

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Sensible Concepts: Experience and the A Priori

Permalink

<https://escholarship.org/uc/item/1mh1r9mh>

Author

Epstein, Peter Fisher

Publication Date

2017

Peer reviewed|Thesis/dissertation

Sensible Concepts: Experience and the *A Priori*

By

Peter Fisher Epstein

A dissertation submitted in partial satisfaction of the

requirements for the

degree of Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge

Professor John J. Campbell, Co-chair

Professor Barry Stroud, Co-chair

Professor Geoffrey Lee

Professor Tania Lombrozo

Summer 2017

Sensible Concepts: Experience and the *A Priori*

This dissertation, in its entirety, is the original work of the author.

Copyright
Peter Fisher Epstein
2017

Abstract

Sensible Concepts: Experience and the *A Priori*

by

Peter Fisher Epstein

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor John Campbell, co-chair

Professor Barry Stroud, co-chair

In this dissertation, I develop a novel account of spatial experience that—unlike most contemporary theories of perception—situates our experience of space within a broader context of non-sensory cognitive activities. On my account, to perceive an object as square is, in part, to deploy an *a priori* Euclidean concept of squareness – a concept that features in, but is not derived from, experience. I use this *rationalist* analysis of spatial experience to shed light on three issues: the connection between Euclidean proof and our perception of physical objects; the distinction between primary and secondary qualities; and the challenge posed to the veridicality of our spatial experience by the findings of relativistic physics.

In light of the discovery of consistent non-Euclidean geometries and the empirical evidence that our own universe is not perfectly Euclidean, many have rejected the Kantian idea that Euclidean proof gives us *a priori* knowledge of physical space. But there does seem to be a *cognitive* connection between the theorems we prove in Euclidean geometry and the spatial features we perceive physical objects to have: having proven the Pythagorean theorem, a carpenter will expect a particular relation to hold among the lengths of the sides of a right triangle she is constructing from wooden beams. This suggests that we *represent* the empirical objects we perceive as subject to the results we obtain in the domain of Euclidean geometry, even if we no longer think that such representations are guaranteed to be correct.

In order to account for this phenomenon, I develop a view on which the concepts we employ in Euclidean proof are *a priori*, but also feature in perception. After setting out the motivations for this view and offering a brief sketch of its contours, in Chapter 1, I go on to defend its central claims in Chapters 2 and 3. I argue that our use of spatial concepts in Euclidean geometry shows that they cannot be *derived* from experience: certain aspects of these concepts, such as the idea of continuity

built into our concept of a circle, outstrip anything we can glean from our sensory cognition. At the same time, I suggest, an experience of an object as square is one that deploys our *a priori* concept of squareness, and does so simply in virtue of its phenomenal character. That is, such an experience, independent of any causal relations it might bear to objects in the external world, represents its object as instantiating the very geometrical property about which we reason in Euclidean proof. It is this fact that explains why the carpenter takes the results of her *a priori* reasoning to apply to the empirical objects she perceives – her experience of shape has specific geometrical content built into it, in virtue of the *a priori* Euclidean concepts that feature in its content.

This account of spatial experience helps shed light on a topic with a long philosophical history: the distinction between primary and secondary qualities. In Chapter 4, I argue that, lacking any *a priori* grasp of a secondary quality like redness, we can represent that property only by way of its role in experience – as *whatever* property plays the relevant role in generating experiences of red. Since perception does not inform us *which* specific property is playing that role, we are left in the dark about the *nature* of the secondary qualities. By contrast, in the case of a primary quality like squareness, we are not constrained to represent the property by way of its role in experience. When we experience an object as square, we grasp the nature of the property represented, in virtue of our *a priori* concept of squareness – the property of having four equal sides joined at four right angles. Color and shape, then, feature in our cognitive lives in very different ways; these *conceptual* and *experiential* differences are, I contend, the real basis of the distinction between primary and secondary qualities.

In the final chapter of the dissertation, I consider how, on my account, we should evaluate the veridicality of our experience of shape in light of Einstein's special theory of relativity (STR). According to the standard interpretation, STR reveals that no purely spatial properties are instantiated in our universe; instead, all that objectively exists is a four-dimensional *spatiotemporal* manifold. Since, on my account of spatial perception, our experience represents the presence of purely spatial properties—the Euclidean properties about which we reason in performing abstract proof—STR might seem to imply that our spatial experience is never veridical. Against this, I argue that STR gives us no reason to abandon the idea that the Euclidean spatial properties represented in our experience are in fact physically instantiated. Instead, what Einstein's discoveries show is that those properties are instantiated in a particular *manner*: namely, relative to various inertial frames of reference. This analysis allows us to hold onto the intuitive idea that there is a tight connection between our *a priori* geometrical reasoning and our experience of space, without being forced to accept that our spatial experience is universally illusory.

ACKNOWLEDGEMENTS

The completion of this dissertation is bittersweet, as it signals the end of a very happy and fulfilling nine years at Berkeley. I could not have asked for a better community of friends, colleagues, and advisors.

I owe debts of many kinds to many people, but first and foremost, I would like to thank the members of my dissertation committee: Barry Stroud, John Campbell, and Geoff Lee.

My time at Berkeley has been profoundly shaped by Barry's wisdom, advice, and kindness. During my first semester as a graduate student, Barry told me that, despite the seemingly vast chasms separating different subfields of contemporary philosophy, properly understood, the subject is "all one question." He never let me in on what that question is. But he did show me how to approach philosophical questions of all kinds with a proper appreciation for their breadth and depth, to go slowly, and to be cognizant of the underlying point of the enterprise.

When I think back to the writing of my dissertation itself, what comes first to mind is sitting in John's office for untold hours, working out answers to his endlessly probing questions about my latest draft (questions whose significance I often came to appreciate only much later). Though he came to our discussions with intuitions and commitments radically different from my own, John always worked through my arguments with an open mind. I am grateful for his uncanny ability to zero in on the key idea in a massively complex web of philosophical argumentation, and for his willingness to share his insight with me (even when I was too stubborn—or, as he often informed me, "too much an internalist!"—to see the wisdom of his suggestions).

Geoff provided a much-needed counterweight to the "Berkeley school" of philosophy of mind I encountered when I first arrived, helping me feel like less of a philosophical outlier within the department. His knowledge about an extraordinarily wide range of topics—from metaphysics to cognitive psychology to space-time physics—was an invaluable resource as I found myself wading into unfamiliar intellectual waters.

I also want to thank Mike Martin, who served as an unofficial fourth member of my committee, offering much-needed advice about both my work and how to navigate the treacherous terrain of a rapidly-professionalizing discipline. I benefited immeasurably from Mike's attention to argumentative detail, his encyclopedic knowledge of all things philosophical (and otherwise), and his unparalleled ability to see through to the core logical structure of a paper (even when the paper's own author did not).

Over the course of this dissertation's genesis, I have benefited immensely from conversations with many other unofficial advisors, including Lara Buchak, Michael Friedman, Ned Hall, Paolo Mancosu, Susanna Siegel, and Daniel Warren. I also had the good fortune to receive a generous grant from the Mabelle McLeod Lewis Memorial Fund, which supported my research during the 2015-2016 academic.

It is difficult to imagine what my time at Berkeley would have been like without my fellow graduate students, whose feedback, friendship, and support kept me both on my toes and on my feet. I am grateful for endless conversations, in Moses Hall and elsewhere, with Austin Andrews, Adam Bradley, Zack Bruce, Caitlin Dolan, Max Gee, Jim Hutchinson, Alex Kerr, Richard Lawrence, Erich Matthes, Janum Sethi, James Stazicker, Klaus Strelau, Dave Suarez, and Stephen Thurman.

I would also like to thank my teachers and advisors from my days as an undergraduate at Harvard, especially Ned Block, Warren Goldfarb, Ned Hall, Jim Pryor, Susanna Siegel, and Alison Simmons. All of the central ideas of this dissertation can be traced, in one way or another, to my time in Emerson Hall (now so many years ago).

I owe a huge debt to my parents for their love and support, and for teaching me the beauty and value of thought, from the moment my own thoughts first emerged.

Finally, and most importantly, I want to thank my partner, Umrao Sethi, who has made both this dissertation and its author immeasurably better.

CONTENTS

Acknowledgements	i
Contents	iii
Introduction	1
1 Internalism Without Magic	6
1.1 The Magic Worry	8
1.2 The Problem of Internal Acquaintance	11
1.3 Separatism	18
1.4 The Magic Worry, Again	22
1.5 Internalist Half-Responses	26
1.6 The Line-Drawing Problem	31
1.7 Primary-Quality Contents and <i>A Priori</i> Concepts	35
1.8 Conclusion	42
2 <i>A Priori</i> Concepts in Euclidean Proof	43
2.1 Euclidean Proof as an “Axiomatic System”	43
2.2 Strawson’s “Picturable Meanings”	48
2.3 Diagrammatic Reasoning	51
2.4 Geometrical Concepts and Perceptual Appearances	52
2.5 Visual Limits	54
2.6 Conclusion	58
3 <i>A Priori</i> Concepts in Spatial Experience	62
3.1 Literal and Metaphorical Spaces	64
3.2 Commonality	71
3.3 Structuralism: An Inadequate Explanation of Transfer	75
3.4 An Account of Spatial Experience	86
3.5 Conclusion	97
4 The Distinction Between Primary and Secondary Qualities	98
4.1 The Simple View of Color	98
4.2 A Case of Disagreement	102
4.3 Error Theories	103
4.4 Alternatives to the Simple View	104
4.5 The Challenge: What is Greenness?	107

4.6	Another Case of Disagreement	109
4.7	Reversing the Conceptual Connection	115
4.8	Roloc	120
4.9	The Distance Between Shape Concepts and Experience	127
4.10	Conclusion	131
5	Shape Perception in a Relativistic Universe	134
5.1	Introduction	135
5.2	Chalmers' Argument Against Shape Presentationalism	140
5.3	Shape Presentationalism in a Relativistic Universe	153
5.4	Is Shape Intrinsic?	159
5.5	Conclusion	165
	Conclusion	167
	Bibliography	169

INTRODUCTION

“How can it be that mathematics, being after all a product of human thought which is independent of experience, is so admirably appropriate to the objects of reality? Is human reason, then, without experience, merely by taking thought, able to fathom the properties of real things?”

Albert Einstein¹

Recent work in the philosophy of mind has tended to proceed along a highly circumscribed path, offering analyses of perceptual experience that make little contact with reflection on other aspects of our cognitive lives. There is a widely-held assumption that the content of our experience must come from our empirical interactions with the world: according to the most common version of this view, the *causal* relations between the external world and our experiences determine what those experiences represent. On such a picture, questions about experience are cordoned off from consideration of other, more intellectual aspects of our cognitive lives, such as our mathematical reasoning.

But the topology of our minds is not discontinuous. Consider the following kind of case. A carpenter, who, in her spare time, enjoys doing elementary mathematical exercises, works through a proof of the Pythagorean theorem. Emerging from her study into her workshop, she sees a set of wooden beams, which she intends to use in constructing the frame of a right triangular ramp. Having performed the proof, the carpenter will take herself to know the following: if the beams serving as the legs of her triangle are three feet long and four feet long, respectively, then she will need to cut the beam serving as the hypotenuse to a length of five feet.

The wooden beams are objects of the carpenter’s perceptual awareness – they show up in her experience of the world. But she takes the knowledge she has of right triangles from having performed the Pythagorean proof—knowledge that seems to be a product of *a priori* mathematical reasoning, rather than any experiential contact with the world—to be directly applicable to those empirical objects.

As this simple case shows, highly abstract (even, perhaps, *a priori*) cognitive activities have obvious and immediate connections to experience. My working hypothesis is that we can make headway in debates about the nature of perception by attending to the fact that sensory experience does not exist in a vacuum. In particular, I think that a proper understanding of *spatial* experience requires attention to the connection

¹ Einstein (1922, p. 15).

between the practice of Euclidean proof and the way our experience represents the spatial properties of the objects we perceive.

Historically, this connection between our *a priori* and empirical representations of space has received much philosophical attention, particularly in the work of Kant. But, in light of the discovery of consistent non-Euclidean geometries and the empirical evidence that our own universe is not perfectly Euclidean, many have rejected the Kantian idea that Euclidean proof gives us *a priori* knowledge of physical space. Still, there does seem to be a *cognitive* connection between the theorems we prove in Euclidean geometry and the spatial features we perceive physical objects to have. This cognitive connection underlies the carpenter's expectation that a certain relation will hold among the lengths of her wooden beams: we *represent* the empirical objects we perceive as subject to the results we obtain in the domain of Euclidean geometry, even if we no longer think that such representations are guaranteed to be correct.

In what follows, I develop an account of spatial experience that—unlike most contemporary theories of perception—situates our experience of space within a broader context of non-sensory cognitive activities. On my account, to perceive an object as square is, in part, to deploy an *a priori* Euclidean concept of squareness – a concept that features in, but is not derived from, experience. I use this *rationalist* analysis of spatial experience to shed light on three issues: the connection between Euclidean proof and our perception of physical objects; the distinction between primary and secondary qualities; and the challenge posed to the veridicality of our spatial experience by the findings of relativistic physics.

OUTLINE OF CHAPTERS

CHAPTER 1: INTERNALISM WITHOUT MAGIC

There are familiar reasons to accept *internalism* about perceptual states – the thesis that the character of our perceptual states is determined solely by our own internal constitution. But our perceptual states are also *world-directed*: they represent a world of mind-independent objects and properties. These two thoughts sit uneasily with one another. For it seems that if our perceptual states are determined purely internally, properties of the mind-independent world—such as the colors and shapes of material objects—could show up in perception only as the unknown external causes of a subject's self-contained inner life. On this picture, a color or shape property could be represented in perception only *opaquely* – as the property, *whatever* it might be, that typically causes experiences of a certain type. This is the placeholder view.

In this chapter, I suggest that such a view makes sense when it comes to experiences of traditional secondary qualities, such as color. But, I argue, there is an important difference when it comes to experiences of traditional primary qualities, such

as shape: in the case of spatial experience, we have *a priori* concepts of the relevant properties; so, when such properties are represented perceptually, we are not constrained to represent them by way of their role in generating experiences (that is, via a placeholder). Instead, our experience deploys our *a priori* concepts of spatial properties; such experiences thus provide representations of the primary qualities that reveal their geometric natures.

CHAPTER 2: *A PRIORI* CONCEPTS IN EUCLIDEAN GEOMETRY

The picture sketched in the introduction relies on the claim that we have *a priori* spatial concepts that feature in the contents of perceptual experience. Defending this claim requires two things: establishing that we do indeed *have* spatial concepts that are *a priori*; and arguing that those very concepts feature in the contents of perceptual experience.

In this chapter, I turn to the first of these tasks, arguing that our use of spatial concepts in the practice of Euclidean proof shows that those concepts could not be derived from experience. For over two millennia, Euclid's *Elements* was seen as a paradigm of *a priori* reasoning. With the discovery of non-Euclidean geometries, and the eventual realization that our universe is itself non-Euclidean, the status of our geometrical knowledge was radically undermined. In the wake of this upheaval, philosophers adopted two revisionary interpretations of Euclidean proof. Some suggested that we understand Euclidean proof as a purely formal system of deductive logic – one not concerned with specifically geometrical content at all. Others suggested that Euclidean proof employs concepts derived from our sensory experience or imagination. I argue that both interpretations fail to capture the true nature of our geometrical reasoning. Euclidean proof is not a purely formal system of deductive logic, but one in which our grasp of the content of geometrical concepts plays a central role; moreover, our grasp of this content is *a priori*, rather than being derived from experience.

CHAPTER 3: *A PRIORI* CONCEPTS IN SPATIAL EXPERIENCE

In this chapter, I turn to the second task set out in the introduction, arguing that, although the concepts employed in Euclidean proof are not derived from experience, there is nonetheless an important connection between spatial experience and our *a priori* geometrical reasoning. We automatically apply our *a priori* spatial concepts to the objects we perceive – as when the carpenter applies the Pythagorean theorem to her wooden beams. This is the phenomenon of TRANSFER. In order to account for TRANSFER, I argue, we must acknowledge that the very concepts we use in Euclidean proof—our *a priori* concepts of spatial properties—feature in the contents of our perceptual experiences. This is the thesis of COMMONALITY.

In the second half of the chapter, I explain how COMMONALITY is possible, by sketching an account of spatial experience that can make sense of cases like the carpenter's. We have an innate, primitive grasp of basic spatial structure—a set of spatial

“proto-concepts”—that features in our cognitive lives in two ways. On the one hand, these proto-concepts are deployed in our experience of the world around us, representing the presence of particular spatial properties. On the other hand, our spatial proto-concepts are explicitly articulated in the practice of Euclidean proof, where we explore in detail the features of the spatial properties we innately represent.

CHAPTER 4: THE DISTINCTION BETWEEN PRIMARY AND SECONDARY QUALITIES

On the account developed in Chapters 1-3, primary and secondary qualities are represented in very different ways in experience. Our *a priori* grasp of spatial properties allows for such properties to be represented in experience in a way that reveals their nature. But, because we lack such *a priori* concepts of colors, our experience of a color property fails to reveal that property’s nature – experience can represent a color only by way of a placeholder.

In this chapter, I address an externalist challenge to this account of color experience, which insists that our experience presents colors, just as much as shapes, in a nature-revealing way. On this externalist picture, the character of an experience of color is itself constituted by the nature of the color perceived; so such an experience necessarily reveals the nature of that property. I reject this account because it fails to make sense of various *spectrum inversion* scenarios, which make clear that the conception of the secondary qualities we get from perceptual experience does *not* reveal the intrinsic natures of those properties. When different subjects’ color experiences diverge, disagreement eventually gives out – we are left with no conception of what the disagreement is a disagreement *about*. This is because, being derived from experience, our concepts of color properties give us no non-experiential grip on what those properties are. I contrast such cases of spectrum inversion and disagreement about color with parallel cases involving shape, in order to bring out the special role our *a priori* concepts play in giving our perceptions of spatial properties the nature-revealing content they have.

CHAPTER 5: SHAPE PERCEPTION IN A RELATIVISTIC UNIVERSE

In the final chapter of the dissertation, I address an internalist challenge to my account of spatial perception, which insists that shapes, just as much as colors, are represented in experience only via placeholder contents. The challenge here stems from reflection on Einstein’s special theory of relativity (STR). According to the standard interpretation, STR reveals that no purely spatial properties are instantiated in our universe; instead, all that objectively exists is a four-dimensional *spatiotemporal* manifold. Since, on my account of spatial perception, our experience represents the presence of purely spatial properties—the Euclidean properties about which we reason in performing abstract proof—STR would seem to imply that our spatial experience is never veridical. But this is implausible: we normally think that we do have veridical spatial experience.

