objects (or “proto-objects”) in the visual field. Phenomenal qualities can be assigned to the visual indexes associated with particular (proto-) objects, and thus can be ‘bound to’ (or ‘predicated of’) those objects in the same way that they can be bound to locations.

Through mechanisms like feature maps and visual indexes, sensory states possess a rudimentary type of intentionality that could be described as *sensory ostension*: a type of gesturing or pointing via sensory systems. Indeed, Pylyshyn describes FINSTs as functioning something like natural language demonstratives, allowing for reference without concepts. Like demonstratives, these mechanisms allow sensory systems to refer to objects without explicitly representing any of their properties. Similarly, Clark describes feature placing as a kind of basic sensory reference, with features maps referring in a way roughly analogous to real maps. Moreover, “master maps” of feature map locations allow sensory systems to represent spatial locations independently of any features that happen to be instantiated there.

These types of mechanisms can be said to “ground the intentionality” of sensory states. They are basic mechanisms of sensory reference by virtue of which our experience is ‘about’ the world. They act as “referential pointers”, picking out objects and locations directly via a kind of sensory ostension. Their referents can then be characterized using a sort of predicate—a phenomenal quality. However, the story about sensory representation is still not entirely completely: there is one additional element that has not yet been discussed: *binding*.

3.7 Binding

The capacity to characterize a location or object as possessing a certain phenomenal quality is the basis of sensory representation. Importantly, this capacity requires more than mere sensory reference—it requires binding. Specifically, on the representationalist proposal argued for here, the (Fregean) content of a sensory experience is characterized in terms of two factors: (i) a phenomenal mode of presentation, characterized as a position within a certain quality space, and (ii) a referred location or object—a *sensory individual*— that is characterized in terms of feature maps or visual indexes. However, while the referential mechanisms described above can ground the sensory *intentionality*, it is the relation between these factors which make sensory states *genuinely representational*. Specifically, the binding of a phenomenal quality to a particular sensory individual—such as an object or location—can be evaluated for veridicality, depending upon whether or not the object/location in question actually has that feature.

In other words, it is in virtue of the capacity to bind a phenomenal quality to a sensory individual that sensory systems bear the hallmark of genuine representation: their states can be assessed for *veridicality*. Roughly, a sensory representation is veridical iff the location or object picked out by the sensory system has the property that normally causes that quality under the current perceptual conditions. However, it's important to note that on this proposal, *the content of a sensory representation is not identified with the actual, normal, or optimal cause of that sensory state*. Rather, recall that on the current proposal, sensory content is *Fregean*, not Russellian. That is, whereas Russellian contents constitutively include the referents of sensory representations, Fregean 'bound' contents are more akin to a description that imposes conditions that the demonstratively-picked out referents must satisfy if the experience is to be veridical.

3.8 Chapter Summary

In this chapter, I have presented the outlines of my version of reductive internalist representationalism. Specifically, I have attempted to describe not only the basic psychosemantic structure of the Fregean content of sensory experience, but in addition, I have attempted to give a preliminary account of the neural mechanisms that underlie the ability to discriminate and represent properties, spatial locations, and objects. (Specifically, in terms of opponent process theory, feature maps, and visual indexes, respectively.) Next, in the following three chapters, I will attempt to elucidate the empirical underpinnings of these mechanisms in much greater detail, examining the methods of the psychological and neuroscientific explanations that invoke them. In this way, I will demonstrate not only how my version of internalist representationalism is *reductive*, but moreover, how it can be integrated with empirical theories of sensory experience, thereby motivating the "methodological representationalism" described in chapter 1.

Chapter 4

4 Quality Space Theory

In this chapter I describe the “Quality Space” approach to phenomenal qualities that I utilize in my representationalist theory. In part one, I describe Quality Space Theory (QST) and explain the empirical methodology behind the construction of a quality space (using psychophysics and neuroscience). In part two, I examine how this approach provides an *explanation* for phenomenal qualities by showing how quality space theory fares from the perspective of philosophy of science; more specifically, (i) how QST fits into the tradition of *structuralism* in the philosophy of science, and (ii) how QST is an example of *mechanistic explanation* that has become popular in contemporary philosophy of neuroscience (e.g., Bechtel 2008, Craver 2007). In part three, I use the QST approach to address traditional problems in the philosophy of mind; specifically, in the context of the central complaint that scientific explanation necessarily ‘leaves something out’.

4.1 An “Objective Phenomenology”

Thomas Nagel’s 1974 classic “*What is it Like to be a Bat*” captured the view of sensory phenomenology that has dominated thinking in philosophy of mind for decades: that scientific (“objective”) accounts of sensory experience necessarily *leave something out*; namely, the subjective, ‘what it’s like’ -ness of the experience—its phenomenal qualities.[60]

As Nagel puts it,

> “At present we are completely unequipped to think about the subjective character of experience without relying on the imagination—without taking up the point of view of the experiential subject. This should be regarded as a challenge to form new concepts and devise a new method—an *objective phenomenology* not dependent on empathy or the imagination. Though presumably it would not

[60] The general problem takes many forms: David Chalmers (1996) refers to this as the ‘Hard Problem’ of consciousness; Joseph Levine (1983) refers to it as the ‘Explanatory Gap’.

> capture everything, its goal would be to describe, at least in part, the subjective character of experiences in a form comprehensible to beings incapable of having those experiences.
>
> […]
>
> Apart from its own interest, a phenomenology that is in this sense objective may permit questions about the physical basis of experience to assume a more intelligible form. *Aspects of subjective experience that admitted this kind of objective description might be better candidates for objective explanations of a more familiar sort."* (p.449, emphasis added)

In what follows, I will examine a proposed theory of phenomenal qualities that promises to offer just the kind of "objective phenomenology" that Nagel calls for: namely, *Quality Space Theory* [QST]. Furthermore, I will show how—as Nagel suggests—objective explanations of the subjective features of sensory experience are made possible by this theory.

4.2 Quality Space Theory

In any sensory modality (or sub-modality), phenomenal qualities stand in certain structural relations such as *relative similarity and difference*. For example, red is more similar to orange than it is to blue; a guitar sounds more like a piano than a car crash; strawberries taste more like raspberries than roast beef. A *quality space* is a spatial ordering of the global structure of similarity and dissimilarity relations that hold between phenomenal qualities in which relative similarities between qualities are represented by their relative distance.

The similarity relations in question may vary across multiple dimensions within a single sensory modality: for example, colour qualities vary along three dimensions: hue, saturation, and lightness. By representing each of these dimensions spatially such that the relative similarity of colour qualities in a given dimension corresponds to their relative distance across that dimension, each colour quality can be characterized as a location in a three-dimensional space, the totality of which constitutes the so-called “colour solid”.

According to quality space theory, phenomenal qualities *simply are* locations in quality space. Of course, what that claim amounts to is at least somewhat open to interpretation, as I will discuss below. But at essence, quality space theory is primarily intended to be a radically *empirical* theory of phenomenal qualities—an “objective phenomenology”. It is an inherently third-person approach to phenomenal qualities, which provides a way in which they can be investigated and described by the empirical sciences and explained in objective terms.

Quality space theory has enjoyed somewhat of a resurgence in contemporary philosophy of mind—Rosenthal (2010), Clark (1992, 2000), and Churchland (1989, 1995, 2005), have all recently endorsed versions of the theory. The idea has also been explored in detail by Palmer (1999), Hardin (1988), and Goodman (1977), but its history goes as far back as the logical positivists: Carnap (1928) and Schrodinger (1920) both used the framework of quality space theory to characterize sensory phenomenology. Indeed, the basic idea goes as far back as Issac Newton (1704).

In this chapter I propose to examine quality space theory in two respects: roughly, how it fares as *philosophy of science*, and how it fares as *philosophy of mind*. More specifically:

First: I will show how QST purports to provide an ‘objective phenomenology’ that can both describe and provide objective explanations for phenomenal qualities. This will require an examination of the empirical methodology of constructing a quality space (based on psychophysics and neuroscience) and an account of how QST fits into philosophical debates about the nature of scientific explanation (specifically, *structuralism* about scientific theories and *mechanistic explanations* in neuroscience).

Second: I will show how the ‘objective phenomenology’ provided by QST can be used to address traditional problems in the philosophy of mind; specifically, in the context of the central complaint that scientific explanation necessarily ‘leaves something out’.

In the next section I will describe exactly what a quality space *is*, and highlight an important distinction between two different types of quality spaces.

4.3 Two Types of Quality Spaces

As mentioned above, a quality space is a spatial ordering of the global structure of similarity and difference relations that hold between phenomenal qualities in which relative similarities between qualities are represented by their relative distances. More specifically, a quality space is a *space* in the mathematical sense of the term: it is a *multidimensional order*. An order is simply a set and a relation which orders that set, and to say that order is multidimensional is to say that the relation orders the set along two or more dimensions of variation.

Moreover, a space is a *quality* space by virtue of the fact that it is an ordering of the qualitative features that are present in sensory experience. However, even this formulation is insufficiently precise. There are two different types of "qualities" that can be ordered, and it is important to distinguish between them. The first are the *qualities that stimuli present.* The second are the *qualitative features of sensory states*. The former I will refer to as "appearance properties", the latter, "phenomenal qualities".[61] It is important to distinguish between these two different types of 'qualities' and their respective 'spaces', and thus I will examine their differences below.

Appearance properties are simply the properties that objects appear to have. Basically, these properties are colours, sounds, tastes, smells, and other perceptible properties that we attribute to things in the external world. Importantly however, how an object appears to a perceiving subject is highly dependent upon the internal states of that subject.[62]

[61] As is often the case, this same distinction has been made by other philosophers using different terminology. For example, Clark (1992, 2000) refers to appearance properties as "phenomenal properties", and to phenomenal qualities as "sensory qualities". Likewise, Rosenthal (2010) uses the term "perceptible properties" for appearance properties and "mental qualities" for phenomenal qualities.
[62] For example, objects appear differently to individuals with different forms of colour vision deficiency, due to the absence of one or another type of retinal cone cells. (Individuals with *deuteranopia*—in which the "green" photoreceptors are absent—and individuals with *protanopia*—in which the "red" photoreceptors are absent—will perceive the same stimuli as having very different appearance properties.)

Phenomenal qualities are qualitative features of sensory states in virtue of which objects in the external world appear as they do.

Put differently, terms like "red" or "cold" characterize how external objects appear, and thus refer to appearance properties. Phenomenal qualities are not themselves red or cold; rather, phenomenal qualities are the properties of sensory states in virtue of which a subject perceives external objects as having the appearance properties (such as 'red' or 'cold') that they do. Of course, appearance properties are also not themselves red or cold; rather, *objects* are red or cold, in virtue of the fact that they instantiate certain appearance properties.[63] (Compare with "solubility"—the property of solubility is not itself soluble; rather, objects are soluble in virtue of possessing that property.)

It's important to note that appearance properties cannot be identified with any intrinsic physical properties of the stimuli that present those qualities, for the simple reason that *physically identical stimuli can present different qualities and physically different stimuli can present the same quality*. (See chapter 2 for the argument against physicalist Russellianism.) Rather, appearance properties are *relational* properties; that is, relations between physical properties of objects and perceivers (though the relations will be quite complex; as noted above in section 2.2, colour vision involves a minimum of six relata!) Appearance properties might also be usefully thought of as *dispositional* properties; that is, dispositions to produce such-and-such a response in certain types of sensory usefully systems under certain conditions.[64]

---

[63] I am using the term "object" here in a very loose and liberal way, such that, e.g., the *sky* counts as an object.

[64] To put it somewhat differently, consider the distinction between appearance properties and phenomenal qualities in *representationalist* terms. According to Russellian representationalism, phenomenal qualities *are* appearance properties, and can be identified with (response-independent) physical properties of external objects. In contrast, according to Fregean representationalism, phenomenal qualities are identified with *modes of presentation* of appearance properties. The Fregean view holds that appearance properties are constitutive of the Russellian content of the sensory states, but it denies that Russellian contents are phenomenal contents. Likewise, the Russellian view can hold that sensory states possess "properties in virtue of which external objects appear as they do", yet it denies that those properties are phenomenal qualities. (Indeed, there is nothing precluding the Russellian representationalist from holding that sensory states possess Fregean contents or MoPs; however, they will deny that those MoPs should be identified with phenomenal content. Of course, Russellian representationalism is also compatible with the claim that phenomenal qualities are entirely non-representational properties.)

Given the difference between appearance properties and phenomenal qualities, we can then define two different types of quality spaces:

A *phenomenal quality space* is a type of psychological-neuroscientific model of sensory phenomenology. It is an ordering of the qualitative features present in sensory experience. It is this sort of quality space that is of central interest here.

An *appearance property space* typically has more practical purposes. The most common examples of such spaces are *colour order systems*; i.e., arrangements of colour stimuli used for, e.g., pigment mixing or digital displays. Examples of these sorts of spaces include the Munsell colour space, the Swedish NCS space, the CIE-lab space, the OSA space, sRGB, etc. These different spaces represent different types of stimuli and employ different ordering relations to order those stimuli, depending on practical considerations.

As Clark (1992, 2000) points out, although these sorts of 'practical' colour spaces aren't particularly interesting to philosophers and/or cognitive scientists, they *can* tell us interesting things about the nature of a creature's internal sensory states. Specifically, the construction of a psychophysical phenomenal quality space—that is, a psychophysically-defined space which attempts to model the structure of phenomenal qualities—requires first constructing an appearance property quality space, then extrapolating the structure of the former from the latter. The next section will describe this process.

4.4 Methodology pt.1: Psychophysics

To construct a quality space that describes the structure of phenomenal qualities in a given sensory modality, one begins with psychophysics. Methodologically, this involves varying the physical properties of test stimuli along some particular dimension(s) and observing the resulting effects on a subject's discriminative capacities.

The construction of a quality space begins with purely behavioral measures like *discriminability* (and judgments of relative similarity, etc.) because these are the empirical entry point for investigating the phenomenology of sensory experience. According to quality space theory, phenomenal qualities are the properties of sensory states in virtue of which a subject perceives external stimuli as having the appearance properties that they do. Assuming that we discriminate stimuli on the basis of their appearance properties, it follows that phenomenal qualities are those properties by virtue of which we make the perceptual discriminations that we do**.** Therefore, by investigating the range of discriminative capacities in a given modality we are indirectly studying properties by virtue of which those discriminations are possible.[65]

There are typically three behavioral measures which are used to gather the data necessary to construct a quality space: *matching* (i.e., *indiscriminability*), *just noticeable differences* (*JNDs*), and *relative similarity*. That is, one begins constructing a quality space for a given sensory modality by utilizing one or more of the following measures: (i) repeatedly presenting subjects with two stimuli and asking if those stimuli match (i.e., are indiscriminable), (ii) measuring the threshold at which a perceptible difference between two stimuli becomes noticeable by the subject (JNDs), or (iii) repeatedly presenting subjects with three stimuli—a target and two foils—and asking for judgments of relative similarity (i.e., "is *x* more similar to *y* or to *z*?").

By gathering this kind of behavioral data—discriminability data, relative similarity judgments, etc.—we discover which types of stimuli the subjects perceive as being similar and different. This allows us to construct a unique ordering of the qualities presented by the stimuli: an appearance property quality space.

---

[65] Indeed, many have argued that psychophysics as a discipline should be properly regarded as investigating the relationship between stimuli and sensory phenomenology. (Horst,2005) However, this view has been challenged by those who argue that psychophysics deals with *discriminative capacities* and nothing more. (To be clear, nobody denies that psychophysics investigates and describes a set of discriminative capacities; the argument is rather about whether or not that is *all* that it describes.) Often, the limiting claim is motivated by a tacit commitment to the legacy of behaviorism and its attempt to make psychology a 'properly scientific' field of inquiry. (Indeed, 'discriminative capacities' is just another way of saying 'behavioral dispositions'!) To be sure, it was the development of psychophysics under practitioners such as Fechner (1877, 1882) and Stevens (1951, 1975) that lead the way in making psychology scientifically respectable. However, the 'behaviorist' interpretation of the subject matter of psychophysics would not have been one that those theorists would have endorsed: both seem to have thought that psychophysical laws related stimulus properties to genuine *phenomenological* properties (Horst, 2005). Alternatively, some argue that phenomenology cannot be the subject matter of psychophysics since discriminative capacities outrun conscious awareness. This is indeed true; however, it does not pose a problem for either psychophysics or quality space theory more generally. I will return to this issue again at the end of the chapter.

Importantly, in order to construct a quality space on the basis of behavioral data, the data must constrain the structure such that there is a *unique interpretation* of that data. That is, we require that each quality in the space be distinguishable from every other quality. Only then can we say that any two given qualities are different (or identical). For example, as Clark (1992) points out, "matching" data alone is insufficient to construct a quality space because the relation of 'indiscriminable' is *non-transitive*: i.e., it may be the case that, of three stimuli *x*, *y*, and *z*, (i) *x* is indiscriminable from *y*, (ii) *y* is indiscriminable from *z*, but (iii) *x* is nevertheless discriminable from *z*. If there is a stimulus *z* that can be distinguished from *x* but not from *y*, it follows that *x* and *y* cannot present the same quality—there must be some feature of their appearance properties which distinguishes *x*'s from *y*'s, and thus *x* and *y* cannot be qualitatively identical. In order to have qualitative identity, we require something stronger: *global indiscriminability*.

Nevertheless, a unique ordering can be generated from matching data with the addition of simple rules. For example, Carnap (1928) and Goodman (1977) both rely on *betweenness* (or *betwixtness*) as their central relation, using (roughly) a rule of the form: *if* x *is indiscriminable from* y *and* y *is indiscriminable from* z, *but* x *is discriminable from* z, *then* y *is 'between'* x *and* z. By application of this rule, a unique ordering of the qualities can be generated. Similarly, a unique ordering can be generated from relative similarity data using a statistical technique known as *multidimensional scaling* (MDS), which is used to discover the structure of similarity relations in data (including its dimensionality). Clark (1992) describes MDS in terms of maps of inter-city distances: given a map of the major cities in North America, one can easily measure the distance between each city. MDS proceeds in the reverse direction: given a list of measurements of inter-city distances, it reconstructs the map. Likewise, given psychophysical data about relative similarity judgments, MDS can construct a quality space for that sensory modality.[66]

---

[66] It's important to note that MDS is somewhat more limited in the case of quality space than in the case of the map. Specifically, in the map case, measurements values are *metric*, in the sense that the units are in miles or kilometers, and 0.5 is twice the value of 0.25. In the case of constructing a quality space, however, all we having is *ordinal* values. That is, in psychophysical ranking of relative similarity, we might know that x is more similar to y than to z, but we do not know the *degree of similarity*. (Analogously, it's like

Of course, different types of behavioral data will produce psychophysical quality spaces with different structures. Likewise, different types of ordering relations applied to that data will affect the structure of the resulting quality space in different ways. Thus, one might wonder about which quality space 'really' captures the structure of the subject's discriminative capacities. However, this issue isn't something that should worry us.

For one, it's important to note that the psychophysical data is inherently *statistical*. Specifically, sometimes a subject might report that two particular stimuli are indiscriminable, sometimes not. Sometimes subjects may swear that two stimuli are completely indistinguishable, yet in forced-choice sorting trials they can reliably sort the two stimuli into different categories at a probability level well above chance. Or, a subject may on one occasion rate *x* as being more similar to *y* than to *z*, while doing the opposite on another occasion. The trick, as Clark (2000) puts it, is to "assess whether that distribution of responses differs from one produced by chance alone." (p.5)

In other words, the topology of a psychophysically-defined quality space is inherently probabilistic, in the sense that its structure is the product of averaging over many different measurements, which may in turn involve different procedures. Thus, although the use of different experimental procedures, measurement types, and constructional techniques may result in differently structured spaces, these should simply be regarded as different sources of evidence which can be used to give us a clearer and richer (albeit probabilistic) characterization of the actual structure of the quality space for that modality.[67] Indeed, as

---

knowing that Toronto is closer to New York than Vancouver, but not knowing how much closer it is.) As a result, the constraints on the ordering of the stimuli are much weaker when one can only use ordinal rather than metric measurement values. (For example, in a non-metric quality space, the fact that red is twice as far from green as from yellow does not entail that red is twice as similar to yellow as it is to green.)
Of course, one can use metric MDS—for example, an experimenter might ask subjects to provide numerical values which correspond to the similarity of stimuli ("magnitude estimation"), either by simply asking the subjects to rate the similarity of a pair of stimuli on a scale from, e.g., 1-9, or by providing a sample pair and stipulating a value for its similarity (e.g., "7"), which can then be used by subjects as a benchmark for rating the similarity of other pairs. However, the validity of these numbers is questionable, and furthermore they cannot be used on non-human animals.

[67] An important methodological question arises at this point: *how do we know that different methods are measuring the same set of discriminative capacities?* For example, consider colour perception: how can we be sure that relative similarity judgments, matching judgments, and JNDs all reflect the 'same' discriminative capacity (i.e., the capacity to discriminate colour / chromatic stimuli)? Briefly, the answer to this is that there is a methodological assumption at work: namely, that when the same set of appearance

Wimsatt (2007) points outs, the use of multiple independent methods of determination or access to a particular phenomenon or result leads to a much more reliable and "robust" theory: for instance, not only can a multiplicity of evidential sources help us distinguish signal from noise (i.e., disambiguate real features of phenomena from artifacts caused by a particular measurement procedure, model, or perspective), but moreover, different sources of evidence can also act as independent verifications of the reliability of other methods. In short, as with all good science, evidence from multiple sources is a virtue, not a vice (even when that evidence can be contradictory).