So STR, combined with the plausible idea that our spatial experience is not universally illusory, would seem to undermine my claim that our *a priori* spatial concepts feature in the contents of our spatial experience. Instead, the thought goes, we must accept that shape experience, just like color experience, represents the world via placeholder contents.

Against this, I argue that STR gives us no reason to abandon the idea that the Euclidean spatial properties represented in our experience are in fact physically instantiated. Instead, what Einstein's discoveries show is that those properties are instantiated in a particular *manner*: namely, relative to various inertial frames of reference. This analysis allows us to hold onto the intuitive idea that there is a tight connection between our *a priori* geometrical reasoning and our experience of space—it allows us to insist that shape experience involves more than mere placeholder contents—without being forced to accept that our spatial experience is universally illusory.

CHAPTER 1

INTERNALISM WITHOUT MAGIC

According to Russell, Descartes's great achievement was to reveal that "subjective things are the most certain." A subject can know the contents of her own mind—know, for example, that she is in pain; what her experience is like; what she is thinking—even while entertaining Cartesian doubts about the existence and nature of the external world.¹

This specially-secure knowledge of our own minds is a reflection of the privileged access to our mental states we enjoy when we exercise our introspective capacities. A subject's introspective access is privileged in that it is only available to her; but it is also limited, in that it does not extend to the world external to her own subjectivity.

Such a picture is plausible if we take the mental states to which a subject has privileged access to be constituted by features *internal* to the subject herself. For Descartes, this meant that mental states were features of a subject's immaterial soul, and so not part of the physical world at all. For more materialistically-inclined versions of this kind of internalism about the mind, the idea is that a subject's mental states depend solely on her internal physical constitution—on the configuration of her central nervous system.²

The internalist understanding of the mind has been the default view in modern philosophy since Descartes first articulated his version of it. But, in the last decades of the twentieth century, analytic philosophy of mind began to take what we might call an "externalist turn." Today, the internalist thesis that a subject's mental states are determined solely by her internal constitution is denied by a variety of philosophical accounts of the mind. The details of these various externalist accounts—which include *naïve realism* (or "disjunctivism") and *wide representationalism*—differ greatly, but they are unified by a rejection of the internalist claim that a subject's mental states are constitutively independent of the external world.

¹ See Russell (1912, p. 18). Russell ends up rejecting the fully Cartesian picture, on which knowledge of one's own mind is *infallible*; but he accepts the Cartesian insight that knowledge of subjective facts is, at least, *more* secure than knowledge of facts about mind-independent reality.

² The switch from the metaphorical sense of "internal" at work in Descartes's internalism to the more literal sense endorsed by contemporary materialist versions of the view is emphasized by McDowell (1986, p. 167). The contrast between the two versions of internalism is significant in its own right, but, for the purposes of my discussion, what is important is the contrast between views that accept internalism in *either* sense, and views that deny such a thesis.

The externalist turn has been driven by reflection on what our privileged, introspective access is access *to*. When a subject comes to know about her own mind, *what* she comes to know about is *a point of view on the world*. Our mental states—including, centrally, our perceptual experiences—seem to reveal a *mind-independent world* to us: a world of material objects arranged in space, with various colors, shapes, and other sensible qualities. The reason this *world-directedness* of the mind has led to a rejection of internalism is that any account of subjectivity will have to make intelligible *how* our mental states can represent a mind-independent world, and furnishing such an explanation seems impossible if we take mental states to be internally constituted. How could a mental state that is constitutively independent of the world external to the subject come to have representational bearing on that world? This is the MAGIC WORRY: according to many advocates of externalism, the reason we must abandon the default internalist view is that, on internalism, our minds' world-directedness seems explicable only by appeal to a *magical* connection between internally-constituted mental states and mind-independent reality.

My aim in this chapter is to resuscitate the default internalist view of the mind, by offering a response to the MAGIC WORRY. Though I think the externalist's challenge can be met, I also want to acknowledge that the threat it poses to internalism is quite pressing. The internalist is motivated by the Cartesian idea that we have special access to our own subjectivity; a central feature of that subjectivity—according to some, “the most conspicuous phenomenological fact there is”³—is the *world-directedness* of our perceptual experiences. To do justice to the Cartesian insight, then, the internalist must make intelligible how states that are internally constituted can represent a mind-independent world. And this is an explanatory obligation that internalists have failed to meet.

The externalist's challenge is a contemporary one, but, as I will show in what follows, it can be answered if we attend carefully to a doctrine with a long philosophical history: the distinction between primary and secondary qualities. Properly understood, the distinction can be used to delineate the *extent* of our introspective access to our internally-constituted mental states. We do not, in the end, have insight into *every* aspect of the world-directed contents of our perceptual experiences. In particular, the determinate contents of our *secondary*-quality perceptions are not fully accessible to introspection, since those contents are fixed by the external world. We do, however, have introspective insight into the contents of our perceptions of *primary* qualities because those contents are internally determined, in a straightforward sense: they employ *a priori* spatial concepts that we grasp independently of our experience of the external world.

Before we can appreciate this resolution of the MAGIC WORRY, we will need to explore in detail how the WORRY arises; why the externalist views proposed as a

³ See McDowell (1986, p. 152).

response to it must be rejected; and why extant versions of internalism fail to alleviate it. I turn to these tasks below.

1.1. THE MAGIC WORRY

In “Brains in a Vat” (1981), Hilary Putnam presents a now-classic argument against the traditional internalist view, in the form of a critique of internalism’s treatment of skeptical scenarios. Such scenarios are supposed to illustrate that the contents of our mental states—what those states represent to be the case in the external world—might fail to line up with the way the external world actually is. The brain in a vat, for example—hooked up to a computer that supplies it with peripheral inputs matching those of a normally-embodied brain, and thus in the same internal state as a normal subject—has, according to the traditional view, perceptual experiences that represent various kinds of material objects, when no such objects are in fact present. And so the brain’s perceptual states are systematically illusory: they *misrepresent* the world as being one way, when it is quite another.

For the internalist, the brain-in-a-vat scenario is supposed to draw out certain epistemological worries about our knowledge of the external world. But what Putnam emphasizes about this traditional picture is that, in taking the brain in a vat to be *misrepresenting* the world as containing, e.g., physical trees, the internalist must, implicitly, be relying on some theory of mental content, some account of what makes it the case that the brain’s mental states do indeed *represent trees* in the first place. And, Putnam argues, the account in question could only be based on a kind of magical thinking.

Putnam begins his attack on the internalist’s account of mental content by calling attention to the way that typical *non-mental* representational items, like physical pictures and written words, come to have representational significance. Consider the orthographic string BOOT. It is obvious that this string, considered solely in terms of its intrinsic features (the matching ovals in the center, the lines meeting at a right angle on one side), has no *necessary* connection to what it in fact does represent—namely, a kind of sturdy shoe—in the representational system of the English language. BOOT’s intrinsic features give no hint as to its representational significance, because that significance derives from certain extrinsic connections: the complex set of causal and contextual links between the orthographic string and the represented entity that the sociolinguistic community of English speakers has established. The contingency of these extrinsic links, and of the representational connection they establish, can be brought out by considering the representational significance of the string BOOT in the *German* language. In *that* system, BOOT does not have the relevant sociolinguistic links to boots; it has such links, instead, to *boats*. As a result, the string BOOT represents a boat, not a boot, when it is written out by a German speaker. Thus BOOT—considered just in terms of its *intrinsic* features—is not “destined” to represent what it in fact represents in English (rather than what it

represents in German). And the point generalizes. As Putnam says, “No physical object can, in itself, refer to one thing rather than to another.”⁴ To suppose otherwise is to suppose that a representational item can have a kind of magical connection to what it represents; it is to think that the arcs and angles of BOOT could somehow conjure up a particular kind of item—a boot—all on their own.⁵

In the brain-in-a-vat scenario, the internalist assumes that there is a representational connection between the envatted brain’s internal states—the ones it shares with a normal human subject—and material objects, like trees. On the traditional picture, when the brain in a vat is in the configuration *my* brain is in when I look at a tree, the brain has a perceptual state that misrepresents the presence of a tree. But, as Putnam emphasizes, the state in question—the one that is typically caused by trees in *me*—has no causal or contextual links to trees when it occurs in an envatted brain. So the internalist must, Putnam suggests, be implicitly assuming that the *intrinsic* features of the envatted brain’s mental state somehow establish a representational connection to trees. The internalist’s account of mental representation thus requires that mental states have just the feature that physical representations cannot have: intrinsic representational significance.

This is where the MAGIC WORRY arises. If magic would be required in order for physical representations to have intrinsic representational powers, why would mental representations be any different? John McDowell has suggested that it is in fact an inchoate awareness of this worry that drives Descartes (the traditional internalist) to posit a non-physical soul. In McDowell’s reconstruction, the internalist, faced with the question of how a mind with no links to material objects like trees—a brain in a vat, or a Cartesian ego at the mercy of an evil demon—could nonetheless manage to represent such entities, reasons that “[m]agic might seem to help, and magical powers require an occult medium.”⁶ But positing magical powers does not really become more palatable when we imagine that the *organ* in which those powers supposedly reside is itself magical.

Like McDowell, Putnam’s central objection to the traditional internalist picture

⁴ Putnam (1981, p. 2).

⁵ An important note: When I use the all-caps form BOOT, I am picking out the physical item in question as an orthographic string, and not as a *word*. The word “boot” of the English language and the word “boot” of the German language are different *words*; but they are embodied in a common orthographic string type (namely, BOOT). Thus, I do not mean to be challenging Kaplan’s (1990) view on how *words* are to be individuated, or to advocate for the “orthographic conception” that he rejects. Instead, I simply want to point out that we *can* consider orthographic strings, and orthographic string types, themselves, and ask about their representational capacities. Doing so amounts to considering the items we use as words—a use that requires various *extrinsic* connections to language communities—solely in terms of their *intrinsic* features.

⁶ McDowell (1986, p. 153). Putnam finds a very similar argument for the conclusion that the mind is non-physical in the work of Brentano (see p. 3, p. 17).

stems from the MAGIC WORRY. Here is how Putnam presents the challenge:

What is important to realize is that what goes for physical pictures also goes for mental images, and for mental representations in general; mental representations no more have a necessary connection with what they represent than physical representations do. The contrary supposition is a survival of magical thinking. (p. 3)

[E]ven a large and complex system of representations, both verbal and visual, still does not have an intrinsic, built-in, magical connection with what it represents—a connection independent of how it was caused.... Thought words and mental pictures do not intrinsically represent what they are about. (p. 5)

Putnam's point is that the traditional internalist account of mental representation grants just such intrinsic representational powers to mental states. And that, Putnam would say, is a magical kind of representation.

Driven by the MAGIC WORRY, Putnam rejects the traditional view that a brain in a vat and a normally-embodied human subject, in virtue of their shared internal constitutions, have states that represent the same external-world features. Instead, Putnam argues that the only way to understand how our mental states can represent features of the external world is to accept that the content of a given mental state is itself determined by the external-world features that are causally correlated with that state. This is the thesis of *content externalism*. The idea is that our mental states come to have their representational significance—they come to represent trees and other features of the mind-independent world—in much the same way that all representational items come to represent what they do: not intrinsically (via magic), but in virtue of extrinsic, contingent links to those features. On this picture, the brain in a vat—since it interacts not with material bodies, but rather with elements of a computer program—has states that (veridically!) represent elements of a computer program, rather than (erroneously) representing material bodies.

So Putnam, in rejecting the “magical” theory of mental representation on which the traditional treatment of the brain-in-a-vat scenario depends, articulates an alternative theory of mental content that implies that the *epistemological* worry allegedly brought out by the scenario—the worry that there could be a widespread *mismatch* between the contents of a subject's mental states and the features of her world—can be dismissed as an impossibility. The very nature of non-magical representation—the fact that what a representational item represents is always its normal cause, whether that cause is a tree or a computer-programming element—means that there could never be such a mismatch between mind and world. Even brains in vats get things right: their mental states are veridical representations of their world.

Putnam's externalism would thus seem to have two key advantages over the tradi-

tional internalist view it seeks to overthrow. First, it avoids the MAGIC WORRY, by giving mental representation the same kind of causal-correlation analysis that seems so unproblematically non-magical in other contexts. Second, it gives us a kind of guarantee against skeptical doubts: on Putnam’s view, we can know that there is no widespread mismatch between the contents of our mental states and the world in which we live, because such a mismatch is a conceptual impossibility. Despite these advantages, however, Putnam’s account of the mind is not sustainable. As I will argue in the next section, the advantages themselves contribute to a fatal flaw: on Putnam’s picture, we lose our grip on the contents of our own minds.

1.2. THE PROBLEM OF INTERNAL ACQUAINTANCE

In this section, I argue that Putnam’s account of mental representation leads to what I call the PROBLEM OF INTERNAL ACQUAINTANCE. The argument proceeds in two stages. First, I show that the knowledge of the external world we get on Putnam’s account is, in Mark Johnston’s evocative phrase, “schematic and bloodless”⁷: it does not include the special kind of understanding of the external world—knowledge of what the features of that world are like in themselves—that we intuitively take perceptual experience to provide. I then argue that this lack of acquaintance with the *external* world, together with Putnam’s account of mental representation, yields a further problem: on Putnam’s picture, we don’t even know what we *take* the world to be like; we lack acquaintance with the contents of our own minds. This is the PROBLEM OF INTERNAL ACQUAINTANCE.

1.2.1. *Acquaintance with the External World*

As noted above, Putnam’s causal-correlation account of mental representation offers a kind of guarantee against traditional skeptical doubts. But that guarantee has struck many as unsatisfying.⁸ The reason is that there is a kind of epistemic contact with the external world—*acquaintance* with its perceptible features, knowledge of what those features are like in themselves—that we naturally take our perceptual experience to provide. This special epistemic contact—not just our confidence in the truth of our beliefs—is threatened by the possibility of skeptical scenarios. And Putnam’s anti-skeptical argument does nothing to counteract this second threat.

The notion of *acquaintance* is notoriously hard to pin down. Acquaintance is supposed to be a particular kind of epistemic or cognitive relation that a subject can bear to an object or property, and it is supposed to be distinct from, at least, two other epistemic phenomena: propositional knowledge, and knowledge by description. There are various ways that the phenomenon of acquaintance can be characterized:

⁷ Johnston (1996a, p. 190).

⁸ See, for example, McDowell (1986), Nagel (1986), Johnston (1996a), Campbell (2002).

acquaintance is (1) knowledge *which* (as opposed to knowledge *that*); (2) knowledge of an object or property *itself* (as opposed to knowledge of truths *about* that object or property); (3) knowledge of the *nature* or *essence* of an object or property; (4) knowledge of what an object or property *is like* (in itself).⁹ Perhaps none of these formulations fully captures what acquaintance is supposed to amount to¹⁰; in what follows, I will try to elucidate the notion by way of example.

I once had the privilege of spotting a tiger in its natural habitat. Looking at a patch of brownish-green underbrush, I suddenly became aware of a brilliant wash of color: not just the characteristic orange, but also a shockingly radiant white, in the shape of a massive pair of paws, sticking out beneath a tangle of vines. My experience informed me that there was a tiger present, and that it had a certain set of properties – the size, shape, and colors that allowed me to identify it. But my experience seemed to do more than merely represent (veridically) the presence of these properties. I came to know (or so it seemed) what tiger-orange and tiger-white are *like*; the *nature* of those properties seemed to be revealed.

Not every form of representation provides this kind of epistemic contact with the features it represents. Prior to spotting the tiger, I heard a distinctive alarm call from a rhesus macaque in a nearby tree. I had been informed (by a reliable source) that this call was a definitive sign of the presence of a tiger.¹¹ But the call did not itself put me in touch with the sensible qualities of the tiger, even though I knew that it was a sure indication of them (even though I could treat the call as a reliable *representation* of the environment as containing tiger-features). I had heard the call several times already, and, each time, I knew that tiger-features were present. If we imagine for a moment that the guide who told me of the macaque’s reliability was God, we can

⁹ On the notion of acquaintance, see Campbell (2011), Johnston (1996a), and Evans (1982); the *locus classicus* is of course Russell (1912).

¹⁰ Formulation (1), in particular, might seem a misleading characterization, since “knowledge which,” on one reading, fails to distinguish acquaintance from knowledge by description. Knowing a uniquely-identifying description of an object is one way to have what Evans (1982) calls “discriminating knowledge” of that object, and is thus (in a sense) one way to know which object is in question. But another reading of “know which” does seem to get at the right idea. I might wonder “Which color is S’s favorite?” even though I can get onto that color via a uniquely-identifying description (namely, “the color S likes the most”). So, in a second sense of “know which,” I don’t know in this case *which color* is in question, even though I have knowledge of it by description. What I lack is *acquaintance* with the color in question, which might be provided if someone informs me that *that*—pointing to a red object—is S’s favorite color. The difference between these two senses of “knowing which”—knowing a uniquely identifying description of a feature, versus having acquaintance with that feature—will play a central role in the discussion to follow (see §1.5 and §1.7).

¹¹ Not all alarm calls are so reliable: the various deer species in the tiger’s domain also give alarm calls, but they are prone to “false positives” (depending on the species, a deer call might be only a 75- or 90-percent reliable indication of a tiger). The macaque, on the other hand, gives its alarm call *only* when there is actually a tiger present.

even suppose that, when I heard the macaque call and formed the belief that tiger-features were present, I had a guarantee of correspondence between my belief and the actual layout of the environment. But, without perceptual experience of the tiger, I didn't yet know its colors themselves.

What this story illustrates is that most forms of representation—representation via macaque alarm calls, for example—do not acquaint us with the features represented. But perceptual experience seems to be different; that is why *seeing* the tiger gave me access to the nature of the tiger's colors in a way that the macaque call did not.

The reason that most forms of representation fail to supply acquaintance with the features of the environment they represent is that a single kind of state or item can, in most cases, be used to represent distinct features that are intrinsically very different. This is particularly apparent in the case of *linguistic* representation. Recall the discussion of the orthographic string BOOT from the previous section. In the representational system of the English language, that string represents one sort of entity: a sturdy shoe. In the representational system of the German language, it represents a very different sort of entity: a seafaring vehicle. The two sorts of entities represented—boots and boats—are not at all alike, in terms of their intrinsic natures. This shows that there is an impressive degree of *flexibility* in the representational powers of an orthographic string like BOOT. But, precisely because of this flexibility, the string itself can't be taken to reveal the nature of whichever entity it in fact represents in a given instance. Representation by means of the string BOOT can't be enough to know what the environment thereby represented *is like*, because nothing about BOOT itself distinguishes between the natures of entities as different as boots and boats.

The claim that perceptual experience provides acquaintance with the features of the external world is the claim that *perceptual* representation does not function like linguistic representation, in this respect. And it is plausible that some forms of representation do provide a degree of acquaintance with features of the environment. Consider *photographic* representation. If I had seen a photograph of the tiger, rather than simply hearing the macaque call (or being presented with a linguistic tiger-representation, like the orthographic string बाघ), I would not have been totally in the dark about the nature of the (visual) tiger-features in my environment. Seeing a photograph of the tiger, and taking it to be a veridical representation of my environment, I would have known what those features were like. I might perhaps have wondered whether the photograph captured the colors *fully*, or whether they were more brilliant *in vivo*. But the representation itself, unlike a macaque-call or linguistic representation, would acquaint me with tiger-orange at least enough to distinguish it from something with a very different (visual) nature (like, say, elephant-gray). Photographic representation, at least on one plausible account, lacks the *flexibility* of linguistic representation; a given photograph can't be used to represent two different

features with two very different (visual) natures.¹² As a result, it can supply some degree of acquaintance with the single feature it is capable of representing.

The key question now is whether *perceptual* representation can supply this kind of acquaintance. As the difference between my macaque-call representation of the tiger-features and my later experiential representation of those features illustrates, perceptual experience *seems* to be what makes the difference between having acquaintance with the (perceptible) features of the external world, and lacking it.¹³ Thus, if perceptual representation turns out not, in fact, to be the kind of representation that can supply acquaintance, we will be incapable of knowing what the external world is like (even if we know that we are veridically representing it).

On Putnam's account of mental representation, this seems to be just the situation we are in. As explained above, the key distinction between linguistic representation and photographic representation—the distinction that explains why the former cannot supply acquaintance, while the latter can—is that, unlike photographic representation, linguistic representation allows for an enormous amount of *flexibility*. A given type of linguistic-representational item—a given orthographic string—can represent *any* kind of entity; that is why representation via BOOT cannot itself reveal the nature of the entity represented. And the entire point of Putnam's argument against the traditional view's "magical" theory of mental representation—including perceptual representation—has just this kind of flexibility. A course of experience like mine *could* represent the presence of tiger-features—the colors and shape of a material body. But it could also—if it occurred in a brain in a vat—represent the presence of vat-computer programming elements. The natures of those two types of features are wildly different, so representation via perceptual experience—which does not itself distinguish one from the other—could not reveal the nature of the external world thereby represented. Since perceptual experience is the only plausible route we have to acquaintance with the features of the external world (witness, again, the difference between seeing the tiger and representing it via macaque-call), we therefore lack acquaintance with the external world altogether.

Thus, the very feature of Putnam's account of mental representation that allows

¹² An important disclaimer: I do not mean to endorse a particular theory of photographic representation (or of the nature of depiction more generally) here. I am simply sketching one possible view about photographic representation, in order to have a useful model for thinking about perceptual representation. The idea is just that it seems *conceivable* that some forms of representation will not exhibit the kind of representational flexibility that characterizes linguistic representation. The key question I will address is whether *perceptual* representation in fact functions in this way, whether or not we accept that *photographic* representation does.

¹³ This is a point emphasized by Johnston (1996, p. 188): "Perception represents itself as (or is at least spontaneously taken by its possessors as) a mode of access to the perceptible natures of things; a mode of acquaintance with their perceptible properties." Campbell (2002) makes a similar claim.

him to offer a guarantee against traditional skeptical doubts—the REPRESENTATIONAL FLEXIBILITY of perceptual experience, the way it can veridically represent any number of different features—is what ensures that perceptual experience, on Putnam’s picture, cannot supply acquaintance with the external world.

Many commentators on Putnam-style causal-correlation accounts of mental representation have noted this much.¹⁴ One common response to this worry has been something like a shrug of the shoulders. According to defenders of causal-correlation accounts, it is simply *true* that perceptual experience does not give us acquaintance with the external world; perception does not itself reveal the natures of the features in our environment. Demanding that experience provide acquaintance in this way is simply demanding too much.¹⁵

Below, I will express some sympathy with this line of thought: I do not, in the end, think that seeing the tiger gave me acquaintance with its colors. So I should apologize for pulling a bit of a bait-and-switch. The tiger story was meant to draw out an element of the intuitive dissatisfaction many feel when confronted with Putnam’s response to external-world skepticism. But I do not fully endorse the intuition on which the story turns. My own view is unlikely to satisfy those who think that the epistemic contact we achieve with the external world through perception is deeper than Putnam allows. On my own view, perception does not, in itself, grant us acquaintance with the features of the external world.

But I now want to emphasize what I see as a much more serious shortcoming of Putnam’s picture, which will be the real target of my criticism, and the problem I will address at length below. Our lack of acquaintance with the *external* world, combined with Putnam’s account of the nature of representation, leads to the conclusion that we are also not acquainted with the contents of our own *minds*. This is the PROBLEM OF INTERNAL ACQUAINTANCE.

1.2.2. *Acquaintance with Our Own Minds*

Here is the argument that Putnam’s account results in the PROBLEM OF INTERNAL ACQUAINTANCE (I elaborate on the argument below):

- (1) We are not acquainted with the features of the external world, including those that are represented in perceptual experience.
- (2) Our perceptual experiences are (by and large) veridical.
- (3) If a set of representations is (by and large) veridical *and* we are acquainted with the contents of that set of representations, then we are (thereby) acquainted with the features of the external world that are represented.

¹⁴ See Nagel (1986), Johnston (1996a), Langton (1998), Campbell (2002).

¹⁵ See Lewis (2009), Gibbard (1996), Sosa (1996), Chalmers (2010).

- (4) So we are not acquainted with the contents of our perceptual representations.

The argument's first two premises follow from Putnam's causal-correlation account of mental representation. Premise (1) is what I have just finished establishing: on Putnam's account, we lack acquaintance with the external world. Premise (2) is what Putnam himself emphasizes about his view. Since (Putnam claims) what a mental state represents depends on what a subject actually interacts with, mental states will necessarily be (by and large) veridical—they will represent their normal causes. This is simply a statement of the conclusion of Putnam's own anti-skeptical argument.

Premise (3) is what ties these two implications of Putnam's view together and delivers the result that we lack internal acquaintance. The idea is that, given a guarantee of veridicality, we could carry over any acquaintance we might have with the contents of a representation to establish acquaintance with the features of our external environment. The notion of *acquaintance with a representational content* that I have in mind here is the following:

A subject *S* is acquainted with the content of a representation *R* iff *S* knows what the features of her environment *would* be like, were *R* to turn out to be veridical.

Importantly, this kind of acquaintance with a representational content is not *itself* sufficient for acquaintance with the features of the *environment*. This is because representations can be non-veridical.

Recall the example of photographic representation. A photograph of a tiger, I suggested, can provide acquaintance with the visual tiger-features instantiated in the environment, in a way that a macaque call cannot. But that is only true if we assume that the photograph is a *veridical* representation. If the orange, tiger-shaped marks on the photograph I am looking at were generated by an ant walking across some photographic plates in a darkroom, and the environment around me in fact contains no tigers, then that photograph does *not* acquaint me with the features of my environment, since it *misrepresents* that environment.¹⁶ One cannot have acquaintance with

¹⁶ A complication here: It's not clear that the object just described—an image created by the movements of an ant on a photographic plate—is really a *photograph* at all. Plausibly, for something to be a photograph, it must be generated through a particular process—one involving light entering the lens of a camera and causing certain effects on a roll of film or some digital medium. On this picture, it is unclear whether photographs can ever *misrepresent* the environment's visual features, since an object is only a photograph of *x* if *x* played the appropriate role in generating it. But we can consider a somewhat wider class of objects—the set of images on photographic paper, including both genuine photographs and images created through non-standard means—as a class of representational items, whose members *are* capable of misrepresentation. This class of items, treated as a set of representa-

the external world's tiger-orange if the external world doesn't actually instantiate tiger-orange.¹⁷

When veridicality is assumed, however, we *can* move from acquaintance with the contents of the photographic representation to acquaintance with the features of the environment. Seeing the photograph allows me to know what the visual features of the environment are like *if* they are the way the photograph represents them as being; the veridicality of the representation then ensures that the environment is indeed that very way. These two pieces together are what allow for acquaintance with the features of the environment.

The reason this presents a problem for Putnam is that, in the case of perceptual representation, he has given us the second piece—a guarantee of veridicality—but, as we have already seen (in what was argued above), we do not in fact achieve acquaintance with the features of our environment. And so it must be the *first* piece—acquaintance with the contents of our own mental states—that is missing. The failure of acquaintance with the external world we get on Putnam's view stems not from any barrier between our mental states and the features of that world, but rather from a barrier between us and our own mental states.¹⁸

Putnam thus leaves us not only out of touch with the nature of the external world, but alienated from the contents of our own minds. On Putnam's account, we don't even know what our experience *presents* the external world to be like. This is a far deeper kind of alienation than traditional Cartesian skepticism generates. Descartes, even in the midst of his skeptical worries (before God has been summoned to

tions of the environment, has the features relevant here: a given member of the set is a representation of the environment, which is not flexible in terms of what it represents the environment to be like, but it might nonetheless misrepresent that environment, if it is generated in a non-standard way.

¹⁷ Johnston (1996b) emphasizes that the idea of acquaintance with a feature of the external world is more than just the idea of knowing what that feature *would* be like, were it to be instantiated; i.e., it is different from what I have called “acquaintance with a *content*.” As Johnston puts it, “knowledge that could be had whether or not a thing exists is not knowledge by acquaintance with that thing” (p. 223). Acquaintance with a *content*, in my sense, is compatible with the non-existence of the external-world feature represented; so it clearly does not count as “acquaintance with a thing,” in Johnston's sense.

¹⁸ My objection to Putnam here is structurally quite similar to that presented by McKinsey (1999), who argues that, on Putnam's externalism, we lack privileged access to our *thought* contents. McKinsey's argument, like mine, begins by noting that there is some epistemic status that we seem to lack (for McKinsey, *a priori* knowledge that, e.g., the world contains water; for me, acquaintance with the perceptible features of the external world); the argument then notes that, on Putnam's view, there is a route from some piece of knowledge of our own minds to knowledge of the external world (since Putnam gives us a guarantee of correspondence between mind and world); and it then concludes that we must not have the knowledge of our own minds in question, given that we lack the external-world knowledge of the relevant kind.

return the world to him) thinks he can know *how things appear to him to be*. At that stage, Descartes indeed doubts that he is acquainted with the external world—he denies that he knows the natures of the features causing his experiences, since he cannot tell whether those features are the shapes and colors of material objects, or, instead, the whims and fancies of an evil demon—but he is certain, he says, that he knows which features he *takes* the world to instantiate.¹⁹ He is acquainted with the contents of his own mind. Putnam’s picture robs us of this internal acquaintance. That is why his account of perceptual representation must be rejected.

1.3. SEPARATISM

In the previous section, I argued that, on a causal-correlation account of mental content like Putnam’s, we don’t even know what we *take* the world to be like. In this section, I offer a diagnosis: The aspect of Putnam’s account that leads to the PROBLEM OF INTERNAL ACQUAINTANCE is a doctrine that has come to be known as *separatism*.²⁰ It is the view that two central features of a perceptual state are intrinsically separable: on the one hand, *phenomenology*, the qualitative character of a perceptual state (or “what it’s like” to be in that state); and, on the other, *intentional content*, what a given state represents.

Putnam’s separatism results from a combination of two theses about what determines these features. The first thesis (discussed in §1.1) is *content externalism*, according to which the representational content of a subject’s perceptual states depends on the features of the external environment that typically cause those states. The second thesis (on which Putnam places much less emphasis) is *phenomenal internalism*, according to which a subject’s phenomenology depends solely on the constitution of the subject’s central nervous system, and not on any features of the external environment.

1.3.1. *The Diagnosis*

Putnam’s explicit concern in “Brains in a Vat” is to elaborate and defend content externalism, by arguing that the external environment determines the content of propositional attitudes like beliefs. But Putnam does mention perceptual experiences

¹⁹ This is an oversimplification of Descartes’s view. For Descartes, we do not actually have this kind of internal acquaintance with the contents of *all* of our perceptions. Importantly, Descartes takes our grip on the contents of our perceptions of colors and other *secondary* qualities to be irredeemably obscure and confused. Still, he thinks that our grip on our perceptual contents as a whole is firm enough for us to know, for example, that we do not represent the world as merely a set of ideas in the mind of an evil demon. Below, I discuss at length the idea that we are acquainted with only *some* of our perceptual contents (see §1.6 and §1.7).

²⁰ The term is due to Horgan and Tienson (2002).

in addition to beliefs, and he implicitly endorses phenomenal internalism when describing an imagined world inhabited by envatted brains:

The humans in that possible world have exactly the same experiences that we do.... Their images, words, etc., are qualitatively identical with images, words, etc., which do represent trees in *our* world.... Just as a splash of paint might resemble a tree picture without *being* a tree picture, so... a 'sense datum' might be qualitatively identical with an 'image of a tree' without being an image of a tree.... [W]e see that the *qualitative similarity* (amounting, if you like, to qualitative identity) between the thoughts of the brains in a vat and the thoughts of someone in the actual world by no means implies sameness of reference.²¹

Putnam describes the envatted brains' "experiences," "sense data," and "images" as being "qualitatively identical" to our own. That is, he claims that their *phenomenology* is the same as ours. And, presumably, the reason he makes this claim is that we and the envatted brains have the same internal constitution. The implicit premise here is phenomenal internalism: internal constitution fixes phenomenology; brains in a vat share our internal constitution; therefore, brains in a vat share our phenomenology.

Putnam's likely motivation for maintaining this traditional internalist conception of phenomenology is that doing so allows him to respect the Cartesian idea that we have specially-secure knowledge of an element of our own subjectivity. Putnam argues for content externalism as a response to the MAGIC WORRY that arises on the internalist picture. But, precisely because it is an externalist thesis, Putnam's account of mental representation threatens to undermine the genuine Cartesian insight that we have privileged access to our own minds through introspection. For, on Putnam's picture, the facts that determine the contents of a subject's mental states—facts about the causal correlations between those states and features of her environment—are accessible as easily by external observers as by the subject herself. In fact, the lesson of the brain in a vat, for Putnam, is that external observers can know such content-determining facts *more* easily than can the subject. On Putnam-style content-externalist views, it is from the *external point of view* (from the *standpoint of the interpreter*, or via the *third-person perspective*) that one comes to know what a subject's mental states represent. The facts that determine the contents of a subject's mental states are therefore not ones that the subject has any special access to; so the contents of her mental states can hardly be knowable by the subject in a privileged, introspective way.²² To put it another way: the whole point of Putnam's argument is that the

²¹ Putnam (1981, p. 10).

²² The point here is *not* that introspection must be a route to knowledge of some in principle "private" set of facts about a given subject's mind, facts which are unknowable for any outside observer. Rather, the point is that for any set of facts to count as the object of introspection, those facts must in general be of such a kind that the subject has some means

difference between the mental states of the normal subject and those of the envatted brain—the difference in the contents of those states—is not introspectively accessible to the subjects themselves.²³

So Putnam’s view is one on which we do not have privileged, introspective access to the *contents* of our mental states. But, by maintaining the traditional view that *phenomenology* is determined solely by a subject’s internal constitution, Putnam sets aside a corner of our minds to serve as the object of introspection. Being determined by a subject’s own constitution, phenomenology is, plausibly, knowable by that subject in a way that it is not knowable by others. And so there is reason to think that we do maintain a kind of privileged access on Putnam’s account.

When we combine this phenomenal internalism with the content externalism that is Putnam’s main focus, we get separatism. On Putnam’s picture, there are two features of our perceptual states—their content and their phenomenology—that are determined by distinct phenomena. And, thus, those two features can come apart: the external environment can vary independently of the subject’s brain state (this is the lesson of the brain in a vat), so intentional content (depending as it does on the former) can vary independently of phenomenology (depending as *it* does on the latter). This is what leads to the REPRESENTATIONAL FLEXIBILITY that characterizes Putnam’s account of mental representation. The phenomenology of a given perceptual state, being internally constituted, has no built-in connection to any particular external-world features, and so that phenomenology can come to represent any number of different features. So even though we do have introspective access to our *phenomenology* on Putnam’s picture, that access does not give us insight into the *contents* of our perceptual states.

Thus the PROBLEM OF INTERNAL ACQUAINTANCE can be traced to separatism: separatism leads to REPRESENTATIONAL FLEXIBILITY, which leads to the loss of internal acquaintance. In order to avoid the PROBLEM, then, we need to reject separatism; we need to find a way of *linking* the phenomenology of a perceptual state with its representational content, thereby removing the flexibility of the relation between the two.

of access to them that others lack, which (at least sometimes) gives her a special way of knowing those facts. Facts about content, on a picture like Putnam’s, fail to meet this condition, and so they can’t be taken to be the objects of introspection. The discussion here echoes a point emphasized by McDowell (1986, p. 159).

²³ Putnam himself explicitly endorses the claim that introspection provides no insight into the *contents* of mental states (see pp. 17-18). In a similar vein, McKinsey (1991) argues that privileged access to mental content is incompatible with Putnam’s externalism about the mind. Burge (1988) and McLaughlin and Tye (1998), by contrast, argue that content externalism *is* compatible with our having privileged, introspective access to the contents of our own thoughts. I won’t delve into these debates here, but it seems to me that Putnam is much more clear-sighted than some of his followers in understanding the limits content externalism imposes on our introspective access to our own minds.

1.3.2. *Phenomenal Externalism*

One way to secure such a link would be to follow the externalist turn a step further: we could take phenomenology, as well as content, to be *externally* determined. On this kind of phenomenal externalist view—advocated, in very different forms, by contemporary *wide representationalists* and *naïve realists* (or *disjunctivists*)—the features of the environment with which a subject interacts determine not only what that subject’s perceptual states represent, but also what it’s like for the subject to be in those states.²⁴ Phenomenology and content cannot, on this type of view, vary independently, since they are determined by the very same set of features. There will thus be no REPRESENTATIONAL FLEXIBILITY when it comes to our perceptual states, and so the threat of losing internal acquaintance will have been defused.

If we extend Putnam’s externalism to include phenomenology, however, there will be *no* corner of the mind left to serve as the object of our specially-secure introspective knowledge. On the phenomenal externalist picture, both content and phenomenology are determined by factors external to the subject herself. Thus, both facts about *what the subject’s mental states represent* and facts about *what it is like to be in those states* are accessible as easily by external observers as by the subject herself. There is now *nothing* that a subject can know about her own mind in a way that others cannot; thus, the essential Cartesian insight, and the commonsense idea that we have special access to our own minds, has been lost.²⁵

The PROBLEM OF INTERNAL ACQUAINTANCE concerns the access we have to our

²⁴ Strictly speaking, for *naïve realists*, perceptual states do not *have* any representational contents at all. On this view, external-world features determine what it’s like to be in a particular perceptual state not in virtue of being the representational contents of those states, but in virtue of being *constituents* of those states. Still, mental states that *do* have representational content—propositional attitudes like beliefs and desires—will, on this kind of picture, derive their contents from experience. So there will still be an “inflexible” link between the phenomenology of perceptual experiences and the contents of, e.g., perceptual *judgments*.