Furthermore, as Hardin (1988) points out,

> "It should be understood that the issue is not that there cannot be a consistent scheme of representing phenomenal color, for there can. And indeed, there is reason to suppose that the various representations can be mapped into each other (though the mapping relations will often be quite complex). The point is, rather, that one cannot expect any *single* representation to be serviceable for all purposes, not even if we restrict ourselves to the construction of psychological color spaces rather than, say, spaces whose primary use is in the mixing of colorants." (p.120)

In other words, different sorts of explanatory contexts and different explanatory goals will require different sorts of behavioral measures, and thus, differently structured quality spaces. Furthermore, the choice of ordering relation will be influenced by context-dependent epistemic factors such as efficiency, fruitfulness, simplicity, and consistency.

Let us now stop and take stock of where we are. Using psychophysics we can investigate the discriminative capacities in a given sensory modality for a particular subject. From this data, we can construct a multidimensional ordering of the qualities that subject perceives external stimuli as having—an *appearance property quality space*. Ultimately

---

properties (chromatic stimuli) are stimulating the same types of sensory receptors (retinal cone cells), different types of measurement procedures are targeting the same set of discriminative capacities are at work. Of course, this methodological assumption is a *defeasible* one: for instance, if two different psychophysical measures of colour discriminations reliably produced wildly divergent results, that would be good evidence that they are in fact measuring two different discriminative capacities.

however, the goal is a psychological/neuroscientific model of the structure of sensory phenomenology—a *phenomenal quality space*.

However, from the structure of appearance properties we can extrapolate something about the structure of the subject's internal sensory states. (Clark 1992; Rosenthal 2010) For in order to make the perceptual discriminations that it does, a creature's internal sensory state must have a relational structure that is at least as complex as the relational structure of the appearance properties that it can discriminate. In other words, there must be some property of the creature's internal sensory states in virtue of which the creature perceives external stimuli as standing in the relevant similarity and difference relations, and whose relational structure explains the structure of the perceptual discriminations that the creature makes.[68] These are phenomenal qualities, and they are locations in a phenomenal quality space.

However, it's important to note that, at this stage of the explanation, the extrapolated phenomenal quality space is merely a datum to be explained. That is, a psychophysically-determined phenomenal quality space is an *explanandum*, not an *explanans*. Psychophysics can discover the structure of the quality space for a given sensory modality, yet we still require an explanation of that structure. For this, we turn to neuroscience.

4.5 Methodology pt.2: Neuroscience

Perhaps the most well known example of this type of explanation comes from colour vision: the structure of colour quality space is explained by *opponent processing theory* (Hurvich 1981; Hardin 1988; De Valois and De Valois 1993). According to this theory—which was described above in sections 2.3.1 and 3.3.3—the outputs of the three different kinds of cone cells in the retina are summed and differenced to create three different post-

[68] As Clark (2000) puts it, "The structure of similarity and differences among qualitative properties of sensation must be sufficient to account for the structure of similarities and differences among phenomenal properties." Of course, the structure of internal properties could be *more* complex than the structure of a creature's perceptual discriminations... As Clark puts it, "By no means can there be *less* structure...since such a finding would render miraculous the creature's capacities to discriminate. There could be more structure, but it would be unparsimonious to suppose so." (p.9)

receptor neural channels. The cells in these channels can assume different firing rates, either above or below their 'base' firing rate, and this difference can be referred to as the channel taking a 'positive' or 'negative' value. Two of the three opponent process channels codes chromatic information: one channel corresponds to a red-green opponent relation, whereas the other corresponds to a yellow-blue one. The third channel is an achromatic channel that codes for lightness.

By representing each of the three channels of the opponent processing system as a spatial dimension, every possible state of our opponent processing channels can be understood as corresponding to a point in three-dimensional colour quality space that is discovered by psychophysics. That is, the 'value' of each channel (as determined by its level of activation) can be regarded as a Cartesian coordinate along a particular dimension of qualitative variation within the space.[69,70]

Furthermore, opponent process theory has received some direct neurophysiological support: post-retinal opponent processing cells have been discovered in the LGN that roughly correspond to red-green and yellow-blue dimensions of colour space, although changes in their activity do not directly correspond to changes in the phenomenology of colour experience (De Valois & De Valois, 1993). Similarly, colour-sensitive opponent processing cells have been discovered in various areas of the visual cortex, including V1 and V4 (Zeki, 1993). And while no direct neural correlates of colour experience have been discovered, there is nevertheless good evidence that the brain uses opponent processing mechanisms to encode chromatic information.

Now of course, opponent process theory is just that—a *theory*. It is an as-of-yet unconfirmed empirical hypothesis, for which there is some promising evidence.

---

[69] Note that this isn't quite accurate: by plotting each channel as a spatial dimension, we get the so-called "color cube". However, our opponent processing states typically cannot occupy every possible position within that space, except under abnormal conditions, and then only briefly. The locations in the cube which our opponent processing state can occupy under normal circumstances and for extended periods of time are within a sub-space of the cube: the colour solid.

[70] Similarly, the gustatory (taste) system can also be construed in terms of a quality space built on four dimensions, corresponding to the four different kinds of taste receptors in the mouth. Any possible taste sensation can be characterized as a point in four-dimensional space, as determined by the relative activation levels of those receptors. (Churchland, 1989). Parallel proposals apply to other sense modalities as well.

Nevertheless, opponent process theory does provide a scientific explanation for the structure of sensory phenomenology, by proposing mechanisms whose activity generates that structure.[71] (Or more accurately, by offering theoretical models in which the psychophysical data can be embedded, by virtue of the model having a relational structure that is isomorphic to that of the quality space.) In the next section I will examine in more detail how QST fares as scientific explanation by demonstrating how QST fits into the tradition of *structuralism* in the philosophy of science, and *mechanistic explanations* in neuroscience. However, before turning to those issues there are two additional points to note about the importance of providing a neurophysiological interpretation of a quality space.

First, it's extremely important to note that neuroscience does more than simply provide an explanation for the structural features of the space. For in addition, it also discovers the dimensions along which sensory information is actually encoded in the brain. In other words, although psychophysics can provide us with the number of dimensions of qualitative variation in a given quality space (e.g., that colour space has three dimensions), it cannot provide an *interpretation* for those dimensions. It cannot tell us the *coordinate scheme* by which the brain encodes qualitative variation. For example, in the appearance property hue circle (a two dimensional cross-section of three-dimensional colour space) one typically uses a *polar* coordinate scheme, with hue as the angular coordinate and saturation as the radial one. However, that's not how the brain encodes chromatic information: rather than hue and saturation dimensions, the brain encodes chromatic information using the red-green and yellow-blue dimensions of qualitative variation described above in the discussion of opponent process theory—in essence, a type of Cartesian coordinate system. Without a description of the coordinate scheme by which the brain encodes information in a sensory modality, an explanation of phenomenal qualities would be necessarily incomplete. We might know that the quality

[71] Furthermore—as was mentioned above in section 2.3.1—opponent processing theory not only explains the structure of colour space, but also a number of different phenomena related to the phenomenology of colour perception, including the unitary-binary distinction, complementary after-images, types of deficiencies in colour vision, the Bezold-Brucke phenomenon, and many more. (Hardin 1988; Clark 1992; Pautz 2006; etc.)

space for a given modality contained *n* dimensions of qualitative variation, but we would not know what those dimensions were.[72]

Second, a neuroscientific explanation of a quality space must provide not only an account of the basic structure of the space, but in addition, it should ideally also provide an account of the *dynamics* of state-to-state transformations within the space. Specifically, it should explain how changes to the input of the mechanism (either via external influences on receptoral input or via internal influences from other neural systems) affect transitions from one location in the space to another. Churchland (2005) has provided a detailed account of how opponent process theory explains (and predicts!) such state-to-state transitions in colour experience, and Akins (1996) has provided a meticulous examination of the (nonlinear) response function in thermoreception and its implications for the dynamics of the phenomenology of temperature perception.[73]

4.6 QST as Philosophy of Science: Structuralism and Mechanistic Explanation

As mentioned above, quality space theory is committed to the view that neuroscientific theories explain the structure of psychophysically-determined phenomenal quality spaces by proposing mechanisms whose activity generates that structure. QST is thus an example of *mechanistic explanation* (Bechtel & Abramson 2005; Bechtel 2007; Craver 2007, 2008). Roughly, a mechanism is composed of a set of entities—component parts—whose activities, organization, and causal interaction is sufficient to produce the phenomenon that it is posited to explain. Mechanistic explanation involves constructing a

[72] In other words, the structure of a particular phenomenal quality space can be generated/realized in many different ways, but we want to know how the brain *actually* does it.

[73] In essence, thermoreceptors have two different types of functions—static and dynamic—neither or which involves a linear response to stimulus temperature. For example, the static function of "hot" receptors—those that are responsive to heat—can be described by a curve that plots the firing rate of the neuron against stimulus temperature (the relationship of which is non-linear). On the other hand, the dynamic function of hot receptors is sensitive to temperature *change*, to which it responds with a burst of activity, the magnitude of which is determined by the initial starting temperature before the change (such that, e.g., the higher the starting temperature, the greater the initial burst of activity). Parallel considerations apply to the "cold" receptors as well. In any case, what's important to note for the purposes at hand is that the transition between locations in our temperature quality space is explained by the particular response properties of the neurons which are sensitive to thermal stimuli.

model of the mechanism whose structure (either in whole or in part) is isomorphic to the structure of the target phenomenon.[74]

Opponent process theory provides an example of this sort of explanatory schema: it provides a detailed model of a neurophysiological mechanism—post-retinal neural channels containing opponent processing cells of a certain well-defined type—that underlies our capacity to discriminate colour and whose internal states have a relational structure that is isomorphic to the structure of the psychophysically-determined colour space (the explanandum phenomenon).

It is worth noting that quality space theory is thus an example of *structuralist explanation* in the philosophy of science. The basic idea behind structuralism is quite simple: science tells us only about the structure of the world, and thus our scientific knowledge is, in some sense, knowledge about structure. The version of structuralism that I'm concerned with here is a thesis about scientific explanation: according to this view, scientific theories explain phenomena by offering abstract mathematical structures in the form of theoretical models, within which the structural relations of the phenomena can be 'embedded', by virtue of an isomorphism between the structure of the model (or some part of it) and the structure of the phenomena.[75,76]

---

[74] It should be noted that I am endorsing an *epistemic* view of mechanistic explanation here, in contrast to, e.g., Craver's (2007, 2008) ontic version wherein 'explanations' are the actual physical parts of the world that are causally or constitutively related to the phenomena. In contrast, on the epistemic view, 'explanations' are a kind of representation of those things.

[75] It is worth noting however that the type of structuralist explanation that is employed by quality space theory is importantly different from the type of structuralist explanation that was (and often still is) typically endorsed by proponents of structuralism. Specifically, the 'traditional' structuralist view of science tends to endorse a *nomological* view of scientific explanation. (see, e.g., Carnap 1928, Friedman 1999) On this view, it is the goal of scientific theories to formulate *laws*, which can then be used to explain phenomena. Historically, this view was epitomized by the deductive-nomological (D-N) model of scientific explanation, which held that an explanation was a deductive argument with at least one law of nature among its premises. However, the well-known deficiencies of the D-N model combined with a general skepticism about the universal applicability of the nomological view eventually lead to the search for alternative conceptions of scientific explanation. Specifically, though the nomological view of explanation might be appropriate for the most fundamental level of physics, it very often fails to conform the type of explanation typically found in higher-level sciences such as biology and cognitive science. (Bechtel 2007, Craver 2008)

[76] Todorović (1987) contrasts "isomorphist" theories that require a structural match between the neuroscientific model and the psychophysical data with "nonisomorphist" theories that do not. Clearly, this proposal falls squarely within the former camp.

Structuralism has a long history in philosophy of science; in modern form, its origins can be traced back to at least the logical empiricists (e.g., Russell 1927; Carnap 1928). And indeed, the history of quality space theory has strong connections to the project of logical positivism: much of Carnap's radical empiricist reduction in the *Aufbau* is focused on the domain of sensory experience, wherein he uses a procedure that he calls "quasi-analysis" to derive a multidimensional order from "pair lists"—i.e., psychophysical matching data. In other words, he was constructing quality spaces. Of course, Carnap's motivation to construct quality spaces was part of his phenomenalistic reduction—the attempt to translate all empirical statements into statements about actual or possible sense experience. Clearly, that is not the goal of this project; nevertheless, QST shares with logical positivism not only a commitment to structuralism, but also a commitment to a particularly strong form of *empiricism* about phenomenal qualities.

Indeed, QST is predicated on the very idea that there even can *be* such a thing as an "objective phenomenology" – a notion that some would deny (e.g., Chalmers – see discussion in the following section). More importantly however, quality space theory rejects certain traditional debates about qualia as fundamentally misguided, in a sense that is not unlike the positivist rejection of metaphysics. Specifically, it rejects the idea that there can be differences in qualia that make no empirical difference (zombies, inverted qualia, epiphenomenalism, etc.), as well as the idea that there are facts about qualia that are unknowable *in principle* from a third-person perceptive.[77]

Nevertheless, the worry persists that the theory remains *incomplete*. Namely, QST seems vulnerable to the criticism that—while useful—the sort of structuralist explanation that it employs necessarily *leaves something out*. I will examine this issue and the relationship between QST and traditional problems in the philosophy of mind in more detail in the next section.

[77] Indeed, the 20th century saw philosophy of mind become mired in unhelpful debates about metaphysics, appealing to arguments which relied almost exclusively on intuitions, thought experiments, conceptual analysis, modal arguments, and all sorts of other bad methodological tools. In many ways, QST in its modern form is a direct response to these debates.

4.7 Does Quality Space Theory 'Leave Something Out'?

Consider again Nagel's call for an "objective phenomenology":

> "This should be regarded as a challenge to form new concepts and devise a new method—an objective phenomenology not dependent on empathy or the imagination. *Though presumably it would not capture everything*, its goal would be to describe, at least in part, the subjective character of experiences in a form comprehensible to beings incapable of having those experiences." (p.449, emphasis added)

As indicated in the emphasized text, Nagel seems to believe that even once we possess an 'objective phenomenology', there would nevertheless still be something that scientific explanation leaves out. This is not an unusual position to find among philosophers; rather, it's a commonly held view that empirical accounts of sensory phenomenology are doomed to be incomplete. Perhaps the most well-known proponent of this view is David Chalmers. As he puts it,

> "In general, certain facts about structures found in [neural] processing will correspond to and arguably explain facts about the structure of experience. This strategy is plausible but limited. At best, it…allows us to recover structural properties of experience from information-processing properties, *but not all properties of experience are structural properties.* There are properties of experience, such as the intrinsic nature of a sensation of red, that cannot be fully captured in a structural description. The very intelligibility of inverted spectrum scenarios, where experiences of red and green are inverted but all structural properties remain the same, show that structural properties constrain experience without exhausting it." (1995, original emphasis)

This view is not confined to philosophers; the psychologist Stephen Palmer (1999) argues that empirical approaches to sensory phenomenology are limited by what he calls the *isomorphism constraint*: "Objective behavioral methods can specify the structure of experience up to the level of isomorphism, but no further." (p.941) Furthermore, he goes on to claim that

> "…the nature of individual experiences (beyond the isomorphism constraint) cannot be known from behavior, even in principle. These *nonrelational* aspects of experience lie, by definition, outside the domain of functionalism [i.e., empirical science]" (ibid., emphasis added)[78]

The central thing to note about these passages is that they make a strong claim about the nature of phenomenal qualities: namely, that they are *non-structural* and *non-relational*. In other words, QST is incomplete because phenomenal qualities are presumed to possess an 'intrinsic nature' that goes beyond their structural features. Again, this reflects the traditional view of 'qualia' as intrinsic properties.

However, quality space theory—at least the radically empirical version of it that I'm interested in here—amounts to more than the mere claim that phenomenal qualities stand in certain relations of similarity and differences. Rather, it makes the stronger claim that phenomenal qualities are *entirely* structural, in that their identity is exhaustively determined by their relational structure. So in stark contrast to the traditional view of qualia as intrinsic properties, QST holds that they are in fact relational.[79]

So how do we settle this debate? Consider again the above quote from Chalmers. It is important to note that the argument for the view that phenomenal qualities possess an intrinsic, non-relational "nature" depends on considerations of *conceivability*.[80] In this particular instance, Chalmers uses the conceivability of an empirically undetectable

---

[78] Interestingly, Palmer does seem to think that by establishing an inter-personal equivalence class of the neural correlates of sensory experience, one could have a kind of third-person knowledge about the supposedly extra-empirical "nature" (his term) of another individual's experience. (p.938) Specifically, if individuals A and B are both established to be in the 'same' neurophysiological state (as defined by the equivalence class), each knows what the qualitative "nature" of the other's experience is like, though this knowledge is *not* available to an outside observer. However, given Palmer's commitment to the claim that sensory experiences possess an intrinsic, non-relational "nature" that cannot be specified beyond the level of isomorphism, it's not clear what sort of evidence would justify this claim. In other words, why should isomorphism between neural states be sufficient for sameness of qualitative experience whereas functional isomorphism is not?

[79] As Clark (2000, p.22) puts it, "To the complaint that 'this method [quality space theory] does not explain the intrinsic nature of colour experience' (Chalmers 1996: 235), my response is 'Guilty as charged, Your Honour'. There is no such nature to be explained."

[80] Or in Chalmers' terminology, the argument depends on the "*intelligibility*" of inverted spectrum scenarios.

spectrum inversion to argue that science 'leaves something out'. (Similarly, Palmer (1999) also appeals to inverted spectrum scenarios to support the same conclusion.) However, other sorts of conceivability arguments are routinely deployed to support the same view (e.g., the possibility of philosophical zombies).

However, as Rosenthal (2010) points out, these sorts of arguments *presuppose* that there can only be first-person access to phenomenal qualities. For example, the possibility or conceivability of philosophical zombies depends on the view that the presence or absence of sensory phenomenology can only be determined by conscious introspection. If, for instance, one could know about phenomenal qualities independently of consciousness, then zombies obviously could not be possible. However, on quality space theory, phenomenal qualities are simply the properties of sensory states by virtue of which we make the perceptual discriminations that we do. Thus, quality space theory presupposes that there can be third-person access to phenomenal qualities.[81] Of course, this response is very unlikely to satisfy those sympathetic to conceivability arguments; however, I think there are additional independent reasons to reject the claim that phenomenal qualities are intrinsic properties, which I will discuss below.

The possibility or conceivability of empirically undetectable inverted spectrum scenarios similarly depends upon the view that there are facts about phenomenal qualities that are only available via first-person conscious introspection. However, it too is ruled out by QST, and for reasons that are worth exploring in detail.

As Palmer (1999) points out, the possibility of undetectable spectrum inversion requires that quality space must be *symmetrical* in some way.[82] However, QST individuates locations in a quality space (i.e., phenomenal qualities) wholly in terms of their relative position to other such locations in that space. Thus, in order for a minimal form of

[81] In other words, QST makes the methodological assumption that perceptual discriminations are made on the basis of phenomenal qualities. It is this methodological assumption that allows one to extrapolate a phenomenal quality space from an appearance property discrimination space.
[82] As he puts it, "whether there exist any undetectable color-to-color transformations can be recast into the simple question of whether an empirically accurate model of human color experience contains any symmetries." (p.924)

undetectable spectrum inversion to be possible, there would have to be at least two locations in the space that bear exactly the same pattern of relations to all other locations in the space. *However, if this were the case, we would have no way of distinguishing between those two locations*.[83] Furthermore, the same moral applies to the *dimensions* of the space: as Rosenthal points out, "if there were an axis with respect to which that quality space is symmetrical, it would be impossible to distinguish stimuli on one side of that axis from stimuli on the other." (p.380)

Furthermore, it's not even clear that the idea of two locations in quality space having exactly same relational structure is even a coherent one. Consider a particular colour quality—say, a certain shade of orange. That colour quality bears certain similarity and different relations to all other colour qualities—e.g., orange is similar to red in certain ways and dissimilar in other ways; likewise to yellow, green, and blue. The possibility of spectrum inversion requires that there be a distinct hue—call it 'not-orange'—that is qualitatively distinct from orange while nevertheless *bearing exactly the same structure of similarity relations to red, yellow, green, and blue.*[84]

What's more, the empirical evidence suggests that quality spaces are not symmetrical. For example, in colour space, certain hues can become much more saturated than others—the most saturated yellow is much less saturated than the most saturated red. Thus, the (radial) distance from achromatic grey to red is much greater than from achromatic grey to yellow. (Hurvich 1981) Furthermore, quality spaces tend to vary in 'density'; that is, not all qualities are evenly spaced, and there may be no common unit to express equivalence of distance across different dimensions. Of course, this complex asymmetry is not particularly surprising when one considers that the structure of quality space is a product of our biology, whose results tend to be "messy".