²⁵ Phenomenal externalists sometimes argue that there *is* still an intelligible sense in which a subject has privileged access to her own mind on their picture (the facts of subjectivity so accessible will, for many phenomenal internalists, be essentially *disjunctive* (see McDowell (1986)); at other times, they seem to *applaud* the rejection of the doctrine of privileged access their view entails (such a response is *also* suggested in McDowell (1986)). I do not find either of these responses particularly compelling; it seems clear that it is indeed a major cost of phenomenal externalism that it makes our access to facts about our own subjectivity less secure than we might have thought (this point is acknowledged by Mike Martin in his (2004), where he notes that, on the disjunctivist version of phenomenal externalism, it is in cases of *mis*perception that our lack of access to our own minds is starkest). But this is not the place for a full treatment of the merits of phenomenal externalism (which I address in depth in Chapter 4). My purpose here is only to point out that it is *prima facie* plausible that phenomenal externalism cannot do full justice to the Cartesian insight that we have privileged, introspective access to our own minds.

own mental lives. The PROBLEM is generated by Putnam's separatist picture, on which our mental states' phenomenology and their intentional content can vary independently. But if we attempt to resolve the PROBLEM by making even *more* of our mental lives externally determined, we risk a *complete* loss of the specially-secure grip on our mental states that we take ourselves to have. And so the phenomenal externalist's proposed solution to the PROBLEM OF INTERNAL ACQUAINTANCE looks like no solution at all.

It thus seems worth exploring a different alternative to Putnam's separatism: we can avoid the REPRESENTATIONAL FLEXIBILITY that leads to the loss of internal acquaintance by taking both phenomenology and content to be *internally* determined. That is the only way to vindicate the claim that we have privileged access to our own minds; it is the only real way to solve the PROBLEM OF INTERNAL ACQUAINTANCE.

1.4. THE MAGIC WORRY, AGAIN

There has been a recent surge of enthusiasm within the philosophy of mind for just such a return to a more traditional, internalist account of perception; this is the so-called *phenomenal intentionality research program* (or "phenomenal intentionalism").²⁶ On this view, the phenomenology of a given perceptual state is determined by a subject's internal constitution; and the *content* of that state is, in turn, determined by its phenomenology. So content is *also* internally determined. Thus phenomenology and content are *linked*; they cannot vary independently of each other. Phenomenal intentionalism removes the REPRESENTATIONAL FLEXIBILITY that leads to the loss of internal acquaintance on the separatist picture, without threatening the *total* loss of privileged access phenomenal externalism seems to engender.

This approach seems to me to be the right way to respond to the PROBLEM OF INTERNAL ACQUAINTANCE: we must allow that mental content is determined by features internal to the subject, in order to make room for the idea that we know the contents of our own minds. But phenomenal intentionalism, in returning to the traditional internalist picture, opens itself up to the objection that originally motivated the externalist turn. According to the phenomenal intentionalist, phenomenology is fixed by internal states, which can be as they are with or without causal connections to any given external-world features (again, the brain in a vat). Phenomenology, in turn, determines which particular features a given state represents. So a perceptual state can, on this view, represent particular external-world features, without having any causal connection to those features. Thus phenomenal intentionalism is precisely the kind of internalist account that was the target of Putnam's argument against "magical" theories of representation. How, Putnam would demand to know, could the internally-constituted phenomenology of a perceptual state *itself* create a representational connection to a particular external-world feature?

²⁶ See Kriegel (2012) for an overview of the phenomenal intentionality research program.

This is, I think, *the* crucial challenge for the phenomenal intentionalist and, more generally, for any internalist account of perceptual experience. One could acknowledge that internalism has strong intuitive appeal, and that it seems to hold out the best hope of vindicating our commonsense conception of the special, introspective access we have to our own minds; but one might nonetheless think that such a view should be rejected because it could never make sense of the way that our mental states seem to present us with a world made up of particular, mind-independent features. Almost all views that reject an internalist conception of perceptual experience are driven by the need to make sense of the world-directedness of our minds, combined with the conviction that no internalist theory can do so. The internalist thus stands in need of a response to the MAGIC WORRY. The remainder of this chapter will be devoted to developing that response.

First, though, I want to tease apart two distinct challenges the MAGIC WORRY presents. There are *two* key features of perceptual content that, the externalist alleges, could only be explained by magic if we accept internalism: our perceptual states are WORLD-DIRECTED (they represent features of a *mind-independent* world); and they are DETERMINATE (they represent *particular* features as being instantiated in that world).²⁷

1.4.1. *World-Directedness*

Our experience of the world does not seem to inform us only about our own minds; it seems to provide us with a point of view on a world that is independent of our

²⁷ There is also a third kind of challenge that some might take to be contained in the MAGIC WORRY: One might wonder how an internally-constituted mental state could come to represent *anything* at all. That is, the MAGIC WORRY might be seen as a way of framing a more fundamental “how possible” question, which asks for an explanation of how representational thought—determinate or not, world-directed or not—is possible on the internalist’s picture. This, I want to emphasize, is *not* the worry I am addressing here; but I also want to emphasize that it is likewise not the worry most commonly pressed against internalism by the externalist. Burge (2009), for example, accepts that some kinds of mental content—including “representation of mathematical and mental matters”—are not constitutively dependent on the subject’s external environment in the way that *empirical* contents are. McDowell (1986) also limits his externalist claims to “object-directed intentionality,” and he does not attempt to give an explanatory account of representation as a whole. So externalists, for the most part, are willing to allow that the internalist can account for some (non-empirical) kinds of thought; or, at least, that the advantage of externalism is its ability to make sense of *empirical* content, not to give an answer to the fully general question about the possibility of representational thought. The worry I am addressing here—the worry that Burge and McDowell explicitly press—is limited in scope: it is a worry about how internally-constituted mental states—the kind that, according to Burge, can account for representation of mathematical and mental matters—could have the specific *kind* of content our perceptual states seem to have: content that includes determinate representations of the features of a mind-independent world.

awareness. The challenge the externalist poses is that it seems quite mysterious how our perceptual states, conceived of as the internalist conceives them, could possibly come to have the WORLD-DIRECTEDNESS that they manifestly do have.

This challenge has been presented most forcefully by John Campbell, who argues that the impossibility of explaining WORLD-DIRECTEDNESS on an internalist conception of experience is what drives Berkeley to idealism. Berkeley's key insight, according to Campbell, is that

you can't ground thought about mind-independent objects in the qualitative character of perceptual experience if the qualitative character of perceptual experience is thought of in 'internalist' terms; that is, if we assume that the qualitative character of experience could be as it is no matter what's in one's environment.²⁸

Berkeley himself *accepts* the internalist conception of perceptual experience; he concludes that we simply do not *have* any thought about mind-independent objects. This is his idealism.

Berkeley's idealist conclusion is likely to strike us as simply incredible: it seems all too clear that we do indeed have mental states that represent a mind-independent world. It is the need to explain *how* we can have such states—*how* we can have a conception of mind-independent features—that, by Campbell's lights, constitutes “the key motivation for current disjunctivist or naïve realist views of experience, which criticize the conception of conscious experience as something that is merely an effect of external objects.”

Campbell, himself a disjunctivist, cites two other disjunctivists—Bill Child and John McDowell—who also press this challenge against the internalist. As Child puts it, to accept phenomenal internalism:

is to think of experiences as states of a type whose intrinsic mental features are world-independent; an intrinsic, or basic characterisation of a state of awareness will make no reference to anything external to the subject. But if that is what experience is like, the disjunctivist objects, how can it yield knowledge of an objective world beyond experience, and how can it so much as put us in a position to think about such a world?²⁹

In a similar vein, McDowell writes that the phenomenal internalist's picture is one on which:

²⁸ Campbell and Cassam (2015, p. 16).

²⁹ Child (1994, p. 148).

subjectivity is confined to a tract of reality whose layout would be exactly as it is however things stood outside it, and the commonsense notion of a vantage point on the external world is now fundamentally problematic.... Once we are gripped by the idea of a self-contained subjective realm, in which things are as they are independently of external reality (if any)... our picture seems to represent us as out of touch with the world altogether.³⁰

So views that reject internalism are motivated, in large part, by the worry that no internally-constituted state could possibly put us in a position to represent a mind-independent world. Since we *are* in such a position—we manifestly *do* have perceptual states that are world directed—the externalist’s charge is that phenomenal intentionalism has an explanatory obligation that it cannot meet.

1.4.2. *Determinateness*

The second challenge to phenomenal intentionalism turns on the DETERMINATENESS of our perceptual contents. A perception of, say, a boot seems to represent the presence of a certain set of perceptible features—those of a sturdy shoe—*rather than* any other set of features—including, for example, the features of a seafaring vessel. A desire to capture this DETERMINATENESS is precisely what motivates the phenomenal intentionalist to reject Putnam’s separatism.³¹ On Putnam’s account, a given perceptual state, just like a given orthographic string, has REPRESENTATIONAL FLEXIBILITY—it lacks DETERMINATENESS. That is what leads to the PROBLEM OF INTERNAL ACQUAINTANCE. In insisting that we *do* have internal acquaintance, the phenomenal intentionalist is simply acknowledging that our perceptual states are, *contra* Putnam, DETERMINATE.

The challenge that the externalist presents to the phenomenal intentionalist is that, though she may be motivated by the idea that our perceptual states have DETERMINATE contents, her view cannot explain how this is possible. Recall Putnam’s claim about *non-mental* representation: “No physical object can, in itself, refer to one thing *rather than* to another” (my emphasis). We saw this point vividly illustrated in the case of the orthographic string BOOT. But now consider a *mental* representation—say, a perceptual experience of a boot. The externalist’s claim is that such an experience could not intrinsically represent a boot, rather than a boat, just in virtue of having the internally-determined phenomenology it does, any more than the orthographic string BOOT could do so, just in virtue of having the loops and lines it does. Since the phenomenal intentionalist acknowledges that our perceptual states do represent particular, DETERMINATE features—boots *rather than* boats—and she

³⁰ McDowell (1986, p. 151).

³¹ This motivation is quite explicit in, e.g., Horgan and Tienson (2002).

argues that the representational content of those states is determined solely by their intrinsic phenomenology, she must (the externalist charges) be attributing magical powers of representation to our perceptual phenomenology.

So there are two challenges the phenomenal intentionalist must meet, in order to alleviate the MAGIC WORRY. First, she must explain how an internally-constituted perceptual state could, even though it is constitutively independent of anything outside itself, represent a *mind-independent* world. Second, she must explain how it is possible that perceptual states can *determinately* represent one kind of entity *rather than* another, in virtue of their intrinsic features, when non-mental representational items, like the orthographic string BOOT, cannot. In each case, the externalist's charge is that, once we accept the internalist's conception of perceptual experience, only a magical theory of representation could explain how our perceptual states have the features they do.

1.5. INTERNALIST HALF-RESPONSES

In the remainder of this chapter, I will show how an internalist can respond to the two challenges contained within the MAGIC WORRY. Before I get to my own response, though, I will first need to consider two other internalist strategies. Each strategy is targeted at just one of the two challenges; neither can resolve both. Exploring these strategies, and their failures, will help point the way to a more satisfying internalist view.

1.5.1. *Determinateness Without World-Directedness*

Our perceptual states have DETERMINATE contents: a given perceptual state represents one set of features, *rather than* another. Part of the MAGIC WORRY is that this determinateness looks mysterious, if we conceive of experiences as internally-constituted “episodes in a subject’s mental biography.”³² If a given perceptual state were, as the externalist holds, partially constituted or determined by a particular external world feature, then it would be intelligible how that state could determinately represent that feature. But how could a perceptual experience, just in virtue of having a certain internally-determined phenomenological character, represent any one feature, rather than another? The determinateness of our perceptual states requires that they be *inflexible*. But what, we might wonder, is the *connection* between internally-determined experiences and the properties they represent that prevents those experiences from having the REPRESENTATIONAL FLEXIBILITY characteristic of other representational items, like orthographic strings?

Putnam claims that no representational item can intrinsically represent one partic-

³² See Evans (1980).

ular feature rather than another; no representation is inflexible. But we have actually already seen, within the domain of *non*-mental representation, that there is reason to doubt the universality of this claim. When we consider linguistic representation, it seems clear that Putnam is right: a given orthographic string has no intrinsic connection to any particular feature. But recall the discussion of photographic representation in §1.2. There, I noted that a given photograph—a picture of a tiger, say—does not have the kind of representational flexibility that an orthographic string does: such a photograph could not just as easily represent elephant-gray as tiger-orange.

The reason that photographic representation is less flexible than linguistic representation is that photographic representation requires something beyond the kinds of causal (or historical, or evolutionary, or sociolinguistic) links that characterize most forms of representation (including linguistic). On one plausible view, photographic representation requires a degree of (visual) *resemblance* between the features of the photograph and the features of the photographed, in order to count as veridical.³³ This requirement of resemblance is what rules out the representation of wildly different features (such as tiger-orange and elephant-gray) by a single representational item: a single photograph could hardly bear a strong degree of (visual) resemblance to two features with such different (visual) natures, so it also could not—given the requirement of resemblance—*represent* both features.

Analogously, then, we might try to explain the lack of flexibility characteristic of *perceptual* representation by proposing that *perceptual experiences* represent via resemblance. And Locke, of course, proposed just such an account, at least in the case of the *primary* qualities (like shape). He claimed that our “ideas” of shape properties

³³ Again, two caveats about my use of the example of photographic representation. First, it is best to understand the relevant class of representational items here as including both “genuine” photographs (where the standard photographic process results in an image on photographic paper) and “pseudo-photographs” (where some non-standard process, like an ant wandering on a photographic plate, results in an image appearing on photographic paper). This removes the complication that photographic representation, in the strict sense, may not admit of misrepresentation (see fn. 16 above). Second, I am not *endorsing* the account of photographic representation (or depiction) outlined here (see fn. 12). I am simply supposing, for the sake of argument, that photographic representation requires a degree of resemblance, in order to have a helpful model of an inflexible form of representation. It is also worth noting that I am not disputing Putnam’s claim, exemplified by the ant that traces a “caricature of Churchill,” that resemblance is not *sufficient* for (pictorial) representation. What I am suggesting is that, for certain forms of representation (like photographic representation), resemblance may be *necessary* for a representational item to count as representing some given feature; other factors (including at least a causal connection, and likely some established practice of *use* of photographs) are also necessary. Putnam does explicitly deny that resemblance is necessary for representation; but his example in support of this claim is a *linguistic* one. He does not seem to consider the possibility that, *within a restricted domain*, resemblance may indeed be *one* necessary condition on representation.

“resemble” their objects.³⁴ On such a view, we would indeed have an explanation of the DETERMINATENESS of a given perceptual representation: an experience of the type we have when we perceive square objects, for example, could *only* represent square objects—as opposed to, say, circular ones—because the representational content of the experience is determined by what the experience *resembles*, and (the thought goes) such experiences (or “ideas”) resemble squares (not circles).

Locke’s resemblance theory does, in a sense, solve the problem of explaining the DETERMINATENESS of our perceptual contents. But it makes overcoming the other half of the MAGIC WORRY—the question of how our perceptual states can represent features of a *mind-independent* world—look hopeless.

On Locke’s proposal, experiences of shape represent, DETERMINATELY, via resemblance. In order to be WORLD-DIRECTED, then—in order to represent features of a *mind-independent* material world—shape experiences would have to resemble material objects. But resemblance requires something like *sharing of properties*. A photograph can represent a worldly object via resemblance because the photograph and the object can literally instantiate the very same property: the tiger and the photograph of it can both *be* orange, for example. But shape experiences cannot share the shape properties they represent objects as having; experiences simply aren’t the kinds of things that *have* shapes. What would it be for an experience (or a Lockean “idea”) to itself *be* square, to *instantiate* the very property of squareness it represents?³⁵ There seems to be a kind of category mistake here: as Berkeley famously put it, the idea of

³⁴ There is much dispute about what Locke’s notion of “resemblance” really amounts to. According to Bennett (1971, pp. 106-107), Locke’s claim that our ideas of primary qualities, but not of secondary qualities, “resemble” their objects is just a reiteration of what Bennett labels Locke’s “Causal Thesis”—the claim that our primary-quality ideas are to be causally explained in terms of objects’ primary qualities, while our secondary-quality ideas are not to be causally explained (at least at the “most basic” level) in terms of objects’ secondary qualities. Woozley (1964) claims that, for Locke, “resemblance” just amounts to accurate representation. On this interpretation, Locke’s claim that primary quality ideas resemble their objects amounts to the claim that objects genuinely have primary qualities; correspondingly, Locke’s claim that secondary-quality ideas do *not* resemble their objects amounts to the claim that objects lack secondary qualities (Stroud (2000) offers a similar interpretation). Jacovides (1999) argues for a more literal interpretation of Locke’s resemblance thesis. On his interpretation, Locke genuinely means that our ideas of primary qualities *resemble* the primary qualities of objects; our ideas of shapes literally are shaped. Campbell (2002) similarly argues that Locke must mean “resemble” in the literal sense, if that notion is to do the work he needs it to do. Here, I will (for the sake of argument) follow Jacovides and Campbell in interpreting Locke’s resemblance thesis literally, in order to see where the resulting account—even if it is not precisely Locke’s own—would lead.

³⁵ It might be thought that ideas and material objects can share shape properties because shape properties are merely “structural”: they have to do with the formal features of an object, and so can apply to objects of any kind, whether material or mental. I discuss this kind of “structuralist” proposal at length in Chapter 3.

this kind of resemblance between a mental item and the physical object it represents makes little sense, since “the only thing an idea can resemble is another idea.”³⁶ My visual experience when I saw the tiger, for example, could not sensibly be taken to resemble a mind-independent, material object, since my mind-dependent experience (or “idea”) would not share any relevant properties with such a material object. The internally-constituted features of my experience could only represent other mind-dependent features (perhaps the features of a “tigerish” idea in *God’s* mind). Thus, if we try to account for our perceptual states’ DETERMINATENESS by appeal to resemblance, we are forced to deny their WORLD-DIRECTEDNESS: we must conclude that what our mental states determinately represent is not a *mind-independent* reality, but rather a collection of *mind-dependent* features.³⁷

1.5.2. *World-Directedness without Determinateness*

The Lockean resemblance account of perceptual representation seems to lead to the conclusion that perception could only represent features of a mind-dependent world. But even without the commitment to a resemblance theory, phenomenal internalism has been accused of making WORLD-DIRECTEDNESS inexplicable, of making it impossible to understand how we could have a conception of a mind-independent world at all. In this section, I want to consider a way of spelling out phenomenal internalism that provides a response to this half of the MAGIC WORRY.