[83] Put somewhat differently, it is impossible to pick out a particular location in a symmetrical space by a structural description alone. For example, consider the simplest kind of space: one that has exactly two positions. It is not possible to distinguish between those two positions purely on the basis of how they are related to each other, since each bears exactly the same relation to the other.

[84] Furthermore, it's important to note that if two locations in the space did have the same relational structure, global indiscriminability would fail and thus we would not have a *unique interpretation* of the space, as is required by QST.

The upshot of all this is that “conceivability” arguments that are meant to establish that phenomenal qualities possess a non-relational, ‘intrinsic’ nature do not hold up under scrutiny. Indeed, as Rosenthal (2010) points out, the problem with these arguments is that they take a pretheoretical intuition—such as the possibility of zombies or undetectable spectrum inversion—as a datum that must be accommodated by a satisfactory theory of phenomenal qualities. Thus, the sense in which these phenomena are ‘conceivable’ is more like the sense that it’s conceivable that water is a simple, basic physical substance. That is, the claim may only be said to be ‘conceivable’ *in the absence of a scientific theory of water*.

Nevertheless, even if one grants both that phenomenal qualities are exhausted by their relational structure and that there can be third-person access to phenomenal qualities, it is difficult to deny the intuition that there is still something that empirical science is leaving out. To return to Nagel’s classic example, even with both a detailed map of the structure of a bat’s echolocative quality space and a detailed neuroscientific account of mechanisms which explain that structure, it seems difficult to deny that there would nevertheless still be something that we don’t know about what it’s like to be a bat.

And indeed, it’s true: on QST, *there is a sense in which we do not know what it’s like to be a bat*. However, this admission does not present a problem for the radical empiricism to which quality space theory is committed. Specifically, it does not entail the existence of some metaphysical ‘fact’ that is forever beyond the reach of empirical science. Rather, the sense in which we don’t know what it’s like to be a bat is a harmless one—we do not share the bat’s quality space and thus we do not have the ability to pick out locations in that quality space by first-person means.

Clark (2000) has argued that experiencing phenomenal qualities from a first-person point of view has an essential indexical component that involves a kind of demonstrative identification; a way of picking out a location in quality space in a direct, immediate way that is only available to creatures that have the proper neural machinery. Nevertheless, the location in quality space that is picked out via this first-person demonstrative identification is that same one that we can refer to via neuroscience and psychology. In other words, through empirical science we are studying the same set of facts that are available from the bat's point of view. Thus, there is a sense in which we don't know what it's like to be a bat because, as Clark puts it, "to engage mechanisms allowing unstudied direct observational use of the term, one must have some actual historical episode of a successful demonstrative identification." (p.31)

To illustrate the idea with an analogy, Clark appeals to the so-called "Ozma" problem. (Gardner, 1991) To quote Clark at length:

> "This is roughly the problem of defining left and right. These terms are part of a family (including clockwise and counter-clockwise, east and west, north pole and south pole), any one of which can be defined in terms of the others, but all of which seem to rely ultimately on some successful demonstrative identification. In particular, suppose we begin receiving transmissions from a planet on the far side of the galaxy (so far away that no stars are mutually observable, or at least we cannot tell from the descriptions that they are mutually observable. The aliens have the terms 'lana' and 'rana' which we know to mean left and right, but we don't know which is which. Similarily, they also have rotational terms kana-wise and counter-kana-wise, directions eana and wana, planetary poles nana and sana, and we know that these stand for one or the other of our cognate notions, but we don't know which is which." (p.35)

We can learn *all* of the facts about the geography of the alien planet—for example, that location *x* is 20km eana of location *y* and 40km sana of location *z*, or that location *q* has a latitude and longitude of 19.73 degrees nana and 155.04 degrees wana. However, if you were suddenly transported to this planet and told to make your way to a certain location, you would be unable to do so. For in order to know which way to start walking, you would require a kind of *indexical* knowledge—you need to know that eana is *that* way, or that *this* is your lana hand and *that* is your rana hand. Once you have made a successful demonstrative identification of one of these terms, you can understand the rest, and thus figure out which way to go.

Similarly, just as there are no facts that we don't know about the geography of the alien world prior to our arrival, there are no facts that we cannot know in principle about the phenomenology of bat experience. To use Nagel's (1974) terminology, the "subjective facts" and the "objective facts" *are the same set of facts*. We simply have a different way of accessing them. Or to take a different example, consider Jackson's (1982) 'knowledge argument'. Clark (2000) argues that prior to her release from the black and white room, Mary the colour scientist is in precisely the same epistemic situation as the individual who is suddenly transported to the alien world – both know all the facts but lack what Clark (2000) calls the 'essential indexical component'; i.e., an actual historical episode of successful demonstrative identification. When released from her room, Mary does not learn some new "non-physical" fact (the conclusion the knowledge argument was intended to establish). Rather, she simply learns a new way to identify an old fact.

But what precisely is this "indexical component" that Clark refers to? What exactly does it mean to engage mechanisms that allow for demonstrative reference a location in quality space in the sort of direct, immediate way required by the above view? I argue that the answer lies in *consciousness*.

4.8 QST and Consciousness

Phenomenal qualities are the qualitative features that we are presented with in conscious sensory experience. However, phenomenal qualities are *not* necessarily conscious qualities—they can and do occur unconsciously. (Indeed, as Rosenthal (2010) points out, it's important to note that the construction of a quality space relies solely upon a subject's awareness of the stimuli, and not on their conscious awareness of the phenomenal quality of the experience.)

Indeed, quality space theory will often outrun conscious access to phenomenal qualities: for example, a subject might not be conscious of the difference between two colour stimuli (in the sense that they *report* that the two stimuli are introspectively indistinguishable), while their forced-choice sorting behavior nevertheless demonstrates that they can and do reliably distinguish between them. In other words, our discriminative capacities often exceed our conscious awareness of the qualities by which we make those discriminations.

This suggests that our conscious awareness of phenomenal qualities is often not as of discrete points in a quality space, but rather as regions or volumes with fuzzy boundaries. Through training we can learn to distinguish points more finely – consider, for example, the sophisticated wine taster versus the novice to whom it all 'tastes the same'. Likewise, we probably discriminate locations in quality spaces more finely when attentional processes are involved.

However, it is important to note that there is a sense in which quality space theory does not solve Chalmer's "Hard Problem", nor can it give a complete answer to Nagel's question "*what is it like to be a bat*?"; in order to do so requires a theory of consciousness, as described in section 1.3. However, QST (at least as I'm presenting it) makes no claims to provide such a theory. *QST is a theory of phenomenal* qualities; *it is not a theory of phenomenal* consciousness.[85] Recall the distinction made in section 1.3 between "phenomenology" and "consciousness". The former is a "first-order" (Lycan, 2000) property of sensory states, that of possessing a phenomenal quality. The latter is a "higher-order" property of such states, and describes the *conscious awareness/expereince* of being in a sensory state that has a particular phenomenal quality.

This is why in order to fully explain 'what it's like' to undergo a conscious sensory experience, quality space theory would ultimately need to be supplemented by a theory of consciousness, as described in chapter 1. Rosenthal (2010) argues for a similar role for consciousness in his analysis of quality space theory. According to his HOR theory of consciousness, "If one is not in any way aware of a [phenomenal] quality, there is nothing

---

[85] Thus, I actually agree with Chalmers when he says that QST "is useful for many purposes, but it tells us nothing about why there should be experience in the first place." (1995, p.10) Indeed, QST *by itself* does not even qualify as an explanation of the "Hard Problem" in Lycan's 'first-order' sense described in section 1.3, because although it defines phenomenal qualities as certain structural properties of sensory states (those by virtue of which we make the perceptual discrimination that we do), it is technically neutral with regard to the question of *what kind of properties* they are. Thus, QST can only explain the *existence* of phenomenology given some additional metaphysical claim, such as, e.g., representationalism, identity theory, or functionalism.

it's like for one to be in a state that exhibits that [phenomenal] quality. There being something it's like for one to be in a qualitative state is due to the HOA in virtue of which one is aware of that state." (p.383)[86]

Indeed, it seems perfectly obvious that there is nothing 'that it's like' (in the conscious, higher-order sense described in section 1.3) for a sensory experience to have a particular phenomenal quality in the absence of consciousness of the quality of that experience. However, this is nevertheless compatible with the claim that the subject is undergoing a phenomenal experience in Lycan's first-order sense—i.e., being in a sensory state that has a qualitative phenomenology.

Nonetheless, given this distinction between consciousness and phenomenology, some might wonder about whether or not we are justified in calling phenomenal qualities "phenomenal" at all. However, there are at least two reasons that we should consider these features as 'phenomenal' despite the fact that they can occur unconsciously: the first follows directly from quality space theory, the second comes from Block (2007), who has argued that the neural basis of phenomenology does not constitutively include the neural basis of consciousness.

First of all, as described above, in QST phenomenal qualities are identified in part by the role they play within a system, and that role is one which is strongly correlated with phenomenology. To be sure, they cannot play this role without the rest of the system; nor do they even 'count' as phenomenal qualities in the absence of the rest of the system.

For example, Palmer (1999) objects to quality space theory using the example of a "color machine": a device that processes light in a way that is functionally equivalent to the human opponent processing channels. A machine of this kind could certainly be constructed (Palmer's hypothetical machine uses only prisms, cardboard masks, photocells, and electronic circuits), and could even be designed to output color terms, make comparisons of similarity and difference between inputs, and so forth. However,

[86] Note that Rosenthal's invocation of the expression "[what] it's like" here should obviously be understood in the higher-order sense.

Palmer rightly judges that it would be absurd to think that the machine's internal states are *phenomenal* states, or that they possess phenomenal qualities.

As Viger (1999) points out, the proper response to this objection is simply to note that Palmer's color machine vastly underestimates the complexity of the functional role required for producing phenomenally conscious experiences.

> "It is only by various mechanisms playing a functional role within a very complex system that experience can arise at all, and in such cases it is the *entire system* that is the experiencer. […] Once we expose the requisite complexity for a system to be experientially isomorphic to humans vis-à-vis color experience, the force of Palmer's thought experiment evaporates." (p.975)

Indeed, it is only by virtue of being embedded within a larger system that a mechanism can even be identified as playing a particular role; for example, recall that QST identifies phenomenal qualities as the properties of sensory states *by virtue of which we make perceptual discrimination*. What's more, mechanistic explanation is inherently multi-level, insofar as it must not only model the internal organization of a mechanism's component parts (and their causal interactions), but moreover it must take into account the organization of the system in which the mechanism is embedded, including how the mechanism causally interacts with other parts.

However, it's important to realized that—as Block (2007, 2011) argues—although the rest of the system is *causally* necessary for sensory states to possess phenomenal qualities, it does not play a *constitutive* role in generating the phenomenology of the experience. (In fact, the conclusion that Block seeks to establish is substantially stronger than the claim that low-level sensory states simply have phenomenal qualities. Rather, he argues that such states are in fact *phenomenally conscious* states, and denies that *cognitive accessibility* or *reportability* is necessary to produce phenomenal consciousness. (In short, Block argues for the existence of phenomenally conscious experiences of which the subject is unaware.) However, I am using Block's general argument here merely to illustrate a weaker claim—that the sensory states in question possess a qualitative character. This claim is certainly less contentious, as even Rosenthal (2005, 2010) endorses a limited version of it despite his insistence that only sensory states that are target of a HOT are phenomenally conscious. Thus, the following discussion of Block's argument is greatly oversimplified, as it is intended to be a mostly exegetical explanation of why we should treat low-level sensory states as genuinely *phenomenal* states, rather than a persuasive argument for Block's preferred conclusion.)

Block's argument comes in two parts: the *overflow* argument and the *mesh* argument. The overflow argument attempts to establish that informational/representational storage capacity of phenomenology (or at least "the visual phenomenal memory system") is greater than that of the working memory system that underlies conscious access (or more accurately, *reportability*). The mesh argument attempts to show that the best neuroscientific explanation of phenomenological overflow is that the neural basis of phenomenal states does not include the neural machinery required for conscious awareness (or rather, for cognitive accessibility).

Block's overflow argument is based the classic Sperling (1960) experiments, in which subjects are presented with a 3x4 grid of letters, flashed very briefly. Subjects report having a visual experience of all the letters, yet can only identify a few of the letters if asked to recall them after the grid has disappeared. However, if cued by a tone that indicates the subject should report on a particular row of letters, the subjects are typically able to identify at least 3 of the letters in that row, *even if the tone occurs after the grid of letters has disappeared*. Thus, it would seem that the subject has detailed, low-level sensory representations that not only persist after the stimulus has disappeared, but whose specific contents also can exist without conscious awareness of them.[87]

Importantly however, the overflow argument by itself is not sufficient to establish the existence of un-accessed, unconscious phenomenal states (as Block himself acknowledges). That is, it could still be the case that the neural basis of phenomenology constitutively includes being cognitively accessible and available to consciousness.

---

[87] Indeed, it would seem almost ludicrous to claim that there was specific phenomenology only *after* the cue, but not before, especially given that the cue occurs after the stimulus has already disappeared.

However, Block argues that the neuroscientific explanation for phenomenological overflow necessitates that we should treat the respective neural bases of phenomenology and conscious accessibility as distinct. (This is the crux of the "mesh" argument—showing how results from neuroscience 'mesh' with results from psychology.)

Specifically, Block argues that the type of cognitive accessibility required for conscious awareness of the contents of a sensory state is mediated by connections between prefrontal areas and areas dedicated to lower-level sensory processing. Empirical evidence appears to show that ensembles of neural activation in these sensory areas compete for dominance, such that the ensembles which "succeed" in those competitions are those that manage to activate the prefrontal areas, thereby triggering a synchronized feedback loop between prefrontal areas and the sensory state itself. It is the existence of this recurrent connection that causes sensory states to become conscious. Moreover, these recurrent connections are thus the informational 'bottleneck' between low-level sensory representations and the working memory system that underlies conscious access.

What's particularly important to the "mesh" argument, however, is that there are also neural ensembles in the sensory cortices that narrowly 'lose' the competition for dominance by only a small amount, and fail to establish the recurrent connection with prefrontal areas. As Block (2007) puts it,

> "If we assume that the strong but still losing coalitions in the back of the head are the neural basis of phenomenal states…then we have a neural mechanism which explains why phenomenology has a higher capacity than [the working memory system that underlies conscious access]. If, on the contrary, we assume that the neural basis of phenomenology includes workspace activation, then we do not have such a mechanism. That gives us reason to make the former assumption." (p.498)

What's more, to argue that these 'losing' sensory states aren't phenomenal because they don't trigger prefrontal activity is simply to beg the question – it assumes conscious awareness is necessary for phenomenology without argument. Of course, to reiterate, this is nevertheless compatible with the claim that phenomenology requires the *causal* influence of prefrontal areas—and indeed, one of the main factors which influences which sensory ensembles ‘win’ the competition is feedback from the prefrontal cortex. However, Block’s arguments seem to demonstrate that such prefrontal activation is not *constitutively* necessary for phenomenology.

Thus, there are good reasons to hold that sensory states possess phenomenal qualities, and do so regardless of whether or not those states are phenomenally conscious. Moreover, as the rest of the chapter laboured to demonstrate, these phenomenal qualities are exhausted by their relational structure, which in turn can be reductively characterized and explained in terms of quality space theory. Finally, although there is a sense in which such an explanation inevitably seems to ‘leave something out’, such concerns can be explained away as a matter of different epistemic routes of access to the same set of objective facts, and do not imply that there is some aspect of phenomenal qualities that fails to be captured by QST.

Chapter 5

5 Spatial Representation

In this chapter, I provide an account of the representation of spatial relations in sensory experience. I first describe the importance of providing an account of sensory spatial representation for a theory of sensory phenomenology. I then describe how psychophysics can discover the topographic structure of represented spatial relations in different sensory modalities, and I provide a basic description of the neural mechanisms that (i) underlie our ability to represent space and (ii) which explain the structure discovered by psychophysics: *feature maps*. Then, I examine the need for sensory systems to *coordinate* different maps in various ways, and provide some empirical examples in terms of the sensorimotor integration of auditory and visual spatial information. Finally, I argue that the spatial content of our sensory experiences is the product of such coordinations, and I attempt to provide the outlines of a theory of the (Fregean) spatial content of sensory experience based on this view.

5.1 Spatial Representation in Sensory Experience

Discussion about the phenomenal content of sensory experiences tend to focus almost entirely on *phenomenal qualities*—the colours, sounds, tastes, smells and other such features that seem to give rise to the so-called “Hard Problem” of consciousness (Chalmers, 1996).

However, our sensory experiences represent more than simply isolated features. Rather, they typically represent features (i) as being located at certain spatial locations, and (ii) as features of objects. We don’t simply have sensory experiences of disembodied redness; rather, our sensory experiences represent objects and locations in our environment as instantiating those properties. For example, we experience the red ball as *over there*, in the three-dimensional space surrounding us. We hear a sound (or see a flash of light) as coming from, e.g., as slightly above us and to the right. In other words, our sensory experiences not only have a particular *qualitative* character (like colour or pitch), but moreover they also have a particular *spatial* character.

Someone may potentially object that existence of spatial experience doesn't seem to give rise to the "hard problem" in same way that phenomenal qualities do. The source of this intuition probably has something to do with the difference between secondary qualities (such as colour and smell) as opposed to primary qualities (such as shape and extension).[88] (Locke, 1690) The former were traditionally thought to be 'subjective' in the sense that they only existed in the mind of the observer, whereas the latter were thought to be 'objective', existing independently of any perceiving mind. Admittedly, there is a certain amount of intuitive pull to this objection; whereas qualitative features such as colour and smell seemingly cannot be identified with any response-independent physical properties of objects (due to the problem of bad structural correlation, described in chapter 2 above), spatial properties such as length and distance seem to be directly reducible to objective, perceiver-independent features of the physical world.

However, not only is the nature of spatial experience an interesting and important problem in its own right, but furthermore, the objection brings to light an important point that should not be overlooked: without an explanation of spatial discrimination, an account of the phenomenology of sensory experiences would necessarily be *incomplete*. For even if one were to have the correct theory of phenomenal qualities, such that it could explain the qualitative aspect of sensation, such a theory would still not be able to explain how one can distinguish between two subjectively indistinguishable qualities that are simultaneously presented at different locations. What's more, I will also argue (in this chapter and the next) that we need an explanation of spatial discrimination to *ground the intentionality* of sensory experience.

In what follows**,** I will examine sensory spatial discrimination, using vision and audition as examples. I will first describe each modality separately, and then describe how they are integrated to form a modality-independent, higher-order representation of space.

[88] The use of the term "qualities" here can be somewhat misleading if it is interpreted in a phenomenal sense. Instead, Locke should be interpreted here as talking about different types of *properties*.

5.2 Methodology: Psychophysics & Neuroscience

Methodologically, investigating spatial discrimination involves roughly the same sort of process that is involved in investigating the discrimination of qualitative features such as colour and pitch. That is, one begins by doing psychophysics in order to test the limits of spatial discrimination and to determine the structure or topography of phenomenological space. Then, we can turn to neuroscience to look for the mechanisms which can explain the psychophysical data.

For example, in testing the spatial resolution of the somatosensory system, a common type of task that has traditionally been used is the so-called "two-point discrimination" task, in which an experimenter touches a subject on the skin with either one or two sharp points. The subject must report whether one point or two points were used, and the smallest distance at which the subject can discriminate one point from two at a level sufficiently above chance is known as the "two-point threshold". (Of course, this threshold varies across different bodily locations; at the fingertips and lips, the threshold is as small as 2mm, whereas on the back the threshold can be as high as 40mm.) The addition of other psychophysical measures can provide an even more accurate picture of tactile spatial resolution; for example, the two-point threshold can be supplemented with an orientation discrimination task which measures the ability of the subject to discriminate the alignment of the points (Tong, Mao, & Goldreich, 2013). Similarly, experimenters can also test the ability of subjects to detect the orientation of a grating that is touching the skin (Johnson & Phillips, 1981). The resulting (psychophysically-determined) picture of somatosensory space that develops is the well-known 'homunculus': the grotesquely disproportionate representation of the human body that has excessively large hands, lips, and face relative to the rest of the body.[89] Importantly,

[89] It might seem as though this resulting spatial representation of the body's surface is somehow *inaccurate*; a distortion of its spatial structure. However, a better way to think of it is as defining the 'density' of somatosensory space, or the 'grain of resolution' of our representation of different body parts. Moreover, sensory systems are rarely designed to represent absolute values of stimuli. Rather, as Akins (1996) puts it, sensory systems are designed to be "narcissistic"; to solve particular informational problems, not to represent veridically; i.e., systematically encoding the absolute value of external relations without embellishment.

the proportions of the somatosensory homoculus can be directly explained by neuroscience; specifically, they are a direct result of both the number of sensory receptors in any particular location on the body's surface, and the amount of neural real estate in the somatosensory cortex (and motor cortex) dedicated to processing the signals from those receptors.

Similarly, the unique structural topography of auditory space that is discovered by psychophysics can be explained by the neural mechanisms that underlie our ability to localize sound stimuli in our environment. Contemporary methods of mapping auditory space rely on varying a number of independent factors (such as the azimuth, elevation, and distance of the sound source; the frequency, pitch, intensity, and timber of the sound; the nature of the environment; and so on) along some particular dimension(s) and observing the subsequent effects on the subject's ability to discriminate the spatial location of a sound source. The resulting picture of auditory space has a very specific topographic structure: For example, spatial localization in audition is vulnerable to a particular kind of illusion known as front-back reversals, wherein sound stimuli that are located directly in front of the head and directly behind the head are apt to be confused with one another. [90] The explanation for this phenomenon has to do with the fact that auditory information about the azimuth of a sound source is based mainly on calculating certain types of subtle differences between a sound as it arrives at each ear. When the stimulus is directly in front of or directly behind the listener, those types of differences (which I will discuss in greater detail below) are equivalent, thus leading to front-back reversal. Similarly, humans are much, much worse at estimating the egocentric distance of a sound source than its azimuth or elevation. Again, this can also be explained by the particular processing structure of neural mechanisms that underlie auditory localization (which again, I will discuss in greater detail below). Finally, the accuracy of human auditory localization depends in many ways on the particular spectral frequencies of the

[90] This effect—and indeed, psychophysical investigation of auditory spatial localization in general—has been studied since as early as 1796, when Giovanni Venturi demonstrated that subjects could point in the direction of a sound source. (Venturi demonstrated this by having subjects sit in an open field blindfolded, while he walked around them, intermittently playing notes on his flute and asking subjects to point in the direction of the sound.)

sound, which again is a result of the sensitivity and excitability of particular cell types in the auditory system to sounds in a certain narrow frequency range.

In any case, the general point should now be clear: methodologically speaking, providing an account of the spatial phenomenology of sensory experience requires us to first do psychophysical experiments to determine the topographic structure of the represented spatial relations in a given modality, and then do neuroscience to discover neural mechanisms which can explain that structure. Of course, thus far I have only discussed the uni-modal representation of space; however, one can also do cross-modal experiments in both psychophsyics and neuroscience in order to discover the effects that spatial representations in one modality have on representations of spatial relations in a different modality. And indeed, as I shall argue below, this sort of cross-modal investigation is absolutely crucial to the project of providing an account of the phenomenology of spatial experience, because *the sensory representation of spatial location is inherently multi-modal in nature*. However, for now this preliminary characterization of the methodology will suffice. In what follows, I will focus primarily on the neural underpinnings of spatial localization in order to provide an (admittedly oversimplified) explanation for the structure of spatial phenomenology.

5.3 Spatial Localization and ‘Feature Maps’

One of the most basic mechanisms of spatial discrimination involves what is known as a “feature map”. As described above, a feature map is a cortical structure containing neurons that selectively respond to some particular dimension (or dimensions) of variation in phenomenal features, such as colour or pitch. Moreover, these neural structures are called “maps” because the cells are arranged roughly topographically, so that adjacency relations in the brain mirror adjacency relations in space.

Feature maps are present in nearly all sensory modalities, but the clearest examples of feature maps are found throughout the visual system. For example, V1 famously contains (among other things) a spatial point-to-point mapping of the retina, such that neurons from adjacent points on the retina project to neurons at adjacent points in V1.[91] (Indeed, V1 was at one time called the "cortical retina" by Henschen (Zeki, p.148)) Maps that preserve the topographic organization of the retina in this fashion are referred to as "retinotopic", and they can be found throughout the visual system. Indeed, the visual cortex contains a variety of different topographically-organized feature maps that respond to different visual qualities—hue, orientation, shape, motion, etc.[92]

Furthermore, although these different areas process different kinds of visual information, they strongly modulate one another. For example, V1 and V2 are unique in visual cortex in that they contain cells that respond to all of the different sub-modalities of vision; i.e., they contains cells that are selective for colour, cells that are selective for motion, cells that are selective for orientation, and so forth. Moreover, populations of feature-selective cells at each location on maps in V1 and V2 are strongly inter-connected (in both feed-forward and feed-back fashion) with corresponding locations on maps in other visual areas that process specific kinds of visual features, such as colour, orientation, motion, and so forth.

For example, V4 receives direct projections from V1 and V2 that preserve a rough topographic mapping of the visual field. However, it receives input from only the centreal (foveal) region of V1 (and from V2) and the receptive fields of its cells are overlapping and jumbled. This is because cells in V4 tend to respond most selectively to chromatic features of stimuli, which explains why it only maps the central 40 degrees of the retina—that's where the colour-sensitive cone cells in the retina are located. (Moreover, it has been thought that the large overlapping receptive fields are used to compute global

---

[91] Each hemisphere of V1 receives projections (via the lateral geniculate nucleus, or LGN) from the ipsilateral temporal retina and the contralateral nasal retina, thereby responding to stimuli in the contralateral visual field. In other words, the right hemisphere of V1 receives its input from the outer half of the right eye and inner half of the left eye, and vice-versa. Each side of V1 thereby registers the visual field on its contralateral side. (Zeki, 1993)

[92] However, the topographical organization of spatial relations in visual areas outside of V1 are somewhat more complex; feature maps in these extra-striate areas are known as "second-order" maps or "field discontinuity maps" because adjacency relations are not fully preserved. For example, in V2 the representation of the retina is divided along a horizontal line such that adjacency points on the upper and lower half of the retina do not project to adjacent point in V2. Nevertheless, these maps largely preserve the spatial organization of the visual field—they simply do it in a non-retinotopic way.

colour properties). The topographic organization of V3 is also somewhat messier than in V1 or V2**,** though it is more precise than that of V4. Cells in V3 tend to be selective to *orientation*, and the area is presumed to be important in form perception. Thus, its topographical organization needs to be more precise and of a higher resolution than that of V4, since precise topographic relationships are more important for extracting information about form than they are for surface colour.

By being strongly inter-connected in this way, the visual system is able to both process specific visual features (such as colour and orientation) and also locate them in a spatial array defined by the structure of the sensory receptors—in this case, the retina. Indeed, a spatial map in a given sensory modality is a *feature* map because it is not merely a representation of space, but rather the representation of some particular sensory feature (or set of features) in space.

5.4 Coordination of Maps

It is important to realize, however, that the mere existence of topographically arranged sensory 'maps' does not explain our ability to locate features in space. That is, the simple fact that certain cells in some map (or set of maps) selectively respond to modality-specific stimuli that are in a particular receptorally-defined location does not explain how we *perceive* or *experience* the stimuli as being located in an external three-dimensional space around our body. Nor does it explain how we *represent* spatial contents. In other words, the mere existence of such maps fails to explain the structure of our spatial phenomenology. In order to provide such an account, we first need to examine how the brain *coordinates* different types of maps.

For example, the structure of our phenomenal experience of visual space is not retinotopic, but rather is egocentric. However, topographically-organized retinotopic maps in the visual system cannot locate stimuli in visual egocentric space by themselves, because the same retinal coordinates can correspond to different egocentric locations depending on which way one's eyes are pointing. Thus, one must coordinate information regarding the location of retinal stimulation with information regarding the position of one's eyes in order to begin to locate a stimulus in a visual egocentric space.

Indeed, even the retinotopic map found in V1—the earliest stage of processing the visual cortex—is not simply a mirror image of what is occurring at the retina. Rather, recent work in size constancy (Murray et al., 2006; Fang et al., 2008; Sperandio et al., 2012; Pooresmaeili, et al., 2013) has shown that the degree of eccentricity of activation in V1's retinotopic map more closely reflects the *perceived* size of an image, and not its retinal image size. That is, the size of the image projected onto the retina by a stimulus can remain constant while the perceived size of the stimulus varies widely, with such variations precisely corresponding to the degree of eccentricity in V1 activation. This effect is most likely due to modulation of V1 by other brain areas that process other kinds of visual information (such as linear perspective cues, information about eye position and focus, binocular disparity information, and so on). It is this mechanism which underlies our capacity to perceive a particular object as having the same size irrespective of viewing distance, and to perceive different-sized objects as being of different sizes even when they both project equally-sized images on the retina due to a difference in their respective distances.

This sort of coordinating of different types of visual spatial information is an example of what Grush (2000) calls "stabilization-coordination" (or s-coordination), which involves establishing a relationship between different types of sensory and motor arrays in order to "stabilize" the elements of those arrays for the purposes of forming a higher-order representation of those elements. This sort of coordination typically comes into play when coordinating sensory and motor information for the purpose of constructing a spatial map that is modality-specific. Furthermore, the resulting map is a "higher-order" representation insofar as it involves the construction of a spatial representation whose coordinate scheme is not specifically defined in terms of receptor stimulation.

Grush contrasts s-coordination with what he calls "coincidence-coordination" (or c-coordination), which is the process of coordinating two or more different sensory maps which overlap in their parts or subparts in order to create a higher-order, "virtual" map.[93] This sort of coordination comes into play when, for example, one is coordinating spatial maps in different modalities; i.e., in order to represent the fact that the region of space represented by our visual system and the region of space represented by our auditory system are in fact the *same region of space*. (Or, e.g., that the stimulus in the region of space represented by our auditory system is *to the left of* the region of space that is currently being represented by our visual system.) In one sense, the resulting map is "higher-order" in the sense that it is not tied to any particular modality. However, the map is also "higher-order" in the sense that it may not be actually physically instantiated in any single location in the brain (or anywhere, for that matter). That is, there need not be some topographically organized brain region wherein maps of different sensory modalities are coordinated with one another. Rather, these "virtual" maps exist only as distributed, higher-order representations.

In what follows, I will describe an example of Grush-style multi-modal c-coordination of spatial maps. I have already given a rough sketch of how the visual system *s*-coordinates some kinds of visual information (specifically, the coordination of eye-position with retinal coordinates and the coordination of various types of size constancy information) to 'stabilize' elements of a higher-order map of location in visual egocentric space. Next, I will describe mechanisms of spatial localization in the auditory system, focusing on the way in which different types of auditory information are s-coordinated in order to locate auditory stimuli in space. I will then proceed to show how the visual system and auditory system maps can be c-coordinated (along with *motor* information) to provide a higher-order representation of multi-modal space. (Admittedly, this will be a *vastly* oversimplified account.) Importantly, as we shall see, the crucial ingredient for c-

[93] To illustrate this idea, Grush (2000) suggests the example of c-coordinating a map of California with a map of Oregon: "a map of California and a map of Oregon, so long as each includes at least a bit of the surrounding region, can be coordinated by identifying these regions – the little bit of northern California on the southern end of the Oregon map is identified with the northern California on the California map, etc. One c-coordinates the two partial maps in order to construct a larger, higher order map. This higher order map may be virtual, in the sense that there is no need to actually physically abut the maps. The two component maps might even by at very different scales, and thus impossible to physically join so as to get a viable physical map." (p.67)

coordination of different maps is the existence of a *common frame of reference* that allows the brain to integrate different kinds of sensory and motor information into an inter-translatable coordinate scheme.

5.5 Auditory Spatial Localization

Spatial localization in audition is based on a number of different types of auditory cues. For one, the temporal delay between the sound from a single source reaching the closer ear and the sound reaching the further ear is registered by the brain and can be used to determine the azimuth—or egocentric angle relative to the head—of the sound source. Similarly, the difference in loudness (or intensity) of the sound between the near and far ears can provide information about the azimuth of the sound source. These two types of auditory cues—known as interaural time difference (ITD) and interaural loudness difference (ILD)—can also provide information about the elevation of a sound source by tilting the head.[94] Another type of auditory cue, known as the "head-related transfer function" (HRTF) describes how the spectral signature of sound frequencies are affected by external structures such as the head, torso, and especially the outer ear (or "pinna"). Sound waves are diffracted and reflected by these structures, and the angle at which sounds strike them cause interference patterns which shape the spectral frequencies of the sound in reliable ways. This allows the brain to extract information about both the elevation and azimuth of the sound source from its spectral signature.[95,96]

Of course, auditory space is not simply two-dimensional; in addition to azimuth and elevation, we also experience sounds as being located at varying distances from us. However, although ITD, ILD, and HRTF can provide very accurate information about azimuth and elevation of a sound source, they are much poorer at providing information

---

[94] Note that this itself is a kind of s-coordination: azimuth + head tilt.

[95] It's worth noting that interaural time difference and interaural loudness difference are *binaural* cues (meaning that they involve both ears), whereas the HRTF is a *monaural* cue (involving only one ear). However, the azimuth information obtained from monaural cues is much poorer than that obtained from binaural ones. Furthermore, spatial information can only be extracted from spectral information using the HRTF when the sound has a wide band of frequencies.

[96] Indeed, this is why many animals have large, specially shaped, moveable ears: it allow for improved auditory localization.

about distance. And indeed, psychophysical experiments show that human subjects are much poorer at judging the distance of an auditory stimulus than they are its azimuth and elevation. In practice, some distance information can be extracted from the amplitude of a sound, since amplitude decreases as distance increases. However, this sort of calculation requires some background knowledge about the nature of the sound source. For example, while this approach can be employed to judge the distance of familiar types of sounds (such as speech), we often lack the necessary information about the sound source required to make such judgments. Echoes can also be used to calculate distance; both (i) the ratio of direct sound source to echo and (ii) the spectral ‘coloration’ of the sound wave that results from the interference pattern between echo and source sound can provide some information about distance.[97]

In any case, what is important for the purpose at hand is that all these different types of auditory information must be s-coordinated in order to locate a stimulus in auditory space: information about ITDs and ITLs must be integrated to get an accurate azimuth reading, which in turn must be coordinated with the outputs of HRTF processing in order to get information about elevation.

The neural mechanisms underlying integration of these different kinds of auditory information are relatively well-understood: biaural cues such as ITDs and ILDs are processed in the superior olivary complex (SOC), but neurophysiological and psychophysical evidence suggests that their processing involves separate pathways (Palmer 1995, p.105). The initial processing of ILDs takes place in the lateral superior olive (LSO), where ILD-sensitive neurons are excited by stimulation of one ear and inhibited by stimulation of the other, such that their response strength depends on the relative sound intensities at each ear. In contrast, the initial processing of ITDs takes place in the medial superior olive (MSO). Cells in the MSO are insensitive not only to differences in sound intensity between ears, but moreover, also seem to be insensitive to delays in the onset time of a sound**.** Rather, cells in the MSO tend to respond selectively

[97] Animals that echolocate are of course extremely adept at using echoes to judge distance (as well as other features of stimuli). What’s more, some humans have also learned to use echolocation: see, e.g., Thaler et al. (2011)

to the relative *phase* of the sounds between the two ears (which of course varies as a function of temporal delay).[98] Finally, there is physiological and anatomical evidence of a topographically-organized spatial map in the MSO based on sensitivity to delay (Palmer 1995, p.111). That is, cells that are most responsive to very short (or near-zero) delays are located in more rostal (nearer the back of the head) areas of the MSO, whereas cells sensitive to longer delays are located progressively more caudal (nearer the front of the head). Again, as in the visual system, spatial adjacency relations are preserved in the topographic organization of these sensory areas.

Both the LSO and MSO project to the inferior colliculus (IC), one of the main processing centers of the auditory system in the brain and the locus of s-coordination between ILDs & ITDs. There is ample evidence that cells in the IC respond selectively to sounds coming from specific locations in three-dimensional space, with cells whose receptive fields are selectively 'tuned' to respond to a particular narrow combination of ILD and ITD (as determined by the output of the MSO and LSO processing). Furthermore, there are some indications the IC may contain something like a topographically-organized map, although it is probably "complex" and "nonlinear".[99] The IC also receives projections from the dorsal cochlear nucleus, which is thought to be involved in processing auditory information relating to elevation. The IC is thus perfectly positioned to s-coordinate auditory spatial information in the sort of way that Grush describes. As a result of IC processing, we possess a higher-order map of auditory space that employs an egocentrically-based bipolar spherical coordinate scheme centered on the head with the horizontal plane crossing the two ears (the "interaural axis").[100]

---

[98] "According to the traditional view of ITD processing (Jeffress, 1948) the underlying neural mechanism relies on delay lines in the transmitting of neural signals. However, Jeffress' theory cannot account for certain psychophysical data, and thus it is likely to be not *entirely* correct, although it is probably *partially* correct.

[99] Specifically, Oliver et al. (2003) argue that "it unlikely that any map of ITD in the MSO is transmitted by point-to-point topography to the IC. Thus, if there is a topographical organization of stimulus azimuth in the IC, as suggested by free-field recordings (Aitkin et al., 1985), it is not a reflection of a spatially mapped projection from the MSO."

[100] Ahveninen et al. (2014)

The question now is how this **s**-coordinated representation of auditory space can be **c**-coordinated with the (**s**-coordinated) representation of visual space that was described above. That is, I've now described how both visual information and auditory information can be s-coordinated by various brain areas into modality-specific feature maps. Now, I will look at a way in which both kinds of information can be c-coordinated into a higher-order multi-modal representation of space.

5.6 Visual-Auditory Integration in the Superior Colliculus

The multimodal integration of spatial information from vision and audition is accomplished (at least in part) by the superior colliculus (SC), which has been implicated in the spatial processing of both types of sensory information. In the SC, one finds topographically organized spatial maps in a number of different sensory modalities, including vision and audition. Some of these maps are modality-specific, such that their cells respond only to visual stimuli or auditory stimuli; however, the SC also contains maps with cells that respond to *multi-sensory* stimuli—i.e., they receive both visual and auditory input and respond with roughly equal strength to both auditory and visual stimuli in a particular region of space. The result is that when a stimulus that is multi-modal in nature (i.e., can be discriminated by both the visual and auditory systems) falls within the receptive fields of these multisensory cells, the combined excitatory stimulation of both aural and visual inputs causes an increase in the saliency of the stimulus.

What's more, all of these distinct topographic maps—visual, auditory, multisensory, and others—are *aligned* with one another in the SC; they are literally superimposed over one another in the brain. As Jay & Spark (1987) point out, this implies that these sensory signals have been translated into *a common coordinate system.* Indeed, as was mentioned above, in order to c-coordinate maps in different sensory modalities they must be translated into a common coordinate scheme. Prior to the SC, the location of a visual stimulus is encoded in terms of retinal coordinates + eye position while the location of an auditory stimulus is encoded in head-centered spherical bipolar coordinates of azimuth, elevation, and distance. If the multimodal maps in SC truly do represent spatial locations in a way that is independent of either vision or audition, then it must be in terms of some common, inter-translatable coordinate scheme.

There are, of course, different possible ways which these different coordinate schemes could be translated into a common one. For example, both could be translated into an egocentric Cartesian coordinate scheme centered on the head. (Indeed, given that both coordinate schemes are head-centered to some degree, this makes a certain amount of intuitive sense.) However, it seems that is not how the brain represents visuo-auditory space. Rather, multimodal cells in SC that respond to both auditory and visual information appear to encode spatial relations in terms of *motor* coordinates; specifically, in terms of what is known as *gaze shift*—the change in eye position required to orient towards a stimulus.

Thus, auditory space has been translated into motor coordinates which specify the change in eye position required to look at an auditory stimulus, encoded in terms of retinal displacement. (Jay & Spark, 1987, p.50) Furthermore, in order to maintain the alignment of visual and auditory maps following eye saccade, there is a dynamic re-mapping of receptive field locations for auditory stimuli that occurs in SC such that the spatial location of the receptive field for cells that selectively respond to auditory stimuli is significantly modulated by the position of the eyes in their sockets.

This integration of auditory and visual spatial information with information about the motor system is possible because the SC contains topographic maps of not only visual and auditory space, but also of *motor space* as well.[101] That is, just as one can specify the spatial location of a stimulus in terms of a coordinate scheme defined in terms of the stimulation of sensory receptors (e.g., in terms of retinal location, or interaural time difference, or bodily location of a tactile receptor), one can also specify the relative spatial position (and orientation) of body parts in terms of a coordinate scheme defined in terms of motor activity. As Grush puts it,

---

[101] In fact, the SC also contains maps of somatosensory space that are aligned with the visual, auditory, motor, and multi-sensory maps. However, I will refrain from including it in the discussion here for the purposes of simplicity.

> "…it is possible to specify the location of my hand relative to my torso by giving the angles of my shoulder, and elbow joints. Given that my shoulder has three (actually more than three, but let's keep it simple) and my elbow one degree of freedom, one can specify my hand position relative to my torso as a point in a four-dimensional joint-angle 'space'" (2000, p.66)

The s-coordination of sensory (e.g., visual and auditory) information with motor information is absolutely crucial to set the stage for a subsequent c-coordination of visual and auditory space. Auditory and visual spatial information are only able to be integrated into a multi-modal topographic map because the motor system actively translates auditory spatial location into gaze shift vectors in a coordinate scheme based on retinotopic displacement. As Grush points out, the involvement of motor maps is particularly important in the representation of spatial location because it provides a *common frame of reference* that allows for the integration of different types of sensory information, especially in cross-modal cases.