One way to explain the WORLD-DIRECTEDNESS of our perceptual states is to show how we can, in effect, *construct* a WORLD-DIRECTED content out of the materials available on the internalist’s conception of the mind. As Campbell notes, the internalist conceives of “conscious experience as something that is merely an effect of external objects.” But this already suggests a way to form a conception of those external objects: namely *as the causes of our (internally-constituted) experiences*. Conceived of in this way, objects are not *themselves* mind-dependent (as Berkeley argues they would have to be, on Locke’s resemblance view); they are the *independent causes* of mind-dependent perceptual states. Following David Lewis, we can call this picture the RAMSEYAN VIEW.³⁸ The RAMSEYAN proposes that we understand our conception of the features represented in perceptual experiences in terms of the roles those features play in causing the experiences themselves: a given external-world feature just amounts to *whatever* feature in fact plays the right role in producing a certain sort of experience.

This move does make intelligible how internally-determined perceptual states could have WORLD-DIRECTED contents. But it does so precisely by denying that

³⁶ *The Principles of Human Knowledge*, 8.

³⁷ Berkeley, of course, reached just such a conclusion, and some contemporary philosophers of perception have been led to similar idealist positions by something like this line of reasoning (see, e.g., Foster (2000)).

³⁸ See Lewis (2009).

those contents are DETERMINATE. On the RAMSEYAN view, a perceptual experience of an object as, say, square, represents the property of squareness as the feature—*whatever* feature that might be—that typically causes our experiences of a certain type (those we call “experiences of squareness”).³⁹ And that does not amount to DETERMINATELY representing the world to be some particular way. Perceptual contents, on this account, include a kind of *placeholder*, which can (depending on external circumstances) be filled in by any number of different external-world features. The content of a perceptual experience of squareness—the content *presence of the feature that typically causes experiences of squareness*—will, for a normally-embodied subject, amount to *presence of a square object*. For a subject who is a brain in a vat, by contrast, the content of such a state will amount to *presence of a certain element in the vat-computer’s programming*. This shows that perceptual representation, on the RAMSEYAN view, has REPRESENTATIONAL FLEXIBILITY—experience does not provide a DETERMINATE conception of the external world.

The RAMSEYAN’S account is thus untenable for essentially the same reason as Putnam’s. The RAMSEYAN VIEW amounts to taking onboard Putnam’s separatist picture and then using that picture to construct a connection between mind and world that is accessible from the internal perspective. Thus the RAMSEYAN does succeed, in a way, in making the WORLD-DIRECTED content of a given perceptual state—which, on Putnam’s view, is determined by features external to the subject—something that is intelligibly knowable by the subject through introspection. The account might therefore seem to offer hope of solving the PROBLEM OF INTERNAL ACQUAINTANCE. But Putnam’s separatist picture itself is what generates the PROBLEM; internalizing that picture can be no help in escaping the difficulty embedded in it. Knowledge of our internally-constituted perceptual contents does not amount to knowing which *particular* way we take the world to be, if the contents in question are, as the RAMSEYAN proposes, not themselves DETERMINATE.

RAMSEYANISM is thus simply a way of *acquiescing* in the face of the PROBLEM OF INTERNAL ACQUAINTANCE and acknowledging that we have no grip on the contents of our own minds. Rae Langton, in articulating the RAMSEYAN view (which she attributes both to Kant and to contemporary philosophical orthodoxy), acknowledges precisely this point. As she puts it, on the RAMSEYAN picture, any representation we can have of an external world feature becomes:

the name for a something-we-know-not-what—ominously similar to a Kantian thing in itself. Our contemporary orthodoxy must concede

³⁹ Lewis himself would not want to say that we latch onto squareness solely in virtue of its role in causing *experiences* in us; he thinks we know properties as occupants of particular roles in a broader *scientific* theory. The view that squareness is picked out in virtue of its role in generating our perceptual experiences is closer to that endorsed by Chalmers (2006, forthcoming). Such a view might seem overly restrictive, given the *non-experiential* causal roles squareness plays (on this point, see Bennett (1971)). Of all this, much more later.

that it, too, is faced with a conclusion similar to Kant's—that there are intrinsic features of the world with which we can never become acquainted.⁴⁰

We have now seen one account—the Lockean resemblance view—that can explain our perceptual contents' DETERMINATENESS; and another—the RAMSEYAN VIEW—that can explain those states' WORLD-DIRECTEDNESS. But neither account can answer the full challenge of the MAGIC WORRY. The resemblance view can offer no account of WORLD-DIRECTEDNESS—indeed, it makes stark why explaining WORLD-DIRECTEDNESS looks hopeless on an internalist view—and thus it makes Berkeley's idealism seem inescapable. The RAMSEYAN VIEW cannot explain DETERMINATENESS—indeed, it is a view that is premised on the denial of DETERMINATENESS—and so it succumbs to the PROBLEM OF INTERNAL ACQUAINTANCE. The two halves of the MAGIC WORRY thus seem to reinforce each other: responding to one seems to require acquiescing on the other. In §1.7, I will develop an internalist view that can resist the MAGIC WORRY's dual attack. But, before I can do so, I will first need to draw out one last challenge for the phenomenal intentionalist.

1.6. THE LINE-DRAWING PROBLEM

The MAGIC WORRY is the central difficulty facing phenomenal intentionalism. But there is another challenge for the view, which stems from the fact that, as advocates of phenomenal intentionalism have themselves acknowledged, the content externalism argued for by Putnam and others is not a wholly misguided doctrine. *Some* aspects of mental content are indeed determined by a subject's external environment. But no clear story about how and where to draw the line between “narrow” contents (those determined solely by the subject's internal constitution) and “wide” contents (those determined at least partially by features of the external world) has been articulated. We can call this the LINE-DRAWING PROBLEM.⁴¹

I argued above that we need to reject Putnam's externalist account of mental representation in order to avoid the PROBLEM OF INTERNAL ACQUAINTANCE. But rejecting content externalism entirely would be an over-reaction, since some of the arguments Putnam and others have used to motivate the view are sound. Take Putnam's original Twin Earth case.⁴² Oscar and Twin Oscar are internal duplicates, so they will (according to the phenomenal intentionalist) have matching phenome-

⁴⁰ Langton (1998, p. 176).

⁴¹ Chalmers (2006), Horgan and Tienson (2002), and Loar (2003) offer various thoughts on where the line between narrow and wide contents should be drawn. These thoughts are sometimes presented simply as *lists* of contents, divided into “wide” and “narrow.” But it seems to me that no genuinely *explanatory* account of the narrow/wide distinction has been offered.

⁴² See Putnam (1973).

nology. Suppose we now accept the phenomenal intentionalist's claim that the *contents* of a subject's perceptual states are determined by her phenomenology. We will then have to conclude that the twins' states, given their matching phenomenology, will also have matching *contents*. So if Oscar's perceptual state when looking at a glass of water on Earth represents water (i.e., H₂O), then Twin Oscar's perceptual state when looking at a glass of twater (XYZ) on Twin Earth will *also* have to represent water. And so Twin Oscar's experience will be illusory, as it erroneously represents the liquid in front of him as an entirely *different* liquid—one with which he has never had any contact.⁴³

This seems like a wildly implausible verdict. It seems clear that Putnam is right about this case: the representational contents of the two subjects' respective perceptual states are determined not by their shared internal constitution, but by the different external-world features with which they have causally interacted. Oscar's perceptual state veridically represents water; Twin Oscar's perceptual state, equally veridically, represents *twater*.

So it seems clear that at least *some* aspects of mental content are determined by features of the external world, not by a subject's internal constitution—that is, some mental contents are *wide*.⁴⁴ The central claim of phenomenal intentionalism is that there are *narrow* contents—contents of a subject's mental states that *are* determined solely by her internal constitution. So, in order to defend phenomenal intentionalism, we will need to draw a line between these narrow contents and those—like the contents of the two Oscars' perceptions of their drinking liquids—that are wide.

One natural way to draw such a line would be to group mental contents by the kinds of *states* in which they feature. Above, I suggested that Oscar and Twin Oscar have *perceptual experiences* that represent the presence of different liquids. But we might take a more restrictive view of *perceptual* content; we might think that a “high-level” natural kind property like *being water* simply doesn't show up in the contents of perception at all.⁴⁵ On this kind of view, perception only represents the presence of basic *observational* features, like colors and shapes. And, once we restrict the range of properties that are represented in perception in this way, the Twin Earth case does not show that any *perceptual* contents are wide. Wide contents come in only when we

⁴³ Or perhaps both twins' perceptual states represent *twater* (XYZ)—in which case Twin Oscar's experience is veridical, Oscar's illusory. Or perhaps such experiences represent *schmater* (where *schmater* is some *third* liquid)—in which case *both* twins suffer illusions. The point is that any of these verdicts seems implausible, since it seems arbitrary to insist—as we must, if we claim that *all* content is narrow—that the twins' phenomenally-matching experiences must both represent one particular liquid, rather than another.

⁴⁴ This claim—that *some* aspects of mental content are wide—is one that is generally accepted by phenomenal intentionalists (see, e.g., Horgan and Tienson (2002) and Loar (2003)). It is unclear whether anyone holds that *all* mental content is narrow (though John Searle may have such a view; see his (1983)).

⁴⁵ See Siegel (2010) for an extended discussion of this question.

consider high-level properties—including natural kind properties—and those properties are only represented at the level of propositional attitudes. Oscar and Twin Oscar do have *beliefs* about the liquids they encounter, and those beliefs will have different contents because of the differing environmental circumstances. Thus belief contents are wide. So we might think that the line between narrow and wide mental contents should be drawn at the border between perception and belief.⁴⁶

The problem with this proposed solution to the LINE-DRAWING PROBLEM is that there are plausible “Twin Earth” arguments to show that some contents involving paradigm *observational* properties are wide. Consider Ned Block’s Inverted Earth thought experiment:⁴⁷ There is a planet just like Earth, except that the colors of the various objects on the planet are all “inverted” relative to those on Earth (the sky is yellow, tiger stripes are blue-green, and so on). In addition, the humans on Inverted Earth have “inverting lenses” in their eyes, so that the signals traveling from their retinas when they look at their blue-green tigers exactly match the signals traveling from our retinas when we look at our orange ones (and so forth). Thus, I and my twin on Inverted Earth (call him “Twinny”) will be in the same internal states when viewing objects of different colors on our respective planets.

Given phenomenal internalism, we can conclude that Twinny and I have matching phenomenology when, for example, I look at an orange tiger-stripe on Earth and Twinny looks at a *blue-green* one on Inverted Earth. Now we can ask what the *contents* of our respective perceptual states will be. Presumably, my experience (veridically) represents the tiger in front of me as orange. If the contents of perceptual states are narrow, we will have to conclude that *Twinny’s* perceptual state also represents orange. But Twinny’s tiger is blue-green, so he will be suffering from an illusion. Indeed, we will have to conclude that *all* of Twinny’s color experiences—and all color experiences of Inverted Earthlings in general—are illusory.

As with Twin Earth, this seems like an implausible verdict. Both Twinny and I are representing the features of the objects in our native environments in the way our species were evolved to do. There seems to be no reason to regard Earthlings’ experiences as veridical and Inverted Earthlings’ experiences as illusory, rather than vice versa.⁴⁸ And so, as with Twin Earth, we are led to the conclusion that Twinny

⁴⁶ This kind of view would actually allow that *some* contents of belief states are narrow—namely, those contents that only involve the kinds of observational concepts that feature in perceptual content. The line is to be drawn between those contents that *can* show up in perception, and those that *never* do.

⁴⁷ Block (1990).

⁴⁸ The way Block originally describes the case, the Inverted Earthlings have inverting lenses placed in their eyes at birth. This might suggest that there *is* a reason to say that their color experiences are illusory, while ours are not. If we take onboard a kind of teleological account of content, we might say that the Inverted Earthlings color-perception states have the *function* of representing the colors in the same way ours do, but that the insertion of the inverting lenses prevents those states from performing this function. But the scenario need not be

and I both have veridical perceptions.⁴⁹ But, given the different colors of the tigers we are looking at, this means that Twinny and I must be representing different colors when we are in the same brain state—i.e., the color contents of our perceptual experiences are wide.

Thus, we must allow that even some perceptual contents—some contents involving observational properties—are not fixed by a subject’s internal constitution. This leaves us still searching for a way to draw the boundary between wide and narrow contents.

The LINE-DRAWING PROBLEM is something of a technical worry, or a question of detail, for the phenomenal intentionalist. But it also bears on the central question I have been exploring: How can we explain the specially-secure access we have to our own minds, while acknowledging that what we have access *to* includes the DETERMINATE, WORLD-DIRECTED contents of our perceptual states? In §1.2 and §1.3, I argued that taking those contents to be constituted by factors external to the subject—taking them to be *wide*, as Putnam does—leaves them outside the scope of our introspective access. In acknowledging that some aspects of our world-directed mental content—even within the domain of perception—are indeed wide, then, I have suggested that some aspects of our mental lives are not accessible to introspection. In order to explain our introspective access to world-directed mental content, we will have to look to the other elements of content: those that are *narrow*.

The challenge, then, is to provide a solution to the LINE-DRAWING PROBLEM on which the contents falling on the “narrow” side of the dividing line can plausibly serve as objects of our privileged, introspective access; and, crucially, to do so in a way that makes sense of the idea that these narrow contents—even though they are internally determined—can, without relying on magic, represent particular, *determinate* features of the mind-independent world.

thought of as involving artificially inserted lenses: the “lenses” could be a perfectly natural part of the Inverted Earthlings’ physical constitution, and they could have evolved due to perfectly normal adaptive pressures. Indeed, we could suppose that *we* are actually the ones whose physiology includes something like an “inverting lens,” since the way our retinas process light has no special claim to being “more natural” than any number of other physically-possible processes (indeed, it is often noted that our own visual processing involves a kind of orientation “inversion” of the retinal image at an early stage of processing). So it seems perfectly legitimate to suppose that there could be two species that have equally natural, equally evolved systems for representing color properties that result—under ideal natural circumstances—in members of those species going into the same internal states when viewing objects of different colors.

⁴⁹ Some are instead led to the conclusion that *neither* twin’s experience is veridical, and, by universal generalization, that *all* color experience is illusory (see, e.g., Pautz (2006)). This seems like a counterintuitive verdict, however, and, as I hope to show in what follows, an unnecessary concession to make.

1.7. PRIMARY-QUALITY CONTENTS AND *A PRIORI* CONCEPTS

In this section, I will argue that the line between narrow and wide perceptual contents lies at the border between primary and secondary qualities: primary-quality contents are narrow, secondary-quality contents are wide. I will also argue that, in coming to understand how we have acquaintance with the narrow contents of our primary-quality perceptions, we can see how internally-constituted perceptual states can indeed represent determinate features of a mind-independent world. Primary qualities are properties on which we have a determinate, *a priori* grasp, and they are also ones that can characterize an objective, spatial world. *That* is how we can know which particular properties we take the world to instantiate when our perceptual experience represents the presence of a primary quality like shape.

1.7.1. *Drawing the Line*

I want to begin by looking more closely at the LINE-DRAWING PROBLEM: the question of where the boundary between wide and narrow contents falls. We have already seen that two kinds of contents—natural kind contents and color contents—fall on the “wide” side of the divide. The way I argued that these contents are wide was by constructing “Twin Earth” scenarios: scenarios in which two subjects with matching internal constitutions represent two *different* features, depending on which features they interact with. These “inversion” cases show that the content in question is wide by, in effect, showing that it is indeterminate: the contents in these scenarios are shown to have the REPRESENTATIONAL FLEXIBILITY characteristic of non-acquaintance-conferring forms of representation. As we have seen, such states do not, in themselves, represent any one determinate feature, rather than any other. The perceptual state that represents the presence of water, for example, could just as easily have represented twater.

The representational flexibility of *color* experience is somewhat more surprising, but that is precisely what the Inverted Earth case reveals. My twin on Inverted Earth, seeing a blue-green tiger (and veridically representing its color), would have been representing a property with a different nature from that of the orange *I* was seeing—blue-green and orange surely have different natures—even though he would have been in a state with the same introspectible phenomenological character as mine. So just being in an experiential state with that character cannot be enough to fix a determinate color content. Despite my initial sense that seeing the tiger conferred acquaintance with tiger-orange in a way that hearing a macaque alarm call did not, color experience turns out not to be a means of acquaintance with the colors (nor even with color *contents*). The moral of the tiger story as I originally told it—that perception affords us acquaintance with colors, while other forms of representation

do not—is revealed, by reflection on Inverted Earth, to be a lie.⁵⁰

The worry now is that it is beginning to seem as though *all* contents are wide, and that we therefore never know which particular property a given mental state represents. If an inversion scenario like Twin Earth can be constructed for a given kind of content, then we will have to acknowledge that we do not have a determinate grip on the content in question. So if relevant inversion scenarios can be constructed for *all* perceptual contents, we will have succumbed to the PROBLEM OF INTERNAL ACQUAINTANCE: we will be forced to acknowledge that our perceptual states simply do not have any determinate content at all.⁵¹ What we need, then, in order to have a response to the PROBLEM OF INTERNAL ACQUAINTANCE and a way of explaining the DETERMINATENESS of our perceptual representations, is to identify a kind of content that is not subject to inversion scenarios.

A striking feature of the philosophical history of inversion scenarios is that such scenarios have almost always been invoked in discussions of *secondary qualities*, like colors, rather than *primary qualities*, like shape.⁵² So let us consider the following proposal: *shape* contents are not subject to inversion scenarios because they are *narrow* and determinate. They (and other primary quality contents) are the locus of our internal acquaintance.

The proposal is this. Our perceptions of *shape* properties do not involve REPRESENTATIONAL FLEXIBILITY. When we perceive an object as, say, square, we do know which determinate property we take the object to instantiate: we know the *nature* of the squareness our experience attributes to the object. Such an experience could not just as easily represent an object as, say, circular: circular objects simply do not instantiate the particular, determinate property that we know our experiences of squareness to represent. That property is one with which we have *a priori* acquaintance—an acquaintance that is employed when, for example, we perform a proof of the Pythagorean Theorem utilizing our concept of squareness. We can do such proofs because we know what squareness is like in itself. And so we know what the world would have to be like for our perceptions of squareness to be veridical. We are acquainted with the contents of our shape experiences (and, in general, with the contents of our primary-quality perceptions).

That is my proposed solution to the PROBLEM OF INTERNAL ACQUAINTANCE. It

⁵⁰ Here, I complete the bait-and-switch acknowledged above (see §1.2).

⁵¹ This is one way of seeing what Putnam's argument about brains in a vat is trying to do: it is intended to be a *global inversion scenario*, and thereby to reveal that we have no acquaintance with *any* of our mental contents.