Consider, for example, what sort of s-coordinations are actually required in order to locate a stimulus in visual egocentric space. First (as was mentioned above) one must coordinate visual information about the location of retinal stimulation with proprioceptive information about not only the position of the eyes in their sockets, but also with information about the orientation of the head relative to the torso. This type of proprioceptive information is registered by, e.g., muscle spindles[102] in both the extraocular muscles (which control eye movement) and in the neck muscles. A representation of auditory egocentric space can be formed in an analogous manner; that is, via the coordination of information about the azimuth, elevation, and distance of auditory stimuli with proprioceptive information about the position of the head relative to the torso and vestibular information about the orientation of the body and/or head. (Indeed, precisely this sort of integration of proprioceptive information with audition is

[102] These cells essentially act as stretch receptors, such that their firing rate is a function of the amount of change in muscle length as well as the speed of that change. In this way, they are able to encode information about the position of various body parts relative to one another.

necessary in creatures who move their heads and ears to obtain better spatial resolution!)[103]

It is important to note that the coordination described above is overly simplistic and incomplete. For example, I have said nothing about the vital need for sensory systems to be sensitive to inputs that contain information about *efference copies* of motor commands. Furthermore, even this additional motor information is not yet sufficient for localizing a stimulus in visual or aural egocentric space; rather, one must also coordinate such information with *vestibular* cues about orientation and balance, which in turn is provided by neural mechanisms in the inner ear. (That is, the ordered array consisting of [retinal location + eye position + head position] will correspond to different locations in egocentric space depending on whether one is lying down or standing upright.)

Nevertheless, the general point should be clear: the representation of spatial location crucially depends on the coordination of many different kinds of sensory and motor information. The content of such representations is a product of the joint contribution of all the lower-level, modality-specific sensory and motor maps that play a role in its formation. For example, the spatial content of a sensory representation of an auditory stimulus is determined in part by the change in eye and head position that would be required to look towards it. Similarly, as Grush (2000) puts it,

> "part of the content of, say, a visual stimulus is provided in part by how one would orient towards that stimulus motor action), and how one would move one's arm in order to bring the hand to that point, such as THE THING GRASPABLE BY REACHING THUS. Similarly, part of the content of a felt location is given by how one would visually orient that location, and how the hand would look when the eyes are trained on it." (p.70)

There is ample empirical evidence for this claim, based on well-documented cross-modal influences on spatial perception. For example, in the well-known "ventriloquist effect",

[103] The use of motor feedback is particularly important in *olfaction*, which is not an inherently spatial sensory modality.

subjects misrepresent the spatial location of an auditory stimulus because of the causal influence of a concurrent visual stimulus. Similarly, in the "rubber-hand illusion" (discussed in the next section), subjects come to misrepresent the spatial location of their limbs due to the contribution of certain visual cues.

5.7 The Representation of Spatial Location

Put differently, modality-specific feature maps and motor maps contain different *information*. But carrying information isn't sufficient to possess genuine *intentionality* or *aboutness or content*.[104] In order to have that, information from different maps must be coordinated in the ways described above. What's more, the (Fregean) content of a given spatial representation depends on—and indeed, is *determined by*—the way that different maps are coordinated.

Specifically, the coordination of maps containing different spatial information provides something that was discussed above in chapter 3: *conditions on extension*. That is, the kind of multimodal sensory-motor coordination described above provides conditions that an external spatial location must meet if it is to be considered in the extension of the representation. For example, as described above, the spatial content of a representation of a visual stimulus can be partially linguistically described as something like "*the location in visual space graspable by reaching thusly*", and the spatial content of a representation of an auditory stimulus can likewise be partially described as something like "*the location in auditory space that can be seen by moving your eyes and head thusly*".[105] Of course, these rough linguistic translations are necessarily incomplete, since the conditions on extension imposed by a representation of spatial location will necessarily make reference to all the different types of lower-level maps that go into its construction.

---

[104] Nor does the existence of topographically-organized maps mean that the information/content carried by those maps is *spatial*—for instance, the auditory cortex use topographically organized maps of pitch space, but the content/information carried by such maps is not spatial.

[105] Importantly, terms like "the location in [visual/auditory] space" or "moving thusly" act as placeholders here for specifications of locations in terms of (i) modality-specific, receptorally-defined maps described above, or (ii) sequences of motor commands.

Uncoordinated feature maps alone cannot impose conditions on extension, because they do not allow an organism to make use of the spatial information contained in those maps. Consider an example from Grush (2000) that concerns a type of sensory prosthetic device known as a *sonic guide*. The device works on a principle somewhat similar to echolocation: it transmits an (inaudible) sound with a small speaker, and records the subsequent echoes produced by the sound with a stereophonic microphone. The echoes are then analyzed and translated into different types of audible tones which are presented to the wearer of the device via headphones. Importantly, the tones presented to the listener depend on the nature of the incoming echoes in well-defined ways: for example, the volume and pitch of the tone vary as a function of distance and size (respectively) of the object creating the echoes. Likewise, the azimuth of the echo source is signaled by the difference in onset time of the tones at each ear.

The sonic guide thus provides subjects with relatively detailed spatial information. For example, a tone of 35dB at middle C will *carry information* to the effect that some object is at a such-and-such location in egocentric space. What's more, as Grush points out, groups of cells in the auditory cortex of the wearer that are selectively tuned to fire in response only to tones of 35dB at middle C *will carry exactly the same information*—i.e., that there is some object is at a given location in egocentric space. Of course, the mere fact that such cells carry such information is not sufficient for a subject (or their sensory states) to *represent* that location. Put differently, simply because the cells carry this information does not mean that it is available to the subject. For that information to be *usable*, it must be correctly coordinated with other kinds of spatial information from different sensory modalities and motor systems. The subject must come to (sensorily, non-inferentially) understand that *"35dB at middle C is graspable by reaching thusly"* or *"35dB at middle C can be seen by moving your eyes and head thusly"*. It is only through this sort of coordination that sensory states come to be imbued with genuine spatial content.

On this view, misrepresentation occurs when maps are incorrectly coordinated, due to interference from either external factors in the environment or internal errors in processing. For example, consider the "rubber hand illusion": by brushing a rubber hand that is placed in front of a subject while simultaneously brushing the subject's own hand that is hidden from their view, a sensation of body ownership over the fake hand is produced. The subject comes to identify the fake hand as their own due (in part) to a mis-coordination between visual maps and proprioceptive/motor maps. Likewise, auditory front-back reversals (described above in section 5.2) will cause subjects to fail to orient their eyes to the *actual* spatial location of an auditory stimulus. (Of course, in this latter case the error is not due to an error in the process of coordination, but rather, a kind of "normal malfunction" in a lower-level map.) A similar effect can be achieved by simply fitting subjects with lens that shift the apparent visual location of objects. Initially, subjects have great difficulty with tasks such as navigating their environments and interacting with objects. This is because their maps of visual space are incorrectly coordinated with spatial maps in other sensory modalities, including motor maps. However, after sufficient time wearing the lenses, subjects are able to compensate for the change and their performance on such tasks returns to normal. This suggests that our ability to coordinate spatial maps is flexible, and not fixed.[106]

In any case, it is through the process of multimodal sensory-motor coordination described above that sensory systems can come to have a kind of basic *aboutness* or *referential directedness*; in other words, *intentionality*. Specifically, the coordination of maps allows subjects to have a kind of sensorimotor directedness towards spatial locations that is grounded in the possibility of different types of motor action that allow for a kind of cross-modal engagement with that location, such as moving one's eyes towards a heard stimulus or reaching for a seen object.[107] Importantly however, although they may be *activated* by some particular uni-modal stimulus, spatial representations of this type are

[106] Importantly, spatial representation does not involve localizing stimuli on static maps of space, but rather is a type of on-line processing of sensorimotor spatial localization. Our sensory systems are constantly receiving ever-changing information, both from the external environment, from our internal sensors, from efferent copies of motor commands, from remapping processes, and so on. In this chapter, I have largely talked about spatial representation and coordination of spatial information in *static* terms, but only for the purposes of ease of exposition.

[107] Importantly, one need not actually engage in any of this overt behavior: rather, it is enough that such behavior is *possible* or such engagements are *available* to the subject.

autonomous of any particular sensory modality, and characterize locations independently of any particular features that happen to be instantiated at those locations. (I discuss this issue further in the next chapter).

Chapter 6

6 The Binding Problem

In this chapter I describe the *binding problem*: how phenomenal qualities are represented as features of "sensory individuals" such as locations and objects. I distinguish two parts to the binding problem—the problem of *sensory individuals* and the problem of *sensory integration*—and review some of the major psychological theories on binding that attempt to solve each of them. Finally, I examine the role of binding from the perspective of sensory representation and psychosemantics. I argue that the proposed solutions to the binding problem (FIT and FINST) serve a *referential* role in sensory psychosemantics by functioning something like natural-language demonstratives, and that binding—understood as the capacity to characterize a location or object as possessing a certain phenomenal quality—is the most fundamental kind of sensory representation that is relevant to representationalism about sensory phenomenology.

6.1 The Binding Problem

As noted above, our sensory experiences contain more than simply isolated qualities. Rather, they typically present such qualities (i) as being located at certain spatial locations, and (ii) as features of objects. That is, we don't simply have sensory experiences of, e.g., disembodied redness; rather, our sensory experiences represent objects and locations in our environment as instantiating that property: we experience the redness as being instantiated by *that object*, and we experience that red object as being *over there*, in the three-dimensional space surrounding us.

To put it in slightly different terms (adopted from Clark, 2000), a theory of sensory experience that focuses only on the extraction of information about sensory *features* is necessarily incomplete. Sensory experience also necessarily involves the representation of sensory *individuals* that are the bearers of those features.

Above, I discussed how spatial locations—a type of sensory individual—are represented by sensory systems. However, I have not yet discussed how representations of spatial locations are combined with representations of the features that appear at those locations. Without an account of how this is achieved, we lack an explanation of how we make the perceptual discriminations that we do.

One way to understand the problem can be described as follows: even if one were to have a true theory of the perceptual discrimination of sensory features, such that it could explain how one could discriminate, e.g., red from green, it would still not be able to explain how one can distinguish a red triangle from a green square that are simultaneously present in one's visual field. Nor could it explain how one is able to discriminate a pain in one's foot from a pain in one's hand, or how one is able to discriminate a sound coming from one's left from a different sound coming from one's right. In order to explain these sorts of perceptual discriminations, we require a theory of how sensory systems manage to distinguish between (and represent) distinct sensory individuals, and of how sensory systems link representations of sensory features to representations of sensory individuals in the correct combinations.

In psychology and neuroscience, this is known as the "binding problem", and it can be opertationalized by considering examples of how binding can fail; i.e., under certain experimental conditions, subjects can be made to perceive illusory conjunctions of features and individuals. Here, I will argue that the "binding problem" can be separated into two distinct (but related) problems, which I will refer to as *the problem of sensory individuals*, and *the problem of sensory integration*. The problem of sensory individuals concerns how our sensory systems pick out, track, and represent discrete sensory individuals, potentially across changes in all their perceptible features. It is fundamentally a problem of *reference*, and concerns the nature of the 'representata' (intentional objects) that we experience as instantiating sensory features. On the other hand, the problem of sensory integration concerns how different types of sensory information can be integrated both within and across modalities such that it can be used by other brain systems (e.g., categorization/identification; cognitive processes; to guide action, etc.). In short, the problem of sensory integration is how to establish relations between different types of sensory representations: representations of objects, representations of features, and representations of spatial locations.

However, these two formulations of the binding problem—sensory individuals and sensory integration—are not the only problems that go under the heading "the binding problem". Indeed, Treisman (1996) identifies at least *seven* different versions of the binding problem, and Roskies (1999) lists six "perceptual" and three "cognitive" types of binding, all of which may require at least somewhat different solutions. Moreover, even within any particular version of the problem, binding itself may be further subdivided into a number of sub-processes that are underwritten by separate mechanisms and thus have different explanations. Thus, a variety of different solutions to the binding problem have been proposed, not all of which target the same phenomenon. What's more, even solutions which *do* target the same problem may not even be proposed at the same level of explanation, thus further compounding the difficulty.

However, the panoply of issues that collectively constitute the binding problem are nevertheless deeply interconnected, and—I argue—all share a vitally important characteristic: they are all problems that, at essence, concern the most basic type(s) of sensory representation. Indeed, in this chapter I will ultimately argue that the binding problem is fundamentally a problem involving the psychosemantics of sensory representation at its most fundamental levels. More specifically, I will argue that binding both (i) provides a basis for the *referential directedness* of sensory states by positing mechanisms which represent sensory individuals in a direct, ostensive-like way that can be said to *ground the intentionality* of sensory states, and (ii) allows for sensory experiences to be *genuinely representational*, insofar as binding provides sensory experiences with veridicality conditions by establishing relations among disparate sensory contents (sensory integration).

6.2 Sensory Individuals vs. Sensory Integration (Selection vs. Encoding)

As described above, the "binding problem" can be separated into *two* problems: the problem of sensory individuals and the problem of sensory integration. This distinction between these two types of binding problems is sometimes described as the difference between "parsing" and "encoding" (Treisman, 1999). The former concerns the process or mechanism by which combinations of features are selected to be bound together, whereas the latter concerns the encoding of those combinations of features into a unified representation.[108]

Prima facie however, it might seem that the "parsing" problem is quite different from the problem of sensory individuals. As described by Treisman, the parsing problem concerns how features that are to be bound are "selected and segregated" from features belonging to other objects or location. In contrast, the problem of sensory individuals was described above as the problem of how sensory systems manage to pick out, track, and represent the bearers of sensory features. However, the problems are in fact equivalent. For, as I shall demonstrate below, features are "selected" to be bound together *in virtue of* belonging to the same sensory individual. In other words, the picking out, tracking, and representing of sensory individuals is *prior* to the selection of sensory features that are to be bound.[109] Thus, I prefer to describe this problem as that of "sensory individuals" because the invocation of terms like "parsing" and "selection" could be somewhat misleading if they are taken to imply that the selection of the features to be bound occurs prior to or independently of the representation of the individuals that are the bearers of those features.[110]

Importantly, different psychological theories of binding can propose different solutions to the problem of sensory individuals while nevertheless agreeing about the mechanisms responsible for encoding/integration, and vice-versa. For example, there are two main types of psychological theories that attempt to explain feature binding (both of which focus almost exclusively on the binding of *visual* stimuli). According to Anne Treisman's

[108] There is a third version of the binding problem in addition to selection and encoding, to which Treisman (1999) refers to as a problem of "structural description" or "the within-object binding problem". It concerns "the correct relations specified between the bound elements within a single object", or "determining how parts within an object should be bound and how illusory conjunctions within objects can be avoided". (p.109). For simplicity, I will leave this issue to the side, as it does not affect my main claims.

[109] For example, certain features (and not others) are "selected" to be bound to, e.g., a location on a master map (described below) because *only the feature map locations that project to that particular master map location are available to be so selected*.

[110] In other words, the *wrong* way to understand the problem is to wonder how the brain manages to determine that feature A belongs to sensory individual B, or how it determines that feature X and feature Y are co-instantiated by the same sensory individual Z. For, the way that the brain actually accomplishes this "selection" is by grouping features *via* sensory individual, such that the latter determines the former. Thus, different types of solutions to the parsing problem will determine what sorts of sensory individuals are posited by the subsequent theory of binding.

(1980, 1982, 1996, 1999, 2003) view, the binding of features is *location-based*, such that features that occupy the same spatial position are 'bound' together. On the other hand, Zenon Pylyshyn (1999, 2001, 2003) argues for an *object-based* solution to the binding problem. Specifically, he argues for the existence of "visual indexes" or "FINSTs" (for 'FINgers of INSTantiation') which track individual objects (or "*proto*-objects") in the visual field. On his view, features that belong to the same (proto-) object are bound together.

These two views differ with respect to the parsing problem and thus differ in the types of sensory individuals that are posited to facilitate binding (i.e., locations vs. objects). However, both these views nevertheless hold that encoding (i.e., sensory integration) should be understood as the process of "tagging" features to certain kinds of representational structures known as *object files* (described below). They differ insofar as Treisman holds that object files are attached to locations on a 'master map', whereas Pylyshyn holds that object files are attached to visual indexes.[111] (And even here, there is overlap: Treisman et al. (1992) explicitly acknowledge that FINSTs can attach to object files as well as locations on a master map.)

In what follows, I will briefly describe both Treisman's and Pylyshyn's respective theories of the "parsing" mechanisms that underlie the capacity to pick out, track, and represent sensory individuals, as well their common solution to the encoding/sensory integration problem.

6.3 Treisman's Feature Integration Theory (FIT)

According to Treisman's (1982, 1992, 1996, 1999, 2003) view—known as "feature integration theory", or "FIT"—the mechanisms that underlie binding are ultimately *location-based*, such that features which share a common spatial location get bound

[111] Importantly, use of the expression "attached to" here should not be taken to suggest that object files exist independently of and prior to visual indexing and/or attending to locations on a master map. Rather, as I will described in further detail below, object files are temporary representations whose instantiation is generated by these processes.

together.[112] Thus, on FIT, spatial locations are posited to be (at least one of) the most fundamental type of sensory individual to which features are bound.

Treisman's FIT is based on *feature maps*, which were described above in sections 3.6 and 5.3. According to FIT, different sensory systems (and sub-systems) contain distinct feature maps which respond to specific features of stimuli such as colour, orientation, shape, pitch, temperature, and so on. Importantly, all these different feature maps are connected to a *master map* of locations (also called a "location map"), with equivalent locations on individual feature maps projecting to the corresponding location on the master map. According to FIT, it is this particular structure of connectivity that allows for the possibility that different sensory features which occupy the same spatial location can be 'bound' together.

Importantly however, as Viger et al. (2008) puts it, on the FIT view, "Binding is more than just a collection of features projecting to the same location on the master map of locations. It requires *attentional selection* of that location on the master map in order to identify a search element as having all of those features." (p.269, *emphasis added*) In other words, for feature binding to actually occur, the master map must be scanned by attentional processes. When the "spotlight" or "window" of attention is focused on a particular location on the master map, all of the features at the corresponding locations in individual feature maps are available for encoding, and features at other locations are inhibited or suppressed, in order to prevent erroneous binding.[113]

Thus, on this view, feature integration proceeds in two stages: in the first stage, which is fully 'bottom-up', different individual feature maps process information about stimuli in

[112] The expression "common spatial location" here should be taken to typically indicate a *region*, not a point. According to Triesman, the size of this region is *scalable*, and depends on the particular perceptual conditions and perceptual task. Moreover, it is an empirical question what the constraints on the size of the region must be, and how those constraints are determined. However, it is very unlikely that, for example, the entire visual field (or even a large portion thereof) could be considered as a 'single' region on FIT.

[113] The suppression or inhibition of locations which are not the locus of attention can also be facilitated by the individual feature maps themselves. As Viger et al. (2008) puts it, "Activity on the individual feature maps can aid search by inhibiting locations on the master map where non-target features occur. For example, if the target is a vertical green bar, then the orientation feature map can inhibit locations where horizontal bars occur. Inhibited locations are passed over in the serial attentional scan, reducing the number of locations that need to be searched." (p.268)

their respective brain areas. In the second, ‘top-down’ stage, the master map is scanned by attentional processes. Features at locations on different maps that project to the location on the master map that is currently illuminated by the ‘spotlight’ of attention are thereby made available for binding; that is, selected to be encoded into an active representation of something with those features at that location. On this theory, errors in binding occur only at locations outside the ‘spotlight’. Because attention is not focused on those regions, features can be mistakenly combined in illusory conjunctions.

6.4 Pylyshyn’s ‘Visual Index’ Theory

Pylyshyn’s proposed solution to the binding problem posits the existence of “visual indexes” or “FINSTs” which track individual objects (or “*proto*-objects”) in the visual field. According to this view, features are selected to be bound together on the basis of belonging to the same (proto-) object. Pylyshyn thus argues that these proto-objects are the most fundamental type of sensory individual (in vision) to which features are bound.

There are good reasons for thinking that the visual system requires precisely the kind of object-based mechanism that Pylyshyn posits; after all, the visual system obviously needs to (and successfully does) keep track of individual objects regardless of changes in all their visual properties—shape, size (distance), colour, orientation, etc. Furthermore, it has been experimentally demonstrated that the visual system can track up to 4-5 individual objects at a time (Pylyshyn 2003), thus lending empirical support to the claim that exactly the sort of object-tracking mechanisms that Pylyshyn describes do in fact exist.

According to Pylyshyn, the indexing of features to objects proceeds as follows: first, processing in the early parts of the visual system identify and segregate clusters of features, “which tend to be reliable proximal counterparts of distinct individual objects in a distal scene” (2003, p.146). These feature clusters then compete for access to available visual indexes, of which there are four or five. (The outcome of this competition between feature-clusters for the limited pool of indexes is primarily determined by the saliency of a stimulus, although it can also be mediated to some degree by top-down cognitive influences.) Once a ‘winning’ feature cluster is assigned to a particular visual index, the external object instantiating those features can change all of its visual properties (including spatial location) “within certain as-yet-unknown constraints” (ibid), thus allowing the visual system to track individual objects despite changes in their visual appearance.

6.4.1.1 FIT vs. FINST

Pylyshyn argues that his visual index theory offers a better solution to the selection problem than Treisman’s FIT.[114] For example, he claims that FIT runs into difficulties when it attempts to account for the possibility of objects that occupy the same spatial location, such as a green square surrounded by a red circle. In such cases, we are able to distinguish between the two—by successfully binding redness to the circle and greenness to the square—despite the fact that both objects would seem to occupy the same spatial location. Pylyshyn claims that this suggests that the mechanism(s) responsible for the selection and extraction of features must therefore treat the two stimuli as distinct objects, and not as a single spatial location with a variety of different features.

Of course, there are some obvious ways for the defender of FIT to respond to Pylyshyn’s objections to the theory: for one, it seems clear that there is sense in which a green square surrounded by a red circle can be distinguished from one another on the basis of spatial location of perceptible features alone. Moreover, features can be bound to locations in complicated hierarchies that allow overlapping locations to be treated as partially co-extensive. Furthermore, there are good reasons to think that map-based feature binding is ubiquitous in sensory systems: whereas FINSTs are unique to vision, feature maps are found in many different sensory modalities. For example, audition can attribute features—such as pitch, timbre, etc.—to spatial locations without attributing them to any particular object thanks to internal ‘maps’ of auditory space (in, e.g., the IC and the SC, as described above in chapter 5). Similarly, the somatosensory system contains a detailed (albeit distorted) spatial map of the body’s surface in the so-called “cortical homunculus”. What’s more, although Pylyshyn seems to be correct in claiming that we have modality-

[114] This view is also defended by Matthen (2004, 2005)

specific representations of objects in vision, many other sensory modalities do not seem to (unimodally) represent objects.[115]

Indeed, it's much more likely that Treisman's FIT and Pylyshyn's FINSTs are not in fact mutually exclusive. Rather, they both describe different mechanisms that pick out different sorts of sensory individuals. Indeed, Treisman herself has acknowledged that the visual system contains FINST-like mechanisms that underlie feature binding in addition to her map-based location mechanisms. (Clark, 2000, 2004 has also argued that the two views are not incompatible.)

In any case, it is not my intention here to comment on the merits of Treisman's and Pylyshyn's respective theories (i.e., between FIT and FINST), because the differences between them are in fact not directly relevant to my purposes. Rather, I'm interested in what both theories have in common: *first*, they both rely on a type of representational structure known as "object files" to solve the sensory integration (/encoding) problem. *Second*, both views share a particular view of the *referential function* of their respective mechanisms. I will address these two issues in turn.

6.5 Object Files & the Problem of Sensory Integration

As noted above, there are (at least) two parts to the binding problem: the problem of sensory individuals ("parsing") and the problem of sensory integration ("encoding"). As described thus far, Treisman's FIT and Pylyshyn's FINST both provide different—but equally and independently plausible, and non- mutually exclusive—accounts of the mechanisms that are potentially responsible for solving the former problem; however, we still require an account of the mechanisms responsible for solving the latter.

As noted above, both FIT and FINST share a common view about the psychological mechanisms responsible for encoding the binding of features: *object files*. (Kahneman, Treisman, and Gibbs 1992; Feigenson and Carey 2003; Pylyshyn 1999, 2003) Object files are often described as "temporary episodic representations" (KTG, 1992), or as

[115] For example, we *see* objects, but we don't *hear* them – we hear *sounds*.

"mid-level representations" (FC, 2003) that occur between earlier, modality-specific sensory representations and later, conceptual/cognitive representations.

According to the theory, the primary output of early sensory processing is an object file or set of object files, with each individual file corresponding to a particular sensory individual—i.e., (proto-) object or spatial location—in the perceived scene.[116] These object files contain all the available sensory information that is associated with that object or spatial location. That is, one might think of the object files as containing a 'list' of perceptible features, which is linked to a particular sensory individual via a visual index or location on a master map. Object files can thus potentially solve the problem of sensory integration on the hypothesis that any and all features that are 'tagged' to the same object file are bound together and to the sensory individual linked to that object file.

Furthermore, only the contents associated with a given object file are available to subsequent cognitive processes. For example, the sensory information that is contained in an object file can also be used to facilitate object recognition/identification/categorization by comparing it to information about the perceptual features of known objects stored in memory—essentially a kind of pattern matching.[117] If a match is found, semantic/conceptual information about the object can also be tagged to the file.[118]

On this view, the identity of an object across time, spatial location, and/or changes in perceptible properties is determined by whether or not sensory information about its successive states is assigned to the same object file, rather than by any of its perceptible features. When the sensory input changes, the information in an object file is either (i) updated, if it is determined to be the 'same' object; or (ii) discarded, and a new one is opened (if it is determined to be a 'new' object). For example, consider the so-called

---

[116] The term "object" in "object file" can be understood to apply to spatial locations as well: indeed, as Viger et al. (2008) points out, on the FIT proposal "binding is a result of treating locations as the locations of search elements, i.e. objects." (p. 269)

[117] Furthermore, the reverse is likely true: conceptual information can enhance sensory information, such as in the case of *amodal perception* (our experiential awareness of the features of an object that is partially occluded).

[118] It's important to be clear that the relevant semantic/conceptual information need not be actually be stored in the object file itself. Rather, the processes of 'tagging' can be understood as a way of linking the file to the relevant information so that it is available if/when it is needed.

"colour phi phenomenon" (Kolers & von Grünau, 1976), a perceptual illusion in which two different coloured dots are briefly flashed on a screen in quick succession, with a small spatial separation between them. When the temporal delay between the onset of the first dot (red) and the onset of the second dot (green) is above some certain critical threshold and the spatial separation of the dots is above some other critical threshold, one perceives them as two distinct objects/stimuli that appear in succession. However, *below* those critical thresholds, one stops perceiving the dots as two distinct objects and instead has the experience of a single dot moving and changing colours from red to green. This phenomenon can be explained by positing that, below the critical thresholds, both dots are assigned the same object file, whereas above that threshold they get assigned to different object files.

6.6 Object Files, Content, and Sensory Reference

The colour phi phenomenon provides an example of many different types of binding, including (i) temporal binding, (ii) motion binding, (iii) location binding, and (iv) feature binding, as well as others.[119] However, the colour phi phenomenon also illustrates another very important aspect of "encoding" mechanisms like object files: namely, their relation to "selection" mechanisms such as FINSTs. Perceiving the colour phi illusion depends on the fact that below the threshold, both dots are "indexed" by the same FINST; however, the *content* of the visual representation (i.e., that there is a single dot that moves and undergoes a change in colour) is due to the particular features that are listed in the object file to which that FINST is linked at any given time.[120]

As was argued above, both the selection and encoding versions of the problem fundamentally concern the nature of sensory representation. The encoding problem is a problem about establishing relations between different types of sensory contents—namely, different representations of features, objects, and spatial locations. On the other hand, however, the selection problem is a problem about *reference*—it concerns how

[119] See Treisman (1996) and Roskies (1999) for a detailed list of different types of binding.
[120] Matthen (2005, p.278-282) also discusses the colour phi phenomenon in the context of the binding problem to argue for the existence of FINST-like object tracking mechanisms to which qualities ar bound.

sensory systems pick out and track discreet sensory individuals.[121] Put somewhat differently, the encoding problem concerns *how* something is represented, whereas the selection problem concerns *what* is represented.

What is particularly important about selection mechanisms such as FINSTs and master maps of locations is that they function something like *referential pointers***,** allowing sensory systems to track individual objects and locations without representing any of their features.[122] For example, Treisman's master map (which is scanned by attention) is not a map of *features*; rather it is a map of *locations*, represented independently of whatever features happen to be instantiated at those locations at any given time. Similarly, visual indexes are posited precisely because (by definition) they can track objects despite changes in their visual properties (within certain constraints).

This allows sensory systems to pick out particular individuals regardless of what sensory features the individual might have at a given time. In this sense, these referential pointers function something like natural-language *demonstratives*. Indeed, Pylyshyn (2001) has explicitly claimed that visual indexes should be viewed "as performing a demonstrative or preconceptual reference function" (p. 127). Similarly, Clark (2000, 2004) has argued that Tresiman's master map of locations can also act as a kind of referential mechanism, which he calls "feature-placing". However, Clark also argues that the psychosemantic function of feature-placing mechanisms is somewhat different than FINSTs, and thus "fail in several ways to have the referential powers of visual indices." As he explains it,

> "…a FINST refers in something like the way a demonstrative term in natural language refers; a feature-map refers (or better, "indicates regions") in something like the way a map refers […] the 'placing' in feature-placing provides materiel

---

[121] The selection problem concerns reference because the ability to refer to a particular sensory individual directly (i.e., regardless of any property it has at the time) is what allows for features to be selected for binding (and what determines the features that are selected).

[122] Indeed, Viger et al. (2008, p.269) describes a visual index as "a referential pointer to the object that is without content, except for perhaps the "thatness" defined by its functional role. The content of a visual representation is recorded in the object file." Or as Kahneman & Treisman (1992, p.216) puts it, "FINSTs are perhaps closer to Marr's concept of place tokens-abstract markers that allow the visual system to treat filled locations independently of the particular features or objects that occupy them. For example, place tokens allow certain spatial relations, such as co-linearity, to be made explicit without reference to any other aspect of the elements between which they hold."

> with which to make the step to full-blown reference, but it alone cannot complete that step. It is proto-reference; or, if you like, protean reference". (p.467)

In the previous chapter, I describe the specialized referential capacities of spatial representation in much greater detail. In any case however, my primary point here is unaffected: "selection" mechanisms such as FINSTs and location maps act as referential mechanisms. What's more, the demonstrative-like nature of these types of referential mechanisms is vitally important from the point of view of the psychosemantics of sensory representation. For example, it is instructive to compare the account of sensory representation described above to a *descriptivist* version of sensory representation that takes reference to be fixed by some kind of descriptive content associated with a representation. Such an account is ultimately untenable for at least two reasons:

First, as was mentioned above, the need for a kind of 'direct' reference is particularly necessary in sensory representation because sensory systems must be able to refer to, track, and re-identify objects that could be continuously changing all of their sensible features (colour, shape, pitch, temperature, smell, etc.). A description that picked out sensory individuals solely by reference to those features would be unable to perform this kind of tracking, as it provides no way to determine that two different descriptions are in fact co-extensional (since any variation in a the sensible features of a sensory individual would necessarily change the description required to pick it out).[123]

Second, and more importantly, descriptivism about sensory reference is untenable because it fails to explain how things *inside the head* manage to establish relations to things *outside the head*. That is, it fails to explain how relations holding between internal states can establish relations to external objects and properties. This is known as the *symbol-grounding problem* (Harnad; 1990). The problem (roughly) is that representations—symbols—cannot acquire their semantic properties only from their

[123] Of course, it's true that sensory individuals have to be picked out and tracked by some set of features; however, Pylyshyn is clear that these need not be the features that can be listed in the object file associated with that sensory individual. For example, the *onset* of a visual stimulus is a feature by which an object might be indexed but which need not be listed in the object file. Similarly, mechanisms like *edge detectors* are useful for segmenting proto-objects from the rest of a visual scene but are too low-level to be available to be a feature capable of being tagged to an object file.

connection to other symbols in the representational system, for that would lead to an infinite regress. At some point, the meanings of our representations have to be grounded via connections with the external world. (This problem was first discussed in section 3.4 above, in the context of short-armed FRS.) As Devitt (1990) puts it (perhaps unfairly), purely descriptivist theories of reference seem to have no answer to this problem other than “it’s magic”.

However, both of these issues are easily dealt with on the account of sensory representation described above. The demonstrative-like referential pointers posited by FIT and FINST allow one to refer ‘directly’ to objects and locations in the external world (thereby grounding the intentionality of sensory states) without needing to represent those objects and/or locations under any kind of description.

Importantly however, as I will argue in the next section, the capacity to characterize a sensory individual under a description is the most fundamental type of sensory representation in phenomenal experience. Genuine, veridicality-evaluable representation requires representing an object, property, or state of affairs under some description or other; i.e., as belonging to a certain category or falling under a certain concept, or possessing a certain feature. However, as noted above, in order to be able to do this we ultimately require some way to refer to the things which our representations are about—their intentional objects—that does not rely on a description. Hence the need for the referential pointers described above in FIT and FINST. In other words, thanks to the demonstrative-like mechanisms of sensory reference and the object file –based encoding mechanisms of binding, *qualitative experience is about the world;* phenomenal qualities are not experienced as properties of sensations, but as properties instantiated by external things.

6.7 Binding & Sensory Psychosemantics

As discussed above, the binding problem is, at essence, a problem that fundamentally concerns *sensory representation.* It is the problem of how the brain manages to represent that, e.g., a certain sensory individual (location/object/etc.) has a particular feature or set of features, or that, e.g., a certain object is at a certain location, or any other of the seven-plus variations on the binding problem.[124]

However, I argue that 'bound' sensory representations are of a particularly special kind. Specifically, I argue that binding is the most fundamental type of sensory representation (that is relevant to representationalism about sensory phenomenology) because binding provides sensory systems with the 'hallmark of representation': *conditions of accuracy*.

To employ a linguistic analogy, a 'bound' representation containing both features and individuals functions something like a statement or a proposition, whereas the representation of individual features, locations, and objects are more like words. While the latter are the atomic constituents of linguistic representations, and while they have referents/extensions, they can't be said to provide a set of *truth-evaluable propositions* in the way that sentences or propositions do. Similarly, sensory representations of 'unbound' features, locations, and objects have extensions or referents, yet they lack genuine veridicality conditions. In contrast, 'bound' representations have conditions of accuracy: the binding of a sensory feature to a particular sensory individual—such as an object or location—can be evaluated as either veridical or non-veridical, depending upon whether the object/location in question actually has that feature or not. (For example, illusory conjunctions are one type of misrepresentation.)

Psychosemantically, a 'bound' sensory representations contains two different types of terms: (i) *singular* terms that pick out a particular sensory individual, in the form of an object or spatial location, and (ii) *general* terms that are used to characterize the object and/or location that is picked out. The general terms—sensory features—are the qualitative aspects of sensation, and they act as *predicates*. The singular terms are the mechanisms of sensory reference, and act like *demonstratives.* The process of binding features to objects and locations thus produces something like a "sensory sentence" with

[124] Indeed, Jan Plate (2007) argues that there is "no way of making sense of the binding problem" unless one accepts that it is fundamentally a problem about representation. For one, the output of binding processes "is always characterized, to put it very generally, by some kind of representation." (p.776) It is representing something as having a certain property (which, for Plate, is to be distinguished from representing something that has that property).

a particular type of predicate-demonstrative structure. Empirically speaking, the general, predicate-like terms can be characterized in terms of quality space theory and explained by neural mechanisms such as the opponent processing channels that underlie colour vision (explained in detail in chapter 4). On the other hand, the singular, demonstrative-like terms can be characterized in terms of selection mechanisms such as FINSTs and location maps. Finally, this "sensory sentence" gets stored in an object file that functions to link representations of features to representations of objects and locations.[125]

It is important to recognize that what enables a sensory sentence to *say something*—to have genuine, veridicality-evaluable representational content—is the fact that its constituent parts are capable of cooperating together to mutually discharge their respective psychosemantic functions.

What's more, these constituent parts have the respective psychosemantic functions that they do because (in part) they are available to be bound together in a sensory sentence. In other words: *sensory states can only function as contentful representations of properties, spatial locations, and objects because those states are available to be bound together*. Indeed, recall (from section 3.4) that both the QST-based MoP and the inferential-role based CoE that are jointly constitutive of sensory representations of properties are analyzed entirely in terms of short-armed FRS. However, as described above, relations among internal states alone is not sufficient for genuine intentionality. Similarly, although the sensory states that function as referential pointers to demonstratively pick out locations and objects are intentionally directed, they lack any representational content aside from the "thatness" of their functional role. It is only by virtue of the fact that these two different types of sensory states can be bound together that allows them to be representational—their availability as constituent parts of a larger whole endows those parts with representational or semantic properties that they would lack if they were not in fact available in this way.

[125] Technically, the "sensory sentence" is not stored in an object file, but rather exists as a distributed representation. The object file is the mechanism by which different types of lower-order representations are bound together, thereby constituting the sensory sentence.

6.8 Binding and Phenomenal Contents

As stated above, binding is the most fundamental type of sensory representation that is relevant to representationalism about sensory phenomenology. The reason for this is because *only bound representations can enter into the phenomenal content of sensory experience.* In other words, phenomenal content is bound content. There are a number of arguments that can be made in support of this claim, the first of which is a *representational* one: as described in the previous paragraph, sensory representations of properties, objects, and locations can only function as representations at all because (in part) they are available to be bound. What's more, representationalism about sensory phenomenology is in fact motivated in part by the observation that phenomenal experience can be assessed for veridicality. However, unbound representations of properties, objects, and locations are not evaluable in this way, because although they have extensions, they cannot provide satisfaction conditions.[126]

The second arguments in support of the claim that phenomenal contents are bound contents is a broadly *empirical* one: recall from section 6.5 above that only the contents of an object file are available to downstream cognitive systems and to conscious awareness. Furthermore, in the "colour phi illusion" example, it was demonstrated how the experienced perceptual contents depended entirely upon the features listed in the object file. And indeed, Kahneman & Treisman (1992) explicitly claim that object file contents are perceptual contents, and Triesman (2003) argues binding is necessary for phenomenal experience.

In response to this empirical argument however, someone may object that the existence of certain deficits of binding such as *aperceptive agnosia* count as evidence against the claim that phenomenal contents are bound contents. Subjects with this condition are unable to recognize familiar objects or faces because of an inability to bind together visual information into a coherent whole. Yet despite this seeming failure of binding, patients still have some phenomenal visual experience of form, colour, and location.

---

[126] According to traditional Russellian representationalism, the phenomenal content of a sensory representation is identified with its satisfaction conditions. According to Fregean representationalism, the phenomenal content of a sensory representation is identified with a MoP of its satisfaction conditions.

Similarly, a subject with *Balint's syndrome* or *simultanagnosia* are unable to perceive more than one object at a time or spatially locate it within the context of the whole scene, yet they still have a visual phenomenal experience of the isolated object. Importantly however, the failure of *some* types of binding should not be taken to suggest that in these cases there is *no* binding occurring at all. Indeed, the variety of different dissociable types of binding suggests there is no single mechanism responsible for binding different types of sensory information together.[127] Or as Revonsuo (1999) puts it, "The different kinds of binding and disintegration at the phenomenal level suggest that normally the contents of consciousness are the result of a great variety of binding mechanisms that, to some extent at least, function independently of each other." (p.180) Thus, it is possible for certain types of binding to be impaired while others are preserved. So long as subjects still possess a basic type of binding that connects sensory features with sensory individuals, the generation of phenomenal content is still possible.

Finally, the third argument in support of the claim that phenomenal content is bound content is a *phenomenological* one: the content of sensory experience does not consist of isolated features or multiple parallel experiences of different features or parts of a scene. Rather, phenomenal experience is unified. Without binding, we would lack an explanation of how and why we experience a single, unified, multimodal perceptual field rather than unconnected sensory impressions. According to the explanation provided above, phenomenal unity is explained by *representational* unity: the binding problem is a problem about establishing relations between different types of sensory contents (namely, the different neural representations of features, objects, and spatial locations). The mechanisms responsible for binding operate on certain types of representations as input, and produce other types of representations as output. Importantly however, these latter representations need not be assumed to be "unified" representations in the sense of being integrated, multi-modal representations that directly corresponds to the contents of our phenomenal experience (something like Dennett's (1991) "Cartesian Theater"). Rather,

[127] Indeed, as noted above in section 6.1, Treisman (1996) identifies at least seven different types of binding, and Roskies (1999) lists six "perceptual" and three "cognitive" types of binding.

mechanisms of binding simply act to establish relations between distinct sensory contents.[128]

[128] In other words, we don't need to assume that there is a single, "unified" multi-modal representation somewhere 'downstream' of early sensory processing where different neural processes converge; rather, we can instead posit "distributed" representations wherein mechanisms of binding serve to make connections between different types of sensory contents.

Chapter 7

7 Conclusion

In this chapter, I review the arguments presented in chapters 1-6, where I have attempted to motivate and develop the basic outlines of an empirically-based internalist version of representationalism about sensory phenomenology. First, I will recap the arguments against externalist representationalism in order to show how it fails to provide a reductive explanation for the structure of phenomenal qualities. Next, I will attempt to explain exactly how the version of representationalism that I advocate is both (i) internalist and (ii) reductive. Finally, I will conclude the chapter with a discussion of the representationalist hypothesis and its connection with the cognitive sciences.

7.1 Externalist Representationalism

As described above, externalist representationalism is committed to two fundamentally incorrect claims about sensory representation: (i) Russellianism about phenomenal content, and (ii) the Tracking Thesis. According to the former claim, phenomenal contents are singular contents and constitutively include the referent of the representation, such that phenomenal qualities are identified with represented physical properties of external objects. According to the latter claim, the phenomenal contents of sensory representations are determined by the causal-informational covariance relations that would obtain between external physical properties and internal states in certain individuals under certain conditions. Moreover, these tracking relations ‘ground the intentionality’ of sensory states.