⁵² There have been a few attempts to construct "shape inversion" scenarios in the recent literature (see, e.g., Chalmers (forthcoming), Thompson (2010); as just noted, Putnam (1981) can be seen as presenting an inversion scenario for shape—along with everything else). But these scenarios are far less compelling than the corresponding color inversion cases. In Chapter 4, I discuss at length attempts to use inversion scenarios to argue that we are not acquainted with shape contents.

depends on distinguishing perceptual representation of primary qualities from perceptual representation of secondary qualities in terms of the different kind of grip we have on the natures of the properties represented. With secondary qualities, we do not know *which* particular property a given perceptual state represents. We don't know the *nature* of orange when we see a tiger (in spite of an initial tendency to suppose that we do); we are not *acquainted* with the content of our color experience. We do, however, know the nature of *shape* properties: we know *which* property squareness is—how the world would have to be in itself, for our perception to be veridical—when our experience represents an object as square.

The rest of this dissertation will be devoted to elaborating on and defending this proposal. As a preliminary step along that path, I want to show how the proposal can address the central challenge for the phenomenal intentionalist: the MAGIC WORRY, the question of how, if not by magic, an internally-constituted mental state can represent a particular, determinate feature of a mind-independent world.

1.7.2. *Primary-Quality Concepts*

It will be helpful to return briefly to the Ramseyan story because it frames the challenge before us quite sharply. The Ramseyan's central claim is that the narrow contents of our perceptual states—the ones to which we have introspective access—are all *descriptive*, hooking on to external-world features only via the role those features play in generating experiences. A central motivation for this Ramseyan view is a form of *empiricism*. The *concepts* we have of the features of the external world must, the thought goes, be derived from perception. But then the question arises what perception discloses to us about the natures of these features, and, in turn, what kinds of concepts we can have of them. On a view that is internalist about both phenomenology and content, the features of the external world do not themselves make any direct contribution to the nature of a perceptual state. So any concept of an external-world feature we derive from perceptual experience will have to latch onto that feature indirectly, presenting it as merely the unknown cause of that very experience.

That is why, the Ramseyan insists, we never do have acquaintance with the contents of our own minds; we do not know how our mental states represent the world to be in itself. The concepts we have, which compose the contents of our mental states, including our perceptual representations, pick out external-world features only in a non-nature-revealing way, as the unknown external causes of our internally-constituted experiences. What we need, then—in order to answer the Ramseyan challenge and secure our internal acquaintance—is a set of concepts that pick out external-world features *directly*, in terms of their natures, rather than in terms of their effects on us.

On Locke's resemblance account, perceptual contents do directly pick out particular, determinate features; they do not have the FLEXIBILITY of Ramseyan descriptive

contents. But the way this DETERMINATENESS is achieved makes it impossible for perceptual experiences to be WORLD-DIRECTED. So we need a different way of explaining the determinateness of our perceptual representations, one that makes intelligible how they could represent a mind-independent world. That is, we need a set of concepts that can feature in perceptual contents that are both determinate and world-directed.

My claim is that our concepts of primary qualities are precisely such concepts. They pick out the features they represent *directly*, not via Ramseyan descriptions. And the determinate features they pick out are features of a *spatial manifold*, features that can be instantiated in a mind-independent, spatial world, in a perfectly intelligible way.

Consider, for example, our concept of squareness. The concept of a square object is not just the concept of an object that has the feature, *whatever* it is, that typically causes experiences of the type we call “experiences of squareness.” It is, instead, the concept of a *determinate shape*, of a particular way that a thing can be arranged in space.⁵³ Such concepts do not represent shapes as causes of perceptual experiences. They represent them *directly*, in terms of their determinate spatial natures: our concept of squareness is the concept of a figure *delineated by four equal sides joined at four right angles in a Euclidean plane*.

As is apparent from the specification of the concept of squareness just given, our determinate spatial concepts are, in a strong sense, independent of perceptual experience; their content can be spelled out without making any reference to experience at all. It thus seems appropriate to label them *a priori concepts*. And, indeed, our possession of these concepts is what allows us to perform the *a priori* reasoning of Euclidean geometry. When, for example, we prove the Pythagorean theorem utilizing our concept of squareness, we employ our *a priori* knowledge of what squareness is like—we make use of a determinate, *a priori* shape concept. That is what enables us to grasp the truth of the various propositions contained in the proof, regardless of what kinds of perceptual experiences we might have had (we do not, for example, need to have *seen* any right triangles, or any squares, in order to understand the proof).

The *a priori* character of our spatial concepts is the key to meeting the Ramseyan challenge. The Ramseyan, in an empiricist vein, demands to know how an internally-constituted perceptual experience could give us a concept of squareness as anything more than the unknown cause of that very experience. The reply is that *perceptual experience*, indeed, could not do so; but that this is irrelevant, since our concept of squareness is not *given* to us by perceptual experience at all. We have concepts of spatial properties that are not derived from experience, but are instead elements in an

⁵³ This formulation echoes McDowell (2011), who also argues that primary-quality perceptions do not present spatial features in terms of their relations to our experience.

a priori mode of thought.⁵⁴

Though our shape concepts, being *a priori*, are not *derived from* perceptual experience, they can *feature in* the contents of our perceptual representations. When we have an experience that presents an object as square, that perceptual representation itself, in virtue of having the phenomenology it does, employs the very same determinate spatial concept of squareness on which we have an *a priori* grasp.

The fact that it is the same concept of squareness that shows up in both our *a priori* reasoning and our perceptual contents is what allows us to have acquaintance with the spatial contents of our experiences. We know which particular way a shape experience represents the world to be, because we know—due to our *a priori* grip on spatial properties—what it is for a space to be arranged in a particular way, for various geometrical properties to be instantiated within a spatial manifold. When we perceive an object as square, we understand what the world must be like, if that perceptual representation is to be veridical: the world would have to contain an object *situated in space in a particular manner* (namely, *with a surface delineated by four equal sides joined at four right angles in a Euclidean plane*).⁵⁵

This last point helps us see how our internally-constituted spatial concepts can represent the features of a mind-independent world. The nature of spatial concepts themselves makes them fit to represent such mind-independent features. Strawson famously argued that having a conception of a world as spatial is the *only* way to represent that world as mind-independent. I do not want to take a stand on Strawson's claim. But I do want to insist on something weaker: a conception of a world as spatial—a set of mental states that represents various shape properties as being instantiated—is at least *one* way of representing a world as mind-independent. Being

⁵⁴ Notice that this anti-empiricist move also plays a key role in responding to the other half of the MAGIC WORRY—the question of how internally-constituted perceptual states can be WORLD-DIRECTED. Campbell, in particular, is driven—quite explicitly—by the kind of empiricist assumption about concept formation I have attributed to the Ramseyan. He takes Berkeley to be presenting a formidable challenge because he, like Berkeley, thinks we need to explain how *experience* can provide us with concepts of a mind-independent world. Trying to explain how experience could possibly do so leads Campbell to conclude that experience must be externalistically constituted. But, having rejected the empiricist picture of concept-formation—on which our concepts of the external world must be *derived from* experience—by allowing that primary-quality concepts are *a priori*, we no longer have to meet Campbell's (or Berkeley's) explanatory challenge on its own terms. We can acknowledge that experience cannot itself *ground* concepts of a mind-independent world; but we can nonetheless take our experiences to *deploy* our concepts of a mind-independent, spatial world—concepts whose source lies outside experience. See Chapter 3 for further discussion.

⁵⁵ Note that I am *not* suggesting that actually *articulating* this kind of definition is what gives us our grip on geometrical properties like squareness. We don't have to go through Euclidean proofs to understand what squareness is; rather, our *a priori* grip on what squareness is allows us to do the proof, and to see that the definition of squareness given above is an appropriate one.

situated in space in a particular way—having a surface delineated by four equal sides joined at four right angles, for example—just *is* a way for an object to exist in a mind-independent, spatial world.

The point I am making here can be helpfully illuminated by attending to a contrast between *arithmetical* and *geometrical* objects emphasized by Tim Maudlin. While both kinds of entities can be useful for *modeling* various features of the physical world, only the latter can sensibly be thought of as possible *constituents* of that world. This distinction is often lost in modern discussions of physics because the practice of modeling the world via arithmetical entities—for example, modeling the Newtonian hypothesis that the physical world is a three-dimensional Euclidean space using the set of ordered triples of real numbers—has become deeply entrenched. But it is a crucial distinction to make, if we are to understand how our spatial concepts can feature in perceptual representations that purport to reveal a mind-independent world. The subject matter of physics is the mind-independent, external world in which we live. And that world is intelligibly thought of as itself *geometrical*, in a way that it is *not* intelligible to suggest that the world might itself somehow be *arithmetical*. Here is how Maudlin puts the point:

For several millennia, the actual space we live in was believed to be a three-dimensional Euclidean space, but no one ever imagined that the actual space we live in consisted in ordered triples of real numbers. Such a proposal makes no sense whatever.⁵⁶

Newton presented his physics geometrically because the subject matter—motion in space—is itself geometrical. To put it bluntly, the physical world is not composed of numbers or of entities for which the standard arithmetic operations are defined. The physical world does contain physical magnitudes that have a geometrical structure. Geometry is more directly connected to the physical world than is arithmetic.⁵⁷

The idea here is that our spatial concepts—the very concepts that I have been suggesting are *a priori*—are important because they are concepts of properties that can sensibly be seen as determining a particular *way for a world to be*. As Maudlin notes, for millennia the space in which we live was assumed to *be* a three-dimensional Euclidean space (*be*, not just be *modeled* by; this is the crucial distinction). In the last century, physicists, driven by the discoveries of Einstein, came to doubt the *truth* of this assumption; but the assumption is a perfectly *intelligible* one. Being a three-dimensional Euclidean space is a *possible* way for a mind-independent world to be; so

⁵⁶ Maudlin (2014, p. 8).

⁵⁷ Maudlin (2012, p. 25).

it is intelligible how a perceptual content comprising a set of Euclidean spatial concepts could represent a mind-independent world (and represent that world as being, spatially, some determinate way).

1.7.3. *The Distinction Between Primary and Secondary Qualities*

Before concluding, I want to draw out a feature of my account that has already been implicitly acknowledged. I suggested above that the key to solving the PROBLEM OF INTERNAL ACQUAINTANCE is to reject the Ramseyan's empiricist claim that all our concepts of the external world must be derived from experience. We have spatial concepts that are *a priori*, and these concepts confer a sort of *representational inflexibility* on the contents of the primary-quality *perceptions* in which they feature. Our *a priori* concepts provide an *anchor* from outside of experience that gives perceptual shape-representation determinate content; this is why we are not forced to understand the properties represented in terms of their effects on our experience. But in the case of *secondary* qualities, rejecting the empiricist claim about concept-formation looks wildly implausible: we simply don't have any *a priori* color concepts; our concepts of secondary qualities *will* have to be derived from experience. Since colors themselves are features of the mind-independent world—*pace* Berkeley, it is tigers, and not tiger-experiences, that are orange—our concepts of colors *will* have to have just the kind of indirect structure that the Ramseyan proposes: we know orange only as *the feature, whatever it is, that causes a certain kind of experience*.⁵⁸ And so our perceptual representations of color properties *will* have REPRESENTATIONAL FLEXIBILITY. This explains why inversion scenarios seem so natural when applied to the secondary qualities.

What this account of perceptual content suggests is a way of drawing the line between primary and secondary qualities in terms of the kinds of *concepts* we have of the respective properties, rather than in terms of the *metaphysical natures* of those properties. On this picture, colors are no less real or mind-independent than shapes: orange is *the property that typically causes a certain kind of experience in normal observers (in the actual world)*; and that property will be a perfectly real, perfectly *objective* property.⁵⁹ So the metaphysical status of primary and secondary qualities is the same. What distinguishes the two kinds of qualities is a difference in the kinds of concepts we have of them. In the case of the primary qualities, we have *a priori* concepts that pick out determinate properties directly, not via a description that involves reference to experience. In the case of secondary qualities, we have no such *a priori* concepts; instead, our concepts of secondary qualities pick out external-world features only indirectly, as *the properties that cause certain kinds of experiences in us*.

⁵⁸ We can, through sophisticated scientific investigation, also come to have another kind of concept of orange: something like *reflecting light of bandwidth 590-620 nm*. But that is not the concept of orange that factors into our *perceptual experience* of the color.

⁵⁹ On the contemporary scientific picture, this property turns out to be something like a particular set of spectral-reflectance profiles. Such a property is clearly *not* mind-dependent.

1.8. CONCLUSION

David Chalmers has said that, on the Ramseyan account (which he endorses), “it is because we demand so little that we know so much.”⁶⁰ Chalmers makes this claim in addressing a skeptical challenge to our knowledge of the external world; he is highlighting the fact that the Ramseyan account, just like the Putnamian separatist picture on which it is built, can provide us with a kind of guarantee against traditional skeptical worries about the truth of our world-directed beliefs. But what seems striking about the pictures we get from both Chalmers and Putnam is how much we *don’t* know on such views: we don’t know what the external world is like in itself, and, even more disconcertingly, we don’t know how we *take* the world to be. We achieve a guarantee of correspondence between our mental states and the world only by forfeiting our acquaintance with our own mental states.

In order to secure acquaintance with the contents of our own minds, we need to be more demanding—we need to acknowledge that our perceptual states present the external world to be some determinate way. But *how* can a state represent the world to be a particular way, independently of its relations to that world? This is Putnam’s challenge. The answer is that we have a kind of state—experiential representation of primary qualities—that *does* present the world to be a particular way, independently of what causes it, because it presents the world as instantiating *determinate spatial properties*, the nature of which we firmly grasp in virtue of our *a priori* concepts. We know what squareness is when we do geometrical proof; and we understand that that very property can be instantiated in a mind-independent world. So, in having a state that represents the presence of a square object, we know which particular way we take the world to be. We know the determinate, world-directed contents of our internally-constituted experiences of space.

This picture depends on two crucial claims about our spatial cognition: first, that the spatial concepts we employ in doing geometrical proof are genuinely *a priori*; and second, that these *a priori* concepts feature in the contents of our spatial experience. In the next two chapters, I offer my defense of each claim.

⁶⁰ Chalmers (2010, p. 490).

CHAPTER 2

A PRIORI CONCEPTS IN EUCLIDEAN PROOF

In the Introduction, I suggested that a proper understanding of spatial experience requires attention to the connection between the practice of Euclidean proof and the way our experience represents the spatial properties of the objects we perceive – the connection that explains the case of the carpenter, who applies her proof of the Pythagorean theorem to the wooden beams she sees in her workshop.

My goal in this chapter is to begin to explore this connection by investigating the kind of cognitive activity we engage in when doing Euclidean proof, and the kinds of concepts involved in the practice. This is a question of what Strawson once called “descriptive metaphysics”: it asks about the nature of a cognitive activity that humans in fact perform (and have been performing for millennia), rather than seeking to correct or improve on our existing conceptual scheme.

In addressing a similar question (though with less of a purely descriptive emphasis), Einstein once contrasted what he called the “older interpretation” of the axioms of Euclidean geometry with his preferred “more modern interpretation.” Taking as an example the axiom “Through two points in space there always passes one and only one straight line,” Einstein (1992, p. 16) described the “older interpretation” as follows:

Every one knows what a straight line is, and what a point is. Whether this knowledge springs from an ability of the human mind or from experience, from some collaboration of the two or from some other source, is not for the mathematician to decide. He leaves the question to the philosopher.

As an answer to our descriptive metaphysical question, I propose to defend and elaborate on this “older interpretation” (thankfully, Einstein has left such work to the philosopher!). In particular, I will argue that (a) Euclidean proof is not a purely formal system of deductive logic, but one in which our grasp of “what a straight line is, and what a point is” plays a central role; and (b) our grasp of those notions is *a priori*, rather than being derived from experience.

2.1. EUCLIDEAN PROOF AS AN “AXIOMATIC SYSTEM”

For over two millennia, Euclid’s *Elements* was taken to be a paradigm of *a priori* reasoning. With the discovery that there are consistent non-Euclidean geometries, and the eventual realization that our own universe is not a perfectly Euclidean space,

the *a priori* status of our geometrical knowledge was called into question. There was a sense that Euclidean proof was not properly *rigorous*, and that rigor could only be achieved by separating out the properly *a priori* aspects of the practice—which, according to this new way of thinking, had to be purely formalistic—from the deliverances of perceptual experience. Einstein approvingly describes this “more modern” understanding of geometry—which he calls “axiomatics”—as follows:

Geometry treats of entities which are denoted by the words “straight line,” “point,” etc. These entities do not take for granted any knowledge or intuition whatever, but they presuppose only the validity of the axioms... which are to be taken in a purely formal sense, i.e. as void of all content of intuition or experience.... All other propositions of geometry are logical inferences from the axioms (which are to be taken in the nominalistic sense only).... In axiomatic geometry the words “point,” “straight line,” etc., stand only for empty conceptual schemata. (Einstein 1922, pp. 16-17)

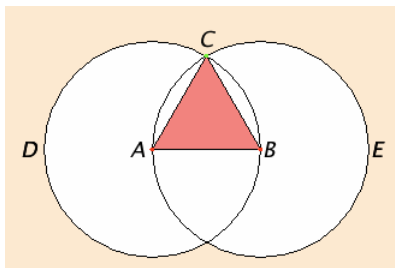
Such an “axiomatic” system of Euclidean geometry was first explicitly formulated by Hilbert in his *Foundations of Geometry* (1899).¹ Hilbert’s work was a hugely important step in the development of modern geometry, and, as Einstein himself says, it was crucial in the development of relativistic physics.² But, as I will explain below, such an axiomatic system was not, in any obvious sense, Euclid’s own: the proofs in Euclid’s text do not appear to meet the standards of axiomatic proof; and the formal tools used by Hilbert in spelling out such a system were developed over two thousand years after Euclid’s time.

As Einstein emphasizes, in order for a proof to be valid in an axiomatic system, each step must be derivable from the preceding steps via rules of logical inference that take account only of the *form* of the propositions in the earlier steps. But, famously, many of Euclid’s proofs, including the very first proof in the *Elements*, fail to satisfy this constraint. Euclid I.1, in which Euclid proves that it is possible “to construct an equilateral triangle on a given finite straight-line,” *AB*, begins thus:

Describe the circle *BCD* with center *A* and radius *AB*. Again describe the circle *ACE* with center *B* and radius *BA*. Join the straight lines *CA* and *CB* from the point *C* at which the circles cut one another to the points *A* and *B*.

¹ Tarski (1959) later showed that such a system could be built on a more modest set of axioms.

² Einstein (1922, p. 17) writes, “I attach special importance to the view of geometry which I have just set forth, because without it I should have been unable to formulate the theory of relativity.”



Euclid goes on to show that the segments AB , AC , and BC are of equal length, thereby proving that the triangle ABC is equilateral.³

Crucially, Euclid's proof requires that line segments be constructed connecting A and B to the point C , where the two circles intersect; this, in turn, requires that there *be* such a point of intersection. But the existence of C cannot be formally derived from the preceding claims (the axioms⁴ and earlier steps of the proof).