Above, I argued that both claims are vulnerable to two different sides of the same problem, which I called *bad structural correlation* (i.e., the bad correlation between the structure of phenomenal qualities and the structure of physical properties “tracked” by our sensory systems). More specifically, the central problem for Russellianism about phenomenal content is that it cannot provide a sufficiently response-independent explanation for the structure of phenomenal qualities (section 2.2). On the other hand, the central problem for the ‘tracking’ view of sensory intentionality is that it delivers the wrong verdict about the phenomenology of certain sensory experiences, such that the representational predictions don't match the behavioral ones (section 2.3). In the next two sections, I will briefly recap these arguments.

7.1.1 The Problem with Russellianism (about Phenomenal Content)

As noted in chapter 2, the central problem for physicalist (i.e., reductive) Russellianism is that the structure of similarity and difference relations that hold between phenomenal qualities in a given sensory modality does not match the structure of similarity and difference relations that hold between the physical properties 'tracked' by that modality. The existence of "metamers" (described in section 2.2.1) demonstrates that properties which are physically very different can produce exactly the same phenomenal experience, but the problem of bad structural correlation goes beyond mere metatmerism: given the identification of phenomenal qualities with external physical properties, externalist representationalism requires that the relational structure of phenomenal qualities must be reductively *explained* by the relational structure of the domain of physical properties. However, externalist representationalism can only individuate external physical properties by reference to their effects on subjects' sensory systems, thereby undercutting both the supposed response-independence of the physical properties as well as their explanatory potential (section 2.3.2).

In response to this challenge, externalist representationalists turn to what Pautz (2013) calls the "unnatural-relations account", which holds that there is some relation *R** that holds between physical properties which explains the relative similarities and differences of their effects on sensory systems, but which does not make essential reference to perceivers (section 2.2.1). Of course, no such relation has ever been discovered. Moreover, Pautz (2013) presents two arguments (the "metasemantic" and "radical indeterminacy" arguments) that demonstrate that even if such a relation did in fact exist, it seems nearly impossible that our *judgements* or *statements* about relative similarity could manage to refer to *R**.

Finally, even *if* these problems could be overcome, externalist representationalism still faces the other half of the problem of bad structural correlation. That is, whereas the problem for Russellianism is that very different physical properties can cause identical phenomenal experiences, it is also the case that the identical physical properties can cause very different phenomenal experiences under different conditions and in different individuals. This is the problem which faces the Tracking Thesis.

7.1.2 The Problem with the Tracking Thesis

It's important to note (as described above in section 2.3) that the Tracking Thesis plays two roles in externalist representationalism. First, it determines the phenomenal/representational content of a sensory state. Second, it grounds the intentionality of sensory states. In this section, I will deal only with the first role. I will address the inadequacy of tracking as a theory of sensory intentionality below in section 7.2.3.2.

According to the Tracking Thesis, the representational content of a sensory state is determined by the causal-informational covariance relation that would obtain between that state and external properties in certain subjects under a counterfactually-specified set of "optimal conditions". Both the relevant subjects and the optimality of perceptual conditions are typically understood in teleological terms, such that members of the same species are subject to the same veridicality conditions (section 2.3). However, this leads externalist representationalism to deny the possibility of *phenomenal variation without misrepresentation*—the possibility that two conspecifics could be veridically representing the same property while having phenomenally different experiences. Thus, the Tracking Thesis can be falsified by the existence of a case in which two phenomenally different experiences track—and veridically represent—the same external property.

Pautz (2006a) provides such an example with his "Maxwell & Twin-Maxwell" thought experiment (section 2.3.1) In this empirically-based example, Pautz demonstrates that it is possible for two individuals to be visually tracking the same physical property (due to the sameness in their retinal photoreceptors) under teleologically optimal conditions while nevertheless having different phenomenal experience (due to differences in their post-receptoral neural wiring). What's more, the externalist representationalist's best response to this problem is either (i) to retreat to a circular argument regarding optimality that threatens their reductive characterization of phenomenal qualities by making them explicable in only phenomenal terms, or (ii) to claim that the two individuals are actually tracking *different* properties, thereby giving up on the reductive, response-independent conception of the tracked properties (section 2.3.2).

Thus, externalist representationalism ultimately fails to provide an account of the structure of phenomenal qualities. Not only do the properties that they specify fail to be response-or perceiver- independent (thereby violating their own criterion of acceptability), but what's more, externalist representationalism fails to provide a *reductive explanation* for the structure of phenomenal qualities. Consequently, phenomenal qualities cannot be reductively explained by identifying them with external physical properties. Nor can it be the case that phenomenal contents are determined by the existence of certain tracking relations that would obtain under certain conditions. In short, externalist representationalism fails to provide a plausible account of sensory phenomenology. In the next section, I describe my reductive internalist alternative.

7.2 The Internalist Alternative

In contrast to the externalist view described above, the internalist version of representationalism that I advocate holds that phenomenal contents are Fregean contents, and that phenomenal qualities are modes of presentation of external properties. Moreover, it claims that the representational contents of sensory states are fixed not by 'tracking' relations that would obtain between those states and external properties in certain subjects under optimal conditions, but rather, by the relations that obtain between that sensory state *and the other sensory states in the system*. In short, this version of representationalism is based on functional role semantics, not tracking semantics. Furthermore, according to this internalist version of representationalism, the intentionality of sensory states is explained not by the existence of tracking relations, but rather, by specialized neural mechanisms that function like referential pointers and which allows subjects to be intentionally directed at objects and spatial locations.

In the following section, I will briefly recap the theory of Fregean content that underlies this version of representationalism. I will then explicitly describe how this theory of content is both internalist (section 7.2.2) and reductive (section 7.2.3). Finally, I will review how sensory intentionality is grounded in neural ostention, and not tracking.

7.2.1 Fregeanism: MoPs, CoEs, FRS, & *r*

According to the view argued for above (section 3.5), there are two distinct parts to the Fregean content of a sensory state: (i) a phenomenal mode of presentation (MoP) and (ii) response-dependent conditions on extension (CoE) that reflect the inferential role of the experience. The CoEs play a referential role by functioning as a mapping from centered possible worlds to extensions, picking out only those properties or objects that normally produce a particular response-type *r* in that subject under the current perceptual conditions (section 3.3). CoEs are thereby the way in which *sense determines reference*. In contrast, the phenomenal MoP itself can be directly identified with the response-type *r*.[129] For sensory representations of properties, this response type can be reductively explicated in terms of quality space theory (QST), which can be psychosemantically analyzed in terms of functional role semantics (FRS) (section 3.4). For sensory representations of objects and spatial locations, *r* can be identified with neural mechanism such as visual indexes and location maps, respectively. These mechanisms act as referential pointers and have no content aside from the ostensive-like “thatness” of their functional role.

In the following sections, I review the reductive explanation of phenomenal MoPs for properties, spatial locations, and objects. Moreover, I will also demonstrate how the proposed content is fully internalist. Before I turn to that however, there is one final important feature of Fregean phenomenal contents that must be noted: although isolated representations of properties, objects, and spatial locations have *extensions*, they lack *veridicality conditions*. Taken individually and alone, they are not sufficient for genuine phenomenal content. For that, representations of sensory features (properties) must be bound to representations of sensory individual (objects and locations) that are the bearers of those features. Psychosemantically, representations of sensory features act like general

[129] It is important to remember that the expression “response type” refers here to the way in which sensory systems are affected by particular stimuli, and does *not* refer to overt behavior.

terms (predicates), whereas representations of sensory individuals act like singular terms (demonstratives). Binding these different types of representations generates what I have called a "sensory sentence" with a predicate-demonstrative structure (section 6.7).

7.2.2 Internalism about Sensory Content

According to the doctrine of phenomenal externalism, the phenomenology of sensory experience is determined (at least in part) by factors that are external to the subject. Representationalism is committed to phenomenal externalism insofar as it is committed to externalism about phenomenal content. As mentioned above (section 2.1), externalist representationalism is committed to content externalism because it is committed to both the Tracking Thesis and Russellianism..[130]Above, I have argued that both tracking and Russellianism are false, and I have provided an alternative, internalist account of sensory content.

According to the Fregean representationalist view defended above, phenomenal contents are identified with a MoP of the referents/extensions/Russellian content of a sensory state. However, the referents themselves play no part in the phenomenal content of the state. Indeed, MoPs are thought to be something that is inherently 'inside the head' (typically, by definition), and are supposed to capture something like the cognitive significance of a representation. However, it is worth examining precisely what these narrow Fregean contents *are*, to show how and why they are in fact internalist.

With respect to properties, according to FRS/QST representationalism, phenomenal contents (MoPs) are identified with locations in quality space (section 3.3.3 & 3.4). On

[130] Importantly, as pointed out in chapter 1, Tracking and Russellianism only entail phenomenal externalism *together with a commitment to reductionism*. For instance, one could be a Russellian about phenomenal content and identify phenomenal qualities with the referents/extensions of sensory representations (i.e., extra-mental, external properties) while denying tracking and holding that Russellian content is determined solely by internal factors. Versions of internalist Russellian representationalism of this sort include non-reductive types such as primitivism and projectivism, and dispositionalism. Or, one could hold a tracking view of sensory intentionality and hold that the referents/extensions of sensory representations is determined by the causal-information relations that would obtain between brain states and external properties under optimal conditions while denying the Russellian claim that phenomenal content is constituted by referents/extensions. Non-reductive versions of internalist tracking-based representationalism of this sort are defended by Chalmers (2004) and Thompson (2009). However, as stated in chapter 1, I am only interested in reductive versions of representationalism here.

this view, the phenomenal content of a sensory representation of a property is exhausted by the structural relations that it bears to other representations within the sensory sub-system, and is therefore fully determined by internal factors.

With respect to locations, according to the account of sensory spatial representation described in chapter 5, Fregean spatial contents are the product of coordinating different types of modality- and motor- specific spatial maps. The contents of the resulting higher-order, distributed sensorimotor representations are determined by the short-armed functional relations that are established (via the process of coordination) between the different types of lower-order internal states that go into their construction. In other words, the Fregean spatial content of sensory experience is ultimately exhausted by relations between different receptoral inputs and motor outputs. Thus, once again, the content is fully internalist.

Unlike sensory representations of spatial location, which have a kind of sensorimotor content described above, sensory representations of objects have no representational content aside from the ‘thatness’ of their functional role. The response type identified with the MoP for objects is the indexing of a feature cluster by a FISNT. Importantly however, the features by which some object is segmented from the background are typically not available to the subject or even downstream consumer mechanisms. Nevertheless, the mechanism is itself fully internalist.

Finally, it must be noted that sensory representation of properties, locations, and objects must be bound together to generate phenomenal content. However, since (i) as was just described, the unbound representations of properties, locations, and objects are fully internalist in terms of their content, and (ii) the process of binding is fully determined by internal factors (described in chapter 6), it follows that the content of the resulting “sensory sentence” is determined wholly by internal factors, and is thereby fully internalist.

It should also be noted that the Fregean content of all three types of sensory representations includes CoEs that determine the extensions of those representations, although the CoEs are not part of the phenomenal content of the experience. However, the CoEs are also fully determined by internal factors: specifically, they can be psychosemantically analyzed in terms of FRS, and reflect the inferential role of an experience, which is itself a type of internal ("short-armed") functional role. Thus, CoEs pose no challenge to internalism about the Fregean content of sensory states.[131]

7.2.3 Reduction

Recall that in chapter 1, reduction was defined as being explicable in non-phenomenal terms. Also, recall that there are two parts to a representationalist theory: (i) a theory of sensory intentionality, and (ii) a theory of sensory content. In order to be reductive, both parts of the theory must be fully explicable in non-phenomenal (i.e., naturalistically acceptable) terms.

I will first examine the different reductive explanations for the Fregean content of sensory representations of properties, locations, and objects, respectively. It is important to recall that on the current proposal, Fregean content has two parts: (i) a phenomenal MoP and (ii) CoE that determine reference. As described above in section 3.2, the CoE is grounded in the inferential role of the experience, which in turn can be reductively explicated in broadly functionalist terms. The CoE therefore pose no challenge to reductionism. Thus, in what follows I will focus only on the reductive explanation of the phenomenal MoP itself.

7.2.3.1.1 Reduction and Sensory Content I: *Properties*

As described above, phenomenal qualities are modes of presentation of external properties, which can be reductively explicated in terms of quality space theory. This reductive explanation for phenomenal qualities proceeds roughly as follows:

First, various psychophysical tests (such as relative similarity, matching, and just-noticable difference) are used to determine the structure of a subject's perceptual discriminations in a given modality (section 4.4). The resulting psychophysical data is

[131] It should be noted here that although the CoE play no role in fixing the phenomenology of the experience (and therefore do not present a challenge to phenomenal internalism), they are nevertheless part of the Fregean content of the experience, which should ideally be captured in wholly internalist terms.

then subject to statistical procedures such as multidimensional scaling (section 4.4), which produces a unique ordering of the features discriminated by the subject: an appearance property quality space.[132] The structure of phenomenal quality space can then be extrapolated from the structure of appearance property space (section 4.5).

The result of this process is a structural characterization of the phenomenal qualities in a given sensory modality. Importantly however, this relational structure still needs to be *explained*. (That is, just because psychophysics can *discover* the structure of phenomenal quality space doesn't mean that one has thereby *reductively explained* phenomenal qualities.) This reductive explanation must be provided by neuroscientific theories that postulate neural mechanisms which (i) underlie our capacities of perceptual discrimination, and (ii) explain the structure discovered by psychophysics. One such type of explanation is opponent process theory, which postulates post-retinal channels with an antagonistic processing structure that code for position along a particular dimension of colour space (sections 3.3.3 and 4.5).

I call this an example of *mechanistic-structural* explanation (section 4.6). That is, a successful reductive explanation of phenomenal qualities must satisfy two criteria: first, because the explanandum is a particular relational structure, whatever neural mechanisms are posited must show how its operations can generate that structure (or how the relations between its states reflect that structure). Second, the mechanism must also be at the right place in the relevant causal chain between sensory receptors and overt behavior/perceptual judgements, such that its operations explain (in part) why we make a particular perceptual judgement given some particular sensory input.

Finally, according to QST, phenomenal qualities are *exhausted* by their relational structure (section 4.7), such that there is no 'intrinsic nature' of phenomenal qualities that is being left out (*pace* Chalmers.) Thus, once one has provided the sort of explanation outlined above (and detailed in chapter 4), one has fully and reductively explained phenomenal qualities.

---

[132] The ordering is said to be 'unique' because it has to satisfy the criterion of *global indiscriminability*, described in section 4.4

7.2.3.1.2 Reduction and Sensory Content II: *Spatial Locations*

The reductive explanation for sensory representation of spatial locations proceeds in a similar fashion. First, psychophysical experiments are used to determine the topographic structure of spatial relations in a given sensory modality (section 5.2). Then, one finds modality-specific neural mechanisms that are both causally relevant (in the appropriate way) to the process of spatial discrimination in that modality and which can explain the structure of those discriminations (section 5.2). Most commonly, these mechanisms are *feature maps*: topographically-organized neural structures whose spatial organization in the brain mirrors adjacency relations of external spatial locations (section 5.3). Furthermore, there are also *motor maps* that specify spatial locations in terms of a coordinate scheme defined in terms of motor activity. (For example, one can use the relative spatial position (and orientation) of body parts to specify a location in "joint-angle space" (section 5.4).)

Importantly however, the existence of topographically-organized feature maps and motor maps alone are not sufficient to explain our capacities for spatial localization (section 5.4). Nor are they sufficient to explain how and why the content of such representations should be regarded as being distinctly *spatial*. To explain these facts, information from maps in different sensory modalities and motor systems must be *coordinated* (section 5.4). This process of sensorimotor coordination can be reductively explicated in terms of a variety of different mechanisms, such as the integration in the SC, where auditory spatial information is coordinated with visual and motor spatial information via multi-modal cells that respond to both auditory and visual stimuli and which encode spatial relations in terms of a motor-based coordinate scheme known as *gaze shift*—the change in eye position required to visually orient towards a stimulus (section 5.6).

7.2.3.1.3 Reduction and Sensory Content III: *Objects*

The response type *r* that is identified with the MoP for sensory representations of objects can be reductively characterized in terms of Pylyshyn's visual indexes (a.k.a., "FINSTs"). As described above (section 6.4), psychophysical evidence had demonstrated that subjects are capable of tracking approximately 4-5 discrete visual objects at a time, across changes in both their perceptible features and in their spatial location. Pylyshyn explains this capacity by positing pre-attentive mechanisms (FINSTs) that automatically lock onto feature clusters which correspond to individual objects. These mechanisms first segment the relevant feature clusters from the rest of the visual scene, via a purely stimulus-driven, bottom-up process. These segmented feature clusters are then assigned to a visual index that can track the object, even across some changes in the feature-cluster which drove the initial segmentation process. Importantly, the assignment of feature clusters to a visual index can be modulated by top-down cognitive influences.

7.2.3.1.4 Reduction and Sensory Content IV: *Binding*

As described above, the binding problem is in fact two distinct but related problems: the "selection" problem of *sensory individuals*, and the "encoding" problem of *sensory integration* (section 6.1). The former problem can be reductively explained by appeal to mechanisms such as feature maps and visual indexes described in the previous two sections. However, it is important to recognize that the representation of sensory individuals *precedes* the extraction and encoding of sensory features instantiated by those individuals (section 6.2), such that features are "selected" to be bound together by virtue of belonging to the same sensory individual. Thus, the representation of sensory individuals plays a crucial role in the process of binding.

On the other hand, the encoding/integration problem is reductively explained by reference to *object files* (section 6.5). As described above, object files are "temporary episodic representations" or "mid-level representations" that function to bind different types of sensory (and also conceptual/semantic) information. The indexing of a feature cluster to a visual index or the focusing of attention to a location on a master map allows for the generation of an object file that is linked to that referential pointer. Sensory information about the indexed object or attended spatial location can then be extracted and encoded in such a way that the file can be said to contain a 'list' of perceptible features that are 'tagged' to the file. Finally, any and all features that are 'tagged' to the same object file are bound together and to the sensory individual linked to that object file.

7.2.3.2 Reduction and Sensory Intentionality

According to what I have called "sensory ostension", sensory intentionality is explained not via the existence of some causal co-variation relation that holds between sensory states and external properties in certain subjects under teleologically-determined optimal conditions (i.e., tracking), but rather by demonstrative-like referential mechanisms such as location maps and visual indexes. These neural mechanisms play the role of the response type *r*—and thus, the MoP—for sensory representation of spatial locations and objects, respectively,

Of course, sensory representations *already* have a referential component in their CoEs, which is itself grounded in the functional/inferential role of the experience (section 3.5). Importantly however, although CoEs determine reference, they are not *intentionally directed* in the way that these mechanisms are, because they lack a direct causal-informational connection with the external world. Thus, they cannot ground sensory intentionality.

Put somewhat differently, because CoEs are psychosemantically analyzed in terms of FRS (section 3.5), their referential powers derive solely from their internal functional role. However, in order to be intentionality directed, an internal state must be grounded via some kind of direct connection with the external world (section 6.6). This grounding is precisely what neural mechanisms such as visual indexes and location maps can contribute to sensory representation. In short, these types of mechanisms are essential to explaining how things *inside the head* manage to establish relations to things *outside the head* (section 6.6), such that the former can be said to be 'about' the latter.

These different mechanisms ground the intentionality of sensory states in different ways. With respect to sensory representations of objects, the intentionality of the sensory state is grounded by visual indexes which act as direct causal connections with objects in the external world (section 6.6). With respect to sensory representations of locations, the intentionality of the sensory state is grounded by the coordination of sensory and motor maps which allow for a kind of sensorimotor directedness towards spatial locations, via a location on a master map.

7.3 Concluding Remarks

In this monograph, I have attempted to present the outlines of a representationalist theory of sensory phenomenology that is compatible with and informed by the basic empirical framework of neuroscience and psychology. More specifically, I have attempted to provide a rudimentary sketch of a psychosemantic theory for interpreting the explanations posited by neuroscientists and psychologists regarding the information-processing mechanisms that underlie the structure of our sensory experience. I have called this view "methodological representationalism" because it is specifically focused on interpreting the empirical methodology of such explanations in psychosemantic terms. In this final section, I will explain some of the reasoning which motivates this view.

As described in chapter 1, representationalism is the claim that the phenomenology of sensory experience is identical with and exhausted by its representational content. Different versions of representationalism will identify phenomenal character with different types of content. As noted above, I do not attempt to argue here for the *truth* of representationalism; that is, I do not make any attempt to evaluate arguments either for or against the representationalist hypothesis. Rather, my goal is to analyze the prospects of a reductive version of representationalism, with a specific focus on the question of whether a reductive representationalist theory should be *internalist* or *externalist*. Thus, I have taken representationalism as a background assumption in the preceding chapters.