A standard way of illustrating this is to show that there are interpretations of the non-logical terms Euclid uses on which the axioms hold, but the claim that there is a point of intersection does not. As Michael Friedman (1992) has noted, given the minimal amount of logical structure in Euclid's formulation of the axioms, such interpretations can involve quite trivial models: even a system with just two points, given the appropriate interpretation, can be counted as satisfying the axioms; and, clearly, no third point C exists in any two-point model.

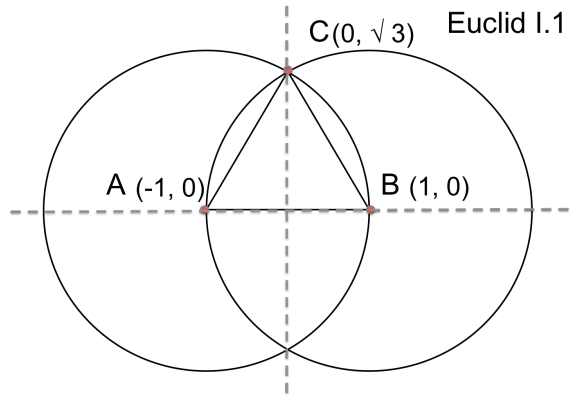
Here, I will be focusing on a somewhat richer model of Euclid's axioms, which will be especially useful for my purposes. It is a model in the so-called "rational plane," the geometrical space that is the result of taking a standard Cartesian coordinate plane, and then removing all points with irrational coordinates.⁵ Suppose we perform the construction of Euclid I.1 in this model, and let the points A and B have coordinates $(-1, 0)$ and $(1, 0)$, respectively. Then the point C at which the circles intersect would have coordinates $(0, \sqrt{3})$. But, since we are working in the *rational* plane, no such point exists: it is one of the irrational points we removed when we shifted from the real plane, \mathbb{R}^2 , to the rational plane, \mathbb{Q}^2 . Thus, in this model, the circles will fail to have a point of intersection. We have now found an interpretation on which the axioms hold, but the conclusion—that there is a point of intersection—does not. And this demonstrates that the proof in Euclid I.1, which depends

³ Translation and diagram from: Joyce, David E. *Euclid's Elements*. Available at: <http://aleph0.clarku.edu/~djoyce/java/elements/bookI/propI1.html>

⁴ The term "axioms" is used for the foundational propositions in Einstein's description of "axiomatics"; Euclid himself gives a set of "definitions," "common notions," and "postulates," rather than "axioms." While these notions differ in various ways, those differences will not be important for my purposes here; for ease of exposition, I will use the term "axioms" in describing Euclid's text to refer collectively to the definitions, common notions, and postulates.

⁵ My discussion here closely follows that in Friedman (1992, Chapter 1).

on the existence of such a point, cannot be valid.



The issue here turns on distinguishing between two models of Euclid’s system: one is a model in \mathbb{R}^2 , which is a *continuous* structure; the second is the model in \mathbb{Q}^2 , in which the plane is merely dense, but not genuinely continuous. The “intended” model, surely, is the former;⁶ and the point of intersection, C, does indeed exist in that model. But there seems to be nothing to rule out the \mathbb{Q}^2 model as an interpretation of the axioms.

In order to salvage the proof, then, we would need to find a way of ruling out the non-continuous \mathbb{Q}^2 model (henceforth, the “defective model”).⁷ And, as it turns

⁶ A caveat: the “intended model” of Euclidean geometry is not, strictly speaking, any kind of *coordinate* system at all. As Tim Maudlin (2014, p. 8-9) emphasizes, despite the contemporary practice of conflating truly *geometrical* objects, like the Euclidean plane, and *algebraic* or *arithmetical* objects, like a Cartesian coordinate system, there are important differences between the two. I am extremely sympathetic to Maudlin’s point; but, for ease of exposition, I will here (reluctantly!) follow the contemporary trend of ignoring the distinction between coordinate systems—which provide useful arithmetic *representations* of geometrical objects—and the geometrical objects themselves.

⁷ Strictly speaking, Euclid’s proofs do not require a fully continuous plane. As Friedman (1992, p. 61), following Tarski (1959), notes, if the constructions are carried out in a coordinate system based on the so-called “Euclidean extension” of the rationals—the set obtained by closing the rationals under the real square root operation, which Friedman labels \mathbb{Q}^{*2} —all of the proofs in the *Elements* can be shown to be valid. Lines and curves in \mathbb{Q}^{*2} , like those in \mathbb{Q}^2 , are merely dense, rather than genuinely continuous; but \mathbb{Q}^{*2} does contain some points not contained in \mathbb{Q}^2 (in particular, \mathbb{Q}^{*2} contains our point of intersection C, with coordinates $(0, \sqrt{3})$). In what follows, I will largely suppress this complication, focusing on the distinction between the defective \mathbb{Q}^2 model and the fully continuous \mathbb{R}^2 model. I offer two justifications for this. First, it seems extremely plausible to me that Euclid’s proofs were conceived of as involving fully continuous lines and curves, rather than the dense lines and curves of \mathbb{Q}^{*2}

out, Euclid’s second postulate does say that a line can be produced “continuously”; so we might think that the resources to rule out the non-continuous model are to be found in that axiom.

But we are evaluating the proposal that Euclid’s geometry is an axiomatic system, of the kind described by Einstein. If it were, each step would need to be derivable from the *form* of the earlier propositions. And the term “continuous” appears without further *logical* analysis in Euclid. It is simply a one-place predicate, to which we can, in evaluating validity within the axiomatic framework, assign any interpretation we like – including, for example, what we would now call “denseness.” So the second postulate doesn’t supply the formal structure that would allow us to rule out the defective, merely-dense model.⁸

At this stage, it might be suggested that the needed formal structure, though not *explicitly* spelled out by Euclid, was nonetheless implicitly intended. Now, I myself think that *something* along these lines is correct: Euclid intended to build into his axioms more content than is contained in the formal structure of the words in which he states them, and the tools to rule out the defective model are to be found in this additional content. But, crucially, I see no way to defend the idea that Euclid intended his axioms to include additional *formal* structure of the kind needed to fill the gap – the kind utilized by Hilbert. As has been detailed at length by Friedman (1992), the relevant formal definition of continuity requires heavy use of the quantifier dependence characteristic of modern polyadic logic, as first spelled out by Frege, over two thousand years after Euclid’s time. It is thus extremely implausible that Euclid’s system relied (even implicitly) on such formal tools.

So, if we interpret Euclid’s *Elements* as an axiomatic system, we must conclude that Euclid I.1 is invalid. Since this is the very first proof in Euclid’s system, and since the proofs build upon each other, any defect in I.1 will carry over to the bulk of the results. Thus, if we regard Euclidean proof as an axiomatic system, in Einstein’s sense, we will have to regard it as a failed one: virtually no results are validly proven in it.

But this would be a puzzling conclusion. For Euclidean proof was massively successful: In the words of Kenneth Manders (1995, p. 80), it was “a stable and fruitful tool of investigation across diverse cultural contexts for over two thousand years.” Euclid’s text itself is “virtually without error... every result has a counterpart in modern mathematics” (Manders 2008, p. 67). We need some account of this undeniable success, and the proposal that it was (in the words of Giaquinto (2011)) “just immense good luck”—that Euclid’s *failed* axiomatic system somehow *happened* to produce only correct theorems—is not a plausible contender.

(support for this claim will be given in §2.6). Second, the arguments I make below rely only on the fact that Euclidean geometry requires a field richer than \mathbb{Q}^2 (this will become clear in §2.5); and so the discussion could be reformulated (though with an extreme loss of expositional clarity) with \mathbb{Q}^{*2} substituted for \mathbb{R}^2 .

⁸ This point is emphasized by Friedman (1992, p. 60).

We are left with this: Euclid's system cannot be seen as a successful axiomatic system; and yet, it was a successful system. The obvious conclusion to draw is that it is not an *axiomatic* system at all: the *form* of the postulates does not exhaust their content.

In a way, this should have been obvious from the start. For we are doing descriptive metaphysics, asking about the nature of Euclidean proof – an actual human practice in which most of us, during our high school education, have taken part. And the kind of purely formalistic reasoning described by Einstein is simply alien to that practice. Consider Hilbert's (1899) description of what it means to regard the axioms of his geometrical system in the way Einstein recommends:

[T]he basic elements can be thought of in any way one likes. If in speaking of my points I think of some system of things, e.g. the system: love, law, chimney-sweep... and then assume all my axioms as relations between these things, then my propositions, e.g. Pythagoras' theorem, are also valid for these things.

Whatever the merits of this way of thinking, it is clearly not the type of thinking we engage in when we actually do basic Euclidean geometry. No high school student, in working through the proof of the Pythagorean theorem, takes herself to be reasoning as much about chimney sweeps as about straight lines and triangles. The subject's own sense of what she is doing needn't be taken as absolutely authoritative. But proving the Pythagorean theorem is something that the student does; and it strains credulity to suggest that, despite her firm introspective judgments to the contrary, she has all along been reasoning about formal features of uninterpreted signs, which are as readily applicable to chimney sweeps as to the triangles she calls to mind when performing the proof.

So it seems we must acknowledge that the practice of Euclidean proof—the practice both of Euclid himself and of the contemporary high school student—is not the kind of formalistic practice developed by Hilbert and praised by Einstein. The *contents* of the concepts employed in the axioms—concepts like “circle” and “line”—play a crucial role in the inferences Euclid draws.

What we need now is an analysis of the nature of those concepts. In particular, I want to explore the question Einstein left to the philosopher: Does the content of the concept “circle”—the content that makes the proof in Euclid I.1 successful, in spite of the gap left by the formal elements in Euclid's text—spring from experience, or from some *non-experiential* “ability of the human mind”? Below, I argue—perhaps surprisingly—for the latter: the relevant concepts are *not* derived from experience.

2.2. STRAWSON'S “PICTURABLE MEANINGS”

In response to worries about the inadequacy of the formal elements of Euclid's

axioms to ground Euclidean proof, many theorists have been drawn to the idea that the “extra material” needed to plug the gaps is something we derive from our sensory experience. Strawson, in his explication of Kant’s theory of geometry, suggests that Euclid’s axioms can be interpreted as concerning something he calls “phenomenal geometry”: the geometry of our visual experience and imagination. According to Strawson, in performing Euclidean proof, we make use of “picturable meanings” to draw conclusions about this phenomenal geometry.

Strawson’s proposal here is notoriously hard to pin down (see Hopkins 1973). But we can perhaps get a better grip on it by looking at what he has to say about the gap in Euclid I.1, and how our “picturable meanings” help fill it. According to Strawson, the key is that

we cannot picture to ourselves any figure which we should be prepared to count as adequate to the sense of [the steps of the construction in I.1 that precede the assumption that the circles intersect] for which this assumption does not hold. The picture of the sense of the description rules out any alternative to this assumption. (1966, p. 284)

So it seems that the “picture of the sense of the description” is supposed to fill the gap we identified above in Euclid I.1: this picture *rules out* any defective models of the axioms in which the circles fail to have a point of intersection.

But this proposal simply cannot be made to work. For the gap in question—the need to rule out the defective Q^2 model—cannot be plugged by any kind of *picture*. The problem is this: carried out in the defective model, the construction of Euclid I.1 would result in two non-continuous but *everywhere dense* curves. These curves would fail to intersect, because the breadthless point at which they “would” intersect is not included in the space of the construction. But there are infinitely many points arbitrarily close to the “missing” point of intersection on each curve. So, if we ask what a *picture* of this defective model would look like, there can be only one answer: it would look *exactly* the same as the picture of the construction carried out in the intended R^2 model. Thus, contrary to Strawson’s claim, there *is* a picture adequate to the sense of the construction in I.1, in which the assumption of intersection does not hold: namely, the picture of the defective Q^2 model, which, *as a picture*, is *identical* to the picture of the construction in the intended R^2 model.

This point in some respects mirrors Descartes’s argument about the inadequacy of imagistic reasoning to capture our concepts of certain polygons. According to Descartes, we have (through the “pure understanding”) a “clear and distinct” idea of a chiliagon (a 1,000-sided figure) even though our *imagination* can generate only a “confused representation” of such a figure. In particular, Descartes (1641/1996, p. 50) notes that his imagistic representation of a chiliagon “differs in no way from the representation [he] would form if [he] were thinking of a myriagon [a 10,000-sided figure].”

But despite the clear similarities, the point I am making about the role of our concept of continuity in Euclidean proof, and the inadequacy of Strawson's picturable meanings to explain its origin, is in a crucial sense more robust than Descartes's argument about our concept of a chiliagon. For the claim that our imaginative capacities cannot explain the origin of our chiliagon-concept might be thought to stem from an overly narrow conception of how we derive concepts from experience. It is true that we cannot imagine a chiliagon in a way that distinguishes it from a myriagon by forming a simple *static* image of a geometrical figure. But we can imagine a certain *course* of visual experiences through which we could identify a figure as, specifically, a myriagon. Imagine visually scanning along the border of a polygon (perhaps with the aide of a magnifying device), marking off sides as you come to them, and making a tally mark in a notebook for each side. When you return to the starting point (i.e., when you first see a side that you have already marked), you can count the number of tally marks and thereby know how many sides the polygon has. This would allow you to distinguish between a chiliagon and a myriagon, by utilizing a (real or imagined) sequence of visual experiences. By taking into account the possibility of such sequences, we can explain how a concept of a chiliagon could, in a sense, be derived from experience.

When we turn to the question of how we might derive the concept of continuity from experience, however, this kind of explanation fails. For, in the case of our concept of continuity, it is not only static experience that fails to get us onto the relevant concept. No imagined sequence of experiences, even if we allow a high degree of idealization of our imaginative capacities, could distinguish a continuous curve from a dense one.

To see this, imagine “zooming in” on the part of the Q^2 diagram of Euclid I.1 at which the curves “would” intersect. No matter how “far down” you zoom, there will always be infinitely many points on each curve arbitrarily close to the gap, which is itself a breadthless point. Thus, even if such a process were to continue over an indefinitely long sequence of experiences, and even if no limit were set on the magnifying power of the “zooming-in” process, the gap in the dense curves is never going to be *visible*. Such a gap, then, can surely make no difference to the *picture* of the construction in Euclid I.1, or to any “picturable meaning.”

Now, it should be noted that Strawson does not take his “picturable meanings” to provide an exhaustive account of Euclidean geometry. He acknowledges that some degree of what he calls “conceptual idealization” will also be needed in spelling out how “phenomenal geometry” fits into the practice of Euclidean proof as a whole (1966, p. 287). But my point is simply that the gap in Euclid I.1 can't be filled by pictures, or by Strawson's “picturable meanings”; if an account like Strawson's is to explain how that gap is filled, whatever is involved in the *non-imagistic* “conceptual idealization” will have to do the real work.

2.3. DIAGRAMMATIC REASONING

More recently, work in the “philosophy of mathematical practice” has emphasized the role of a different class of visual items—the diagrams included in Euclid’s text—in offering an explanation of how the gap in Euclid I.1 can be filled. Manders (1995 and 2008) suggests that an analysis of the role of “diagrammatic reasoning” in Euclid reveals that we are indeed licensed “to attribute the intersection points the diagrams show” – that is, to derive the existence of the intersection point C from the diagram that accompanies Euclid’s text in I.1 (2008, p. 66).

In defending this claim, Manders (2008, pp. 69-70) begins by distinguishing what he calls “exact” features of Euclidean diagrams—such as two lines being of equal length—from “co-exact” features—such as one figure being a subpart of another. Manders’ suggestion is that claims about *co*-exact features may be legitimately inferred from diagrams. Such inferences are legitimate, according to Manders, because the features in question are “insensitive to the effects of a range of variation in diagrams.” By contrast, claims about exact features, such as the equality of the lengths of two line segments “would fail immediately upon almost any diagram variation,” and so cannot be legitimately inferred from diagrams.

In the specific case of Euclid I.1, Manders claims that the existence of the intersection point of the circles is a co-exact feature, since “[a]s we distort the ‘circles’ in I.1, their intersection point C may shift but it does not disappear” (2008, p. 69).

But, as can be seen from the preceding discussion, there is a certain kind of “distortion” of the construction in I.1—the shift from the construction in the real plane to the construction in the rational plane—that would undermine the inference to the conclusion that the circles intersect, even though it involves no radical change to the diagram’s appearance. Indeed, such a “distortion” involves no change *whatsoever* in the diagram’s appearance. As explained above, the *pictures* of the two constructions—in the intended model and in the defective one—would be visually *identical*. The point of intersection thus seems to fit better with Manders’s description of *exact* features: its existence is sensitive to the most minor of variations in the diagram.

Now, to be fair, Manders does not take the point of intersection to be stable under *every* possible type of distortion. He writes, “As long as the ‘circles’ in I.1 *are continuous closed curves*, no amount of distortion can eliminate intersection points” (2008, p. 71). So the kind of “distortion” I just mentioned—shifting from a construction in the real plane to one in the rational plane—is one that Manders means to rule out.

But this simply pre-supposes what is in question, when we are asking whether there is a gap in Euclid I.1, and how it might be filled: the question is what licenses us to assume that the circles *are* continuous closed curves in the first place, such that we may legitimately infer that they do intersect. And, despite his emphasis on the role of diagrams in Euclidean proof, Manders’s real answer to our question is that the needed content is conveyed by the *text* that accompanies Euclid’s “diagram

entries.” Manders writes, “Whenever a diagram entry is made, the text records the exact character (straight, circular) of the element entered. There is thus no need to later judge this from the diagram.” In Euclid I.1, the “diagram entries” include two “circles.” According to Manders, the accompanying text, in using the term “circle,” is specifying that these are to be regarded as “(at a minimum) continuous non-self-intersecting curves” (2008, p. 71). So the continuity of the curves is not, in the end, visually derived from the diagram; it is built into the diagram, *conceptually*, in virtue of the content conveyed by the term “circle” in the accompanying text.

Thus, it turns out that even on Manders’s picture (which is meant to emphasize the role of diagrams in Euclidean proof), the content needed to fill the gap in Euclid I.1 comes not from anything visual in the proof, but from something conceptual that we bring to the practice: our understanding that a Euclidean construction of a circle is to be counted as resulting in a continuous curve. The question of whether the relevant understanding “springs from experience,” then, is not answered by pointing to the role that our visual experience of the diagram plays; it turns on the nature of our *concept* of a circle—the concept used in *constructing* and *interpreting* the diagram—and whether *it* is derived from experience.