However, I do think that representationalism has many good arguments in its favor: it explains the intentionality of phenomenal experience, it explains the *transparency* or *diaphaneity* of experience[133], the classic problem of hallucination reduces to the

[133] It is often claimed that sensory experience is "transparent" in the sense that subjects typically "see through" their experience to external properties and objects (Harman, 1990). The claim, in other words, is that we are not aware of any intrinsic features of experience, but rather, only the features *represented by* the experience. This is supposed to distinguish the representationalist from the "unrepentent qualia freak" (Tye, 2009), who holds that sensory experiences have intrinsic features (qualia) which are the objects of our direct awareness. Importantly, externalist representationalists have often misinterpreted the doctrine of transparency, taking it to simply mean that we are directly aware of external physical properties and objects. However, this interpretation assumes a Russellian view of phenomenal content in which those external physical properties and objects are literally constituents in the content of the state. The *correct* way to understand transparency from the perspective of representationalism is as the claim that we are only directly aware of the *content* of our sensory representations, and not the *vehicles* of those representations.

(tractable) problem of intentional inexistence, and—most importantly for my purposes here—it seems to offer a naturalistically acceptable explanation for sensory phenomenology. That is, assuming that both sensory content and sensory intentionality can be explicated in non-phenomenal terms (section 1.4), sensory phenomenology can be given a reductive explanation.

Indeed, representationalism is a popular view in contemporary philosophy of mind for precisely these reasons. What's more (and more importantly), representationalism is often implicitly assumed in empirical claims about sensory experience. And such an implicit assumption makes perfect sense given the ubiquity of the concept of mental representation in cognitive science: the neural mechanisms posited to explain some psychological phenomenon are routinely characterized as *information processing* mechanisms that guide behavior by realizing states that function as representations of external stimuli. (Indeed, many have argued that the feature of modern cognitive science that distinguished it from its behaviorist predecessors is its invocation of internal representations.) However, despite the frequency of the appeal to notions of representation, neuroscientists and psychologists rarely (if ever!) have any specific psychosemantic theory of content in mind.[134] Thus, there is a need and an opportunity for a philosophical theory of representation and content that can ground the appeal to intentional notions that is implicit in empirical theorizing about sensory experience.

What's more, representationalism about sensory phenomenology is also particularly well-suited to the empirical methodology of cognitive science because it seems to offer a "best of both worlds" scenario with respect to questions of *ontology*: it allows one to make a metaphysical distinction between (i) the physical state that is the *vehicle* of representation and (ii) the *content* of the representation, which is identified with the phenomenology itself. Thus, cognitive scientists can treat the brain state(s) that realize a sensory experience as distinct from the phenomenology of that experience, while nevertheless

---

Clearly, this understanding of transparency is compatible with the Fregean version of representationalism that I have argued for here.

[134] This should *not* be taken as any sort of criticism of psychologists and neuroscientists – indeed, it's simply not their job to wonder about whether mental content is Russellian or Fregean!

both retaining a strong logical connection between brain states and phenomenology and maintaining a commitment to physicalism.

For these reasons and more, I believe that representationalism is an extremely promising theory with which to analyze sensory phenomenology, and I have attempted to provide the basic outlines of how this philosophical view can be integrated with the empirical methodology of psychology and neuroscience. However, it should be noted that the sketch that I have presented in the preceding chapter is just that—a *sketch*. It is woefully incomplete in many, many ways, and should thus be understood as a mere outline of what a fully worked-out representationalist theory should look like. Nevertheless, even in this rudimentary form, methodological representationalism seems to offer something incredibly extraordinary and valuable: a naturalistically acceptable explanation of sensory phenomenology.

Bibliography

Ahveninen, J., Kopčo, N., and Jääskeläinen, I. (2014). “Psychophysics and neuronal bases of sound localization in humans”. *Hearing Research* 307: 86-97

Aitkin, L. M. & Schuck, D. (1985). “Low frequency neurons in the lateral central nucleus of the cat inferior colliculus receive their input predominantly from the medial superior olive”. *Hearing Research* 17:87–93

Akins, K. (1996). “Of Sensory Systems and the "Aboutness" of Mental States”. *Journal of Philosophy* 93 (7):337--372

Armstrong, D. (1981). “What is consciousness?”. In John Heil (ed.), *The Nature of Mind.* Cornell University Press

Bechtel, W. & Abramson, A. (2005). “Explanation: A Mechanistic Alternative”. *Studies in History and Philosophy of the Biological and Biomedical Sciences* 36: 421-441

Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge

Binns, K.E., Grant, S., Withington, D.J., Keating, M.J. (1992). “A topographic representation of auditory space in the external nucleus of the inferior colliculus of the guinea-pig”. *Brain Research* 589(2): 231-42

Block, N. (1990). “Inverted Earth”. *Philosophical Perspectives* 4:53-79

Block, N. (1998). “Conceptual Role Semantics”. In Edward Craig (ed.), *Routledge Encyclopedia of Philosophy*. Routledge. 242-256

Block, N. (2007). “Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience”. *Behavioral and Brain Sciences* 30 (5):481-548

Block, N. (2011). “Perceptual Consciousness Overflows Cognitive Access”. *Trends in Cognitive Sciences* 15 (12):567-575

Boghossian, P.A. & Velleman, J.D. (1989). “Color as a Secondary Quality”. *Mind* 98 (January):81-103

Brogaard, B. (2010). “Strong Representationalism and Centered Content”. *Philosophical Studies* 151: 373–392

Brogaard, B. (2011). “Centered Worlds and the Content of Perception”. In S. Hales, (ed.), *Blackwell Companion to Relativism*. Blackwell: 137-158

Byrne, A. & Hilbert, D. (2003). “Color Realism and Color Science”. *Behavioral and Brain Sciences* 26 (1):3-21

Byrne, A. & Hilbert, D. (2004). “Hardin, Tye, and Color Physicalism”. *Journal of Philosophy* 101 (1):37-43

Byrne, A. & Tye, M. (2006). “Qualia Ain't in the Head”. *Noûs* 40 (2):241-255

Byrne, A. (2001). “Intentionalism Defended”. *Philosophical Review* 110 (2):199-240

Byrne, A. (2004). “What Phenomenal Consciousness is Like”. In Rocco J. Gennaro (ed.), *Higher-Order Theories of Consciousness: An Anthology*. John Benjamins

Byrne, A. (2011). “Sensory Qualities, Sensible Qualities, Sensational Qualities”. In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *The Oxford Handbook of Philosophy of Mind*. OUP Oxford

Carnap, R. (1928). *Der Logische Aufbau der Welt*. Felix Meiner Verlag

Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge University Press

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press

Chalmers, D. (2004). “The Representational Character of Experience”. In Brian Leiter (ed.), *The Future for Philosophy*. Oxford University Press 153-181

Chalmers, D. (2005). “Representationalism Showdown”. *Fragments of Consciousness.* Retrieved from http://fragments.consc.net/djc/2005/09/representationa.html

Churchland, P. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. MIT Press

Churchland, P. (1995). *The Engine of Reason, the Seat of the Soul: A Philosophical Journey Into the Brain*. MIT Press.

Churchland, P. (2005). “Chimerical Colors: Some Phenomenological Predictions From Cognitive Neuroscience”. *Philosophical Psychology* 18 (5):527-560

Clark, A. (1992). *Sensory Qualities.* Clarendon

Clark, A. (2000). *A Theory of Sentience*. Oxford University Press

Clark, A. (2004). “Feature-Placing and Proto-Objects”. *Philosophical Psychology* 17 (4):443-469.

Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press, Clarendon Press.

Davidson, D. (1987). “Knowing One's Own Mind”. *Proceedings and Addresses of the American Philosophical Association* 60 (3):441-458

De Valois, K. & De Valois, R. (1993). “A Multi-Stage Colour Model”. *Vision Research* 33(8): 1053-1065

Dennett, D. (1987). *The Intentional Stance.* MIT Press.

Dennett, D. (1988). “Quining Qualia”. In Chalmers, D (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press

Dennett, D. (1991). *Consciousness Explained*. Penguin

Dennett, D. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Bradford

Devitt, M. (1990). “Meanings Just Ain't in the Head”. In G. Boolos (ed.) *Meaning and Method: Essays in Honor of Hilary Putnam.* Cambridge University Press

Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press

Dretske, F. (1993). “Experience as Representation”. *Philosophical Issues* 13 (1):67-82

Dretske, F. (1995). *Naturalizing the Mind*. MIT Press

Dretske, F. (1996). “Phenomenal Externalism, or If Meanings Ain't in the Head, Where Are Qualia?”. *Philosophical Issues* 7:143-158

Egan, A. (2006). “Secondary Qualities and Self-Location”. *Philosophy and Phenomenological Research* 72(1): 97-119.

Fang, F., Boyaci, H., Kersten, D., & Murray, S.O. (2008). “Attention-dependent representation of a size illusion in human V1”. *Curr Biol* 18:1707-1712

Feigenson, L. & Carey, S. (2003). “Tracking Individuals via Object-files: Evidence from Infants’ Manual Search”. *Developmental Science* 6(5): 568-584

Fechner, G. (1877). *In Sachen der Psychophysik*. Breitkopf & Härtel

Fechner, G. (1882). *Revision der Hauptpuncte der Psychophysik*. Breitkopf & Härtel

Fodor, J. & LePore, E. (1996). “Churchland on State Space Semantics”. In Robert N. McCauley (ed.), *The Churchlands and Their Critics*. Blackwell Publishers 145--158

Fodor, J. (1975). *The Language of Thought*. Harvard University Press

Fodor, J. (1990). *A Theory of Content and Other Essays*. MIT Press

Friedman, M. (1999). *Reconsidering Logical Positivism*. Cambridge University Press

Gardner, M. (1991). “The Ozma Problem and the Fall of Parity”. In James Van Cleve & Robert E. Frederick (eds.) *The Philosophy of Right and Left*. Kluwer

Goodman, N. (1977). *The Structure of Appearance.* Harvard University Press

Grice, H.P. (1957). “Meaning”. *Philosophical Review* 66 (3):377-388

Grice, H.P. (1961). “The Causal Theory of Perception”. *Proceedings of the Aristotelian Society* 121:121-152

Grush, R. (2000). “Self, World and Space: The Meaning and Mechanisms of Ego- and Allocentric Spatial Representation”. *Brain and Mind* 1 (1):59-92

Hardin, L. (1988). *Color For Philosophers: Unweaving the Rainbow*. Hackett

Harman, G. (1987). “(Nonsolipsistic) Conceptual Role Semantics”. In Ernest LePore (ed.), *Notre Dame Journal of Formal Logic*. Academic Press 242-256

Harman, G. (1990). “The Intrinsic Quality of Experience”. *Philosophical Perspectives* 4: 31-52

Harnad, S. (1990). “The Symbol Grounding Problem”. *Philosophical Explorations* 42:335-346

Hilbert, D. (1987). *Color and Color Perception: A Study in Anthropocentric Realism*. Csli Press.

Horst, S. (2005). “Phenomenology and Psychophysics”. *Phenomenology and the Cognitive Sciences* 4 (1):1-21.

Hurvich, L.M. (1981). *Colour Vision*. Sinauer Associates

Jackson, F. (1982). “Epiphenomenal Qualia”. *Philosophical Quarterly* 32 (April):127-136.

Jackson, F. (2007). “Colour for Representationalists”. *Erkenntnis* 66 (1-2):169-185

Jay, M.F. & Sparks, D.L. (1987). “Sensorimotor Integration in the Primate Superior Colliculus”. *Journal of Neurophysiology* 57(1):35-55

Jeffress, L. (1948). “A Place Theory of Sound Localization”. *J Comp Physiol Psychol* 41:35–39

Johnson, K.O. & Phillips, J.R. (1981). “Tactile spatial resolution. I. Two-point discrimination, gap detection, grating resolution, and letter recognition”. *Journal of Neurophysiology* 46 (6): 1177-1192

Kahneman, D., Triesman, A., & Gibbs, B.J. (1992). “The Reviewing of Object Files: Object-Specific Integration of Information”. *Cognitive Psychology* 24: 175-219

Kolers, P.A., & von Grünau, M. (1976). "Shape and Color in Apparent Motion". Vision Research 16: 329–335

Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press.

Levine, J. (1983). “Materialism and Qualia: The Explanatory Gap”. *Pacific Philosophical Quarterly* 64 (October): 354-61.

Lycan, W. (2000). “Representational Theories of Consciousness”. Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition)

Matthen, M. (2004). “Features, Places, and Things: Reflections on Austen Clark's Theory of Sentience”. *Philosophical Psychology* 17 (4): 497-518

Matthen, M. (2005). *Seeing, Doing, and Knowing*. Oxford University Press

McGinn, C. (1982). “The Structure of Content”. In Andrew Woodfield (ed.), *Thought and Object*. Oxford University Press

Murray, S.O., Boyaci, H., & Kersten, D. (2006). “The representation of perceived angular size in human primary visual cortex”. *Nat Neurosci* 9: 429-434

Nagel, T. (1974). “What is It Like to Be a Bat?” *Philosophical Review* 83 (October):435-50

Oliver, D.L., Beckius, G.E., Bishop, D.C., Loftus, W.C., & Batra, R. (2003). “Topography of Interaural Temporal Disparity Coding in Projections of Medial Superior Olive to Inferior Colliculus”. *J Neurosci.*13;23(19): 7438-7449

Palmer, A.R. (1995). “Neural Signal Processing”. In Brian CJ Moore (ed.), *Hearing*. Elsevier. 75-121

Palmer, S.E. (1999). “Color, Consciousness, and the Isomorphism Constraint”. *Behavioral and Brain Sciences* 22 (6):923-943

Pautz, A. (2006a). “Sensory Awareness is Not a Wide Physical Relation: An Empirical Argument Against Externalist Intentionalism”. *Noûs* 40 (2):205-240

Pautz, A. (2006b). “Tracking Intentionalism and Optimal Conditions: A Reply to Byrne and Tye”
<http://philrsss.anu.edu.au/people-defaults/adamp/papers/optimal_conditions.pdf>

Pautz, A. (2013). "The Real Trouble for Phenomenal Externalists: New Empirical Evidence for a Brain- Based Theory of Consciousnes". In Richard Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. Springer 237-298

Plate, J. (2007). "An Analysis of the Binding Problem". *Philosophical Psychology* 20 (6):773 – 792

Pooresmaeili, A., Arrighi, R., Biagi, L., & Morrone, M.C. (2013). "Blood Oxygen Level-Dependent Activation of the Primary Visual Cortex Predicts Size Adaptation Illusion". *J Neurosci* 33: 15999-16008

Proske, U. & Gandevia, S.C. (2009). "The kinaesthetic senses". *J Physiol*. 587(17): 4139-4146

Pylyshyn, Z. (1999). "Vision and Cognition: How Do They Connect?". *Behavioral and Brain Sciences* 22 (3): 401-414

Pylyshyn, Z. (2001). "Visual Indexes, Preconceptual Objects, and Situated Vision". *Cognition* 80 (1-2): 127-158

Pylyshyn, Z. (2003). *Seeing and Visualizing: It's Not What You Think*. MIT Press

Revonsuo, A. (1999). "Binding and the Phenomenal Unity of Consciousness". *Consciousness and Cognition* 8 (2): 173-85

Rey, G. (1998). "A Narrow Representationalist Account of Qualitative Experience". *Philosophical Perspectives* 12 (S12): 435-58

Rosenthal, D. (2002). "How Many Kinds of Consciousness?". *Consciousness and Cognition* 11: 653-665

Rosenthal, D. (2005). *Consciousness and Mind*. Clarendon Press

Rosenthal, D. (2010). "How to Think About Mental Qualities". *Philosophical Issues* 20 (1):368-393

Roskies, A.L. (1999). "The Binding Problem". *Neuron* 24: 7-9

Russell, B. (1927). *The Analysis of Matter*. Routledge

Schrodinger, E. (1920). “Outline of a Theory of Colour Measurement for Daylight Vision”. In MacAdam (ed.) *Sources of Colour Science*. (1970) MIT Press. 134–82

Shoemaker, S. (1994). “Phenomenal Character”. *Noûs* 28 (1):21-38

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

Sperandio, I., Chouinard, P.A., & Goodale, M.A. (2012). “Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage”. *Nat Neurosci* 15:540-542

Sperling, G. (1960). “The Information Available in Brief Visual Presentations”. *Psychological Monographs: General and Applied* 74(11):1-29

Stevens, S.S. (1951). *Handbook of Experimental Psychology*. Wiley

Stevens, S.S. (1975). *Psychophysics: Introduction to its Perceptual, Neural and Social Prospects*. Wiley

Thaler, L., Arnott, S.R., & Goodale, M.A. (2011). “Neural Correlates of Natural Human Echolocation in Early and Late Blind Echolocation Experts”. *PLoS ONE* 6(5): e20162

Thompson, B. (2009). “Senses for Senses”. *Australasian Journal of Philosophy* 87 (1):99-117

Todorović, D. (1987). “The Craik-O’Brien-Cornsweet effect: New varieties and their theoretical implications”. *Perception & Psychophysics* 42: 545-560.

Tong, J., Mao, O., & Goldreich, D. (2013). “Two-point orientation discrimination versus the traditional two-point test for tactile spatial acuity assessment”. Frontiers *in Human Neuroscience* 7 (579)

Triesman, A. & Gelade, G. (1980). “A Feature-Integration Theory of Attention”. *Cognitive Psychology* 12: 97-136

Triesman, A. & Schmidt, H. (1982). “Illusory Conjunctions in the Perception of Objects”. *Cognitive Psychology* 14: 107-141

Triesman, A. (1996). “The Binding Problem”. *Current Opinion in Neurobiology* 6:171-8

Triesman, A. (1998). “Feature Binding, Attention and Object Perception”. *Phil Trans R. Soc London B* 353: 1295-1306

Triesman, A. (1999). “Solutions to the Binding Problem: Progress through Controversy and Convergence”. *Neuron* 24: 105-110

Triesman, A. (2003). “Consciousness and Perceptual Binding”. In Axel Cleeremans (ed.) *The Unity of Consciousness*. Oxford University Press. 95-113

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press

Tye, M. (1998). “Inverted Earth, Swampman, and Representationalism”. *Philosophical Perspectives* 12: 459-78

Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press

Tye, M. (2002). “Visual Qualia and Visual Content Revisited”. In David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press

Tye, M. (2003). “Consciousness, Color, and Content”. Philosophical Studies 113 (3): 233-235

Tye, M. (2009). *Consciousness Revisited: Materialism Without Phenomenal Concepts*. MIT Press

Viger, C. (1999). “The Possibility of Subisomorphic Experiential Differences”. *Brain and Behavioral Sciences* 22 (6):975-975

Viger, C., Bluhm, R., & Mosurinjohn, S. (2008). “The Binding Problem: Achilles in the 21st Century”. In Lagerlund and Yrjönsuuri (eds.) *The Achilles of Rational Psychology*. Springer

Wimsatt, W. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press

Winter, J.A., Allen, T.J., & Proske, U. (2005). “Muscle spindle signals combine with the sense of effort to indicate limb position”. *J Physiol* 568(3): 1035-1046

Wittgenstein, L. (1953). *Philosophical Investigations*. Blackwell

Zeki, S. (1993). *A Vision of the Brain*. Blackwell

Curriculum Vitae

Name:	Matthew Ivanowich
Post-secondary Education and Degrees:	York University Toronto, Ontario, Canada 1999-2004 B.A. (Honours), Philosophy
	York University Toronto, Ontario, Canada 2007-2008 M.A., Philosophy
	The University of Western Ontario London, Ontario, Canada 2008-2015 Ph.D., Philosophy
Honours and Awards:	Province of Ontario Graduate Scholarship 2011-2012, 2012-2013
Related Work Experience	Instructor, “Philosophy of Emotion” 2410F The University of Western Ontario Fall 2015
	Instructor, “Adv. Topics in Phil. of Mind: Perception” The University of Western Ontario Spring 2015
	Instructor, “Business Ethics” The University of Western Ontario Fall 2013
	Instructor, “Big Ideas” The University of Western Ontario Spring 2011
	Teaching Assistant, “Adv. Introduction to Philosophy” The University of Western Ontario Fall 2009 - Spring 2010
	Teaching Assistant, “Business Ethics” The University of Western Ontario Spring 2009

Teaching Assistant, "Environmental Ethics"
The University of Western Ontario
Fall 2008

Publications

Choinard & Ivanowich (2014). "Is the Primary Visual Cortex a Center Stage for the Visual Phenomenology of Object Size?". *The Journal of Neuroscience* 34(6):2013-2014

Ivanowich (2013). "The dlPFC isn't a NCHOT: A Reply to Sebastián". In Richard Brown (ed.) *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. Springer. 445-450

Selected Conference Presentations

"Quality Space Theory". *Annual Meeting of the Canadian Society for the History and Philosophy of Science*, University of Victoria, June 2013

"Phenomenal Qualities, Representationalism, and Cognitive Science". *6th Annual Mind Network Meeting and CoMiC Conference,* University of Edinburgh, October 2012

"Representationalism, Externalism, and the Structure of Phenomenal Qualities". *York University Philosophy Graduate Students Association Annual Conference,* May 2012

"Intentionalism: Tracking vs. Functional Role" *Philosophy Graduate Students Association Colloquium Series,* University of Western Ontario, February 2011

"Defending the Internalist Intuition: Putting Phenomenology back in the Head". Philosophy Graduate Students Association Colloquium Series, University of Western Ontario, March 2010

"A Moderate Representationalism". 1st Annual Consciousness Online Conference, Feb-March 2009.