2.4. GEOMETRICAL CONCEPTS AND PERCEPTUAL APPEARANCES

In recent work, Marcus Giaquinto (2007, 2011) has offered a subtle and complex answer to this question about the nature of our geometrical concepts. Giaquinto begins by distinguishing what he calls our “geometrical concept” of circularity from our “perceptual concept.” Our geometrical concept is the one at work in Euclidean proof. It applies to objects possessing the properties needed to ground the proofs: “perfect” circles with null breadth and genuine continuity. Since *perceptual* circular objects, such as coins and buttons, lack these “perfect” features, Giaquinto suggests that the subject matter of Euclidean geometry is a set of *abstract* geometrical figures. This explains how the practice can be successful, even though it ascribes to its “circles” and “lines” features that no object we encounter in perception actually instantiates. Furthermore, since it is the geometrical concept at play in Euclid, the defective model of Euclid I.1 is ruled out by the content of that concept. In particular, the defective model is ruled out by the fact that our geometrical concept is a concept of a *perfect* circle – a figure that instantiates, in Giaquinto’s words, “closedness, that is, having no gap” (2011, pp. 283-296).

But where, we might wonder, does this *geometrical* concept of a circle come from? Giaquinto’s answer is that our geometrical concept is, in a somewhat complex way, derived from experience, and from the developmentally prior *perceptual* concept. Here is how Giaquinto thinks this process works: In perceiving or imagining circular objects or figures, we form a perceptual concept of circularity. This concept applies to many objects that are not *exactly* circular, according to the strict mathematical definition. But we have a natural tendency to rank those objects as “better” or

“worse” circles. Those that have visible gaps, or are visibly asymmetrical, are ranked worse than those that are less gappy or asymmetrical. At the limit of our acuity, if a circle contains *no* visible gaps or asymmetries, we judge that it “looks perfect.” We then form the concept of a perfect circle as a figure that *is* as a perfect-looking circle *looks*; it is a figure that “has the spatial properties that any perceptual circle must appear to have in order to look perfect” (2011, p. 291). This is our geometrical concept of a circle.

Giaquinto is careful to flag the ways in which the concept we employ in Euclidean proof includes features—such as continuity—that are not possessed by the objects we perceive. But, I want to suggest, Giaquinto’s story about how we come to have our geometrical concepts, like Strawson’s earlier account of “picturable meanings,” still puts too much weight on the role of *visual experience*.

Again, on Giaquinto’s picture, the features included in our geometrical concept are those that a “perceptual circle must appear to have in order to look perfect.” But, I claim, a concept of this sort—one derived from features that perceived circles “appear to have”—could never distinguish between a continuous curve and a merely dense one.

To see why, consider what it would be for a figure to appear to have the property of mere-denseness. A circle that is merely dense, as noted above in the discussion of Strawson, would have no *visible* gaps. So lacking any visible gaps seems to be the way a circle would appear, if it appeared to have the property of mere-denseness. And that is precisely how, on Giaquinto’s own account, a perfect-looking circle looks: it has no visible gaps. Thus, a perfect-looking circle is one that *appears* to be merely-dense. But then, according to Giaquinto’s definition, a circle that *is* perfect would be one that *is* merely-dense. So our geometrical concept of a circle—our concept of a perfect circle, as defined in this perceptual way—would be a concept of a merely-dense curve. But then that concept couldn’t possibly rule out the defective model after all.

Now, Giaquinto might protest that a circle can’t *look* to be merely-dense: to have gaps, but only ones that are invisible. In a sense, this is surely right: since the gaps in question are invisible, they won’t affect the way the circle looks. But it seems equally true, then, that a circle can’t *look* to be genuinely continuous, *rather than* merely-dense: the *absence* of such invisible gaps, which is what distinguishes the continuous curve from the merely-dense one, also won’t affect the way the curve looks. There is nothing in the way a circle looks—even when it “looks perfect”—that privileges the property of continuity over mere denseness. And so a concept of a circle that *is* the way a perfect circle *looks* cannot do the work of ruling out the merely-dense defective model of Euclid I.1.

2.5. VISUAL LIMITS

Giaquinto’s story about how we come to have the “perfect” geometrical concepts at work in Euclidean proof relies on the following idea: we begin with a visual concept of a circle; we note that there is a certain ordering in our visual experiences, in that some circles strike us as “better” than others; and we then consider what the *limit* of this “better than” ordering would be. Giaquinto assumes that the limit in question—which, on his account, defines our concept of a perfect circle—will be a continuous curve. But, as argued above, Giaquinto’s account fails to explain how our concept, understood in this way, could pick out continuous curves, rather than merely dense ones. In this section, I want to investigate further whether a kind of “limit procedure” could explain how we derive our concept of a continuous curve from experience. I will argue for a negative answer: within the domain of experience, the limit of our concepts of curves—the ideal to which “perceptual curves” can be conceived of as approaching—is *denseness*, not full continuity. So no appeal to “limit concepts” can vindicate the claim that our geometrical concepts, which include the concept of circles as genuinely *continuous* curves, are derived from experience.⁹

The kind of proposal I have in mind is discussed by Crispin Wright, who suggests (without endorsing the suggestion) that we might be able to form “perfect” geometrical concepts on the basis of experience, through a certain kind of “idealization.” He writes: “The kind of idealization involved in the notion of perfect circularity... corresponds to a movement to the limit of a scale, as it were, whose intermediate values are ordered by a comparative – ‘is more circular than’” (1986, p. 15).

As it stands, though, this thought doesn’t help much with the question of whether a concept of continuity can be derived from experience; for it requires that we have a grip on what the comparative “is more circular than” has as its limit. If “circular” in “is more circular than” expresses Giaquinto’s *geometrical* concept of a circle—a concept of a continuous curve—then the limit of “is more circular than” is, surely, a continuous curve. But what is at issue is how we come to have the concept of a genuinely continuous curve in the first place, and whether such a concept can be derived from experience. So, in trying to explain how we could get the concept of continuity *from experience*, via a limit procedure, it seems we must interpret “circular” (in “is more circular than”) as picking out a *perceptual* concept – Giaquinto’s perceptual concept of a circle. But now the trouble is that we have been given no reason to suppose that the relevant limit—the limit to which “is more *perceptually* circular than” approaches—will be a continuous curve.

⁹ Note that this does not mean that continuous curves are *excluded* from the category determined by our perceptual limit-concept. Continuous curves are dense; so they, along with *merely* dense curves, would fall into the category determined by the limit-concept of a dense curve. The point is just that this limit-concept does not specifically pick out, from *within* this category, the genuinely continuous curves, as opposed to the merely dense ones (or vice versa).

This might suggest that the idea of a limit procedure can do no work here. But I think we can extract from my earlier discussion of Strawson a way of defining the limit to which experientially-derived concepts (like “is more perceptually circular than”) approach, which does not merely leave us where we started.

The proposal is this: First, we abstract from the contingent imperfections of our own visual capacities and circumstances – the poor acuity, the finite time we have to visually investigate, the minimal tools we have for magnification. We suppose these limiting factors to be steadily removed, bringing us ever closer to a kind of experiential ideal. Then we ask: Given such an idealization of visual capacities and circumstances—granting unlimited time for investigation, endless magnification, etc.—what kinds of properties could make a difference to visual experience? What would visual experience or imagination, thus idealized, allow a subject to pick up on?

The idea is that we can think about the limit to which a perceptually-derivable concept approaches, in a way that takes us beyond the concepts we can read off directly from our own experience, by thinking about what the “limit” of experience itself would be. Note that this proposal is quite expansive in what it counts as derivable from experience, via such a limit procedure. It does not restrict the domain of visually-derivable concepts to features that are actually observable in our experience,¹⁰ nor even to the domain of features we humans could observe, with more powerful tools or more favorable circumstances. The minimal requirement it imposes is that, if we are to count a concept as derivable from visual experience—as the limit towards which a visual concept (like “perceptual circle”) approaches—we must be able to see how the feature picked out by the concept could show up in something recognizable as the “limit” of visual experience.

Using this procedure, we can first consider what such idealized experience of a curve that is *non-dense*, but which has only extremely small gaps (of, say, less than a nanometer), would be like. Our actual visual experience will not pick up on any gaps in such a curve. But, since the curve is non-dense, there will be two points along its length between which there is no third point. So, using our idealized visual imagination, we can “zoom in” on the region that contains those two points and the empty interval between them; and we will thereby observe that the curve does not satisfy the definition of denseness. That is, our idealized imaginative procedure gives us an experience-based grasp of the distinction between dense curves—those along which, between any two points, there is a third—and non-dense curves. So denseness is a concept we could derive from experience, via our limit procedure.

Next, we can consider whether this idealized imaginative procedure can get us onto the *further* distinction, within the category of dense curves, between those that are *merely* dense and those that are genuinely continuous. Consider, first, what the idealized experience of a genuinely continuous curve would be like. Since such a

¹⁰ A condition of this narrower sort is suggested as determining what counts as imaginable in the discussions in Strawson (1966, p. 282) and Hopkins (1973, pp. 22-23).

curve has no gaps at all, zooming in on any part of it will of course never reveal any gaps.

The question now becomes whether idealized imaginative investigation of a *merely* dense curve would be any different from this, such that we could derive from our procedure a way of conceiving of genuinely continuous curves as *distinct* from merely dense ones.

Unlike genuinely continuous curves, merely dense curves do have gaps. But, I want to suggest, even fully idealized visual experience could not pick up on these gaps. Remember that between any two points on a merely dense curve there are infinitely many further points, reaching arbitrarily close to those points. This makes the gaps along the curve—which, again, do exist—inaccessible to vision, even in its fully idealized form. As you zoom in on a merely dense curve, you will simply continue to see more and more points.

Because this point is so central, I want to linger a bit on why a gap in a merely dense curve is not the kind of thing that is accessible via idealized visual imagination. Consider the way I described the idealized procedure in the case of the non-dense curve. There, the idea was that we could find two points between which there is no third point, and, through “zooming in,” observe the gap separating them. This process requires picking out some point, and then observing both that point and the *next* point along the curve. It is the gap *between one point and the next* that is visually accessible, via the idealized procedure.

But the gaps in a merely-dense curve do not lie between some given point and the *next* point along the curve. In the case of dense curves, for any given point, p , there is *no such thing* as the “next” point. We can identify a point p along the curve, and we can note of other points that they are further along the curve than p . But there is no way to define a point p' that is the *next* point past p : because the curve is dense, for any candidate “next” point p' , there will be another point p'' that is between p and p' . In order to visually discern a gap in a curve, a subject needs to be able to perceive (or imagine) that gap as falling between two given points. The gaps in a merely dense curve are not like this: they do not fall “between” any two points in particular. And so they are simply not the kind of phenomenon we can get a grip on through experience, however idealized.

What this suggests is that the dense/non-dense distinction marks off a kind of visual limit: it defines, granting full idealization, the boundary between the kinds of curves we could visually experience or imagine as gappy (curves that have gaps falling between two specific points) and those we could not (either because the curves are genuinely continuous, and so simply lack any gaps; or because they are merely dense, and so have, between any two points, infinitely many further points, making the gaps that do exist inaccessible to experience). The further distinction, *within* the category of denseness, between *mere* denseness and genuine continuity lies beyond this idealized experiential limit. And so we cannot derive a concept of continuity as a visual limit-concept.

At this stage, it will be instructive to consider Delia Graff's discussion of so-called "phenomenal continua." Graff (2001, p. 923) writes:

[I]magine we could have the following two experiences: the first, of a cursor on a computer screen looking to move discontinuously from one pixel to the next (imagine also that the pixels are incredibly small), but in an even way; the second of a cursor looking to move *continuously* from one pixel to the next.

Graff notes that we could not tell the difference between these two experiences. But she suggests that we could *imagine them as* distinct experiences – one where the cursor's motion looks genuinely continuous, one where it does not. So Graff might seem to be calling into question my claim that the distinction between continuity and its absence is not something we can get onto via (idealized) imagination.

I do not want to dispute Graff's claim that we can (through idealization) imagine her two scenarios as distinct. But note that the two cases Graff describes do not just fall on different sides of the continuous/non-continuous boundary; they also fall on different sides of the dense/non-dense boundary. The pixels of a computer screen do not form a dense ordering; for each pixel, there is a determinate *next* pixel. Thus, our ability to imagine the motion of the cursor moving discontinuously from pixel to pixel—as *distinct from* the case of genuinely continuous motion—only shows that, as I suggested above, we can (in idealized imagination) distinguish denseness from non-denseness. In Graff's discontinuous *and non-dense* case, we can imagine picking out two pixels that lie next to each other along a non-dense line of the computer screen, and then zooming in to observe the cursor's "jump" between them; and this contrasts with the way we imagine a cursor moving along a dense line—whether continuous or not—where, for any point the cursor touches, there simply is no "next" point for it to "jump" to.

Indeed, as is almost explicit in her discussion, Graff actually seems to be concerned with the dense/non-dense distinction, rather than a distinction, *within* the category of denseness, between *continuity* and its absence. Strikingly, Graff's statement of what is needed for a process to count as "continuous" is essentially just a definition of *denseness*. In describing a case of an object, o , that changes position over time, Graff writes that, for the change to count as continuous, the following two conditions (which Graff takes to be equivalent) must hold:

1. If o changes its position over an interval, then it must change its position by some lesser amount over some proper part of that interval.
2. Between every two positions o occupies, there is a third position it occupies. (2001, p. 931)

Crucially, this requirement is satisfied by a scenario in which a point, o , moves along a merely dense line, like the rational number line. Consider o 's motion over an interval I that begins at a point B on the line and ends at a point E . In any sub-interval S of I , o will move a (rational) distance that is less than the length of BE . For, no matter what rational numbers B and E might be, there will always be another rational number, E' , between B and E for o to have moved to during S . Thus, meeting Graff's stated condition does not suffice for a process to be genuinely continuous: certain "gappy" processes—like the motion of a point along the merely-dense rational number line—meet Graff's condition.¹¹

That Graff would articulate a distinction between denseness and non-denseness in her discussion of phenomenal *continua*—rather than offering a distinction between genuine continuity and its absence—is, I want to suggest, quite understandable. For Graff is concerned with features of experience - with *phenomenal* continua. And, as I argued above, it is precisely the distinction between non-dense and dense curves (a category that includes both merely-dense *and* genuinely-continuous curves), rather than the distinction between denseness and continuity, that offers a plausible demarcation of the limit of *visual* (or more broadly *experiential*) concepts.

But it is a distinction *within* the category of denseness—between *mere* denseness and genuine continuity—that is needed to ground the proof in Euclid I.1. And so even the "limit-concepts" we can derive from experience fall short of the concepts we utilize in performing Euclidean proof.¹² Thus, the concept of continuity, which plays a central role in Euclid I.1, must be *a priori*.

2.6. CONCLUSION

That denseness, and not continuity, determines the upper bound of the visual might help explain a curious historical fact. As Friedman (1992, p. 60) notes, when mathematicians prior to Dedekind offered formal definitions of continuity, they tended to

¹¹ What Graff says about continuity is, strictly speaking, correct: she presents her condition only as *necessary* for continuity; and, since continuity entails denseness, meeting the conditions Graff lays out—which define denseness—is indeed necessary for continuity.

¹² Here we can see why it is immaterial to my argument that there is a less-than-continuous space, Q^{*2} , that suffices for Euclidean geometry (see fn. 6 above). For consider a (merely-dense) curve in Q^{*2} . Such a curve will, just like a continuous curve in R^2 , be indistinguishable, via our imaginative limit procedure, from the (sparser) merely dense curves of Q^2 . So the difference between Q^2 (a structure that is *not* sufficient to ground Euclidean geometry) and Q^{*2} (the *minimal* structure required to ground Euclidean geometry) lies beyond the limit of visually-derivable concepts: both fall on the same side of the dense/non-dense boundary, beyond which idealized imagination can make no further distinctions. And so the concepts we need for Euclidean proof—whether we take them to be concepts of the genuinely continuous curves of R^2 , or concepts of the curves of Q^{*2} (which, though merely dense, include points absent from those of Q^2)—cannot be derived from experience.

give what we would now consider definitions of denseness. Why did they make this mistake?

We might first ask whether any mistake was actually involved in these definitions of continuity. For one could suggest that the concept of continuity employed prior to Dedekind, lacking any specific formal articulation, was simply indeterminate as between (what we now call) denseness and (what we now call) continuity. So, the thought goes, it was open to theorists to precisify the notion using whatever (internally consistent) formal definition they pleased.

But this seems wrong. Dedekind clearly takes himself to be offering the *proper* definition of the intuitive notion of continuity that we associate with geometrical lines, in contrast to the *distinct* property instantiated by the rational numbers (denseness). Once the two notions had been formally distinguished, the term “continuity” was attached specifically to one, rather than the other. And this does not seem to have been an accident. As Dedekind (1901, p. 5) says, his work was motivated by a

comparison of the domain [Q] of rational numbers with a straight line, [which] led to the recognition of the existence of gaps, of a certain incompleteness or discontinuity of the former, while we ascribe to the straight line completeness, absence of gaps, or continuity.

Earlier mathematicians, then, in failing to give a definition that distinguished the continuity of a line from the mere denseness of the rational numbers, were indeed making an error: they were failing to capture the intuitive notion of continuity.¹³ And so we might ask what could explain their error.

Here is a (somewhat speculative) proposal: The error was due to over-reliance on visual thinking. In their search for a characterization of our intuitive idea of a maximally gapless line—the intuitive notion of continuity at work in Euclidean proof, which Dedekind initially explicates by pointing to *geometrical* concepts¹⁴—

¹³ In response to the idea that there is no saying of a mathematically well-defined notion like denseness whether it “succeeds” as a definition of an intuitive concept like continuity, Maudlin (2014, pp. 2-3) writes: “[A] mathematical subject... is devised in the first place to capture, in a clear and precise language, certain informal concepts already in use. It is only because we begin with some grasp of a subject like geometrical structure that we seek strict definitions in the first place. Those formalized definitions can do a better or worse job of capturing the informal concepts whose names they inherit.” This expresses quite perfectly my point about the concept of continuity, and the failure of the attempts, prior to Dedekind, to formally define it.

¹⁴ Dedekind simply takes for granted that the common conception of straight lines involves thinking of them as genuinely continuous – as more “complete” than the dense set of rational numbers. This supports the suggestion I made above, in fn. 6, that it is reasonable to suppose that Euclid conceived of his lines and curves as figures in a genuinely continuous plane, rather than as constructed in the merely dense Q^* . It is not Q^* that captures our

mathematicians focused (implicitly) on what it would take for a line to be maximally *visually* gapless. Since, as argued above, *denseness* is what is required to meet this specifically visual condition, theorists mistakenly believed that, in coming up with a characterization of denseness, they had hit on a definition of the more demanding notion of continuity. Conversely, Dedekind's *success* in articulating a proper definition of continuity might be attributable, in part, to his determination to capture the continuity of the real numbers in a way "internal" to arithmetic, without recourse to any visual thinking.

The desire (stated explicitly by Dedekind) to divorce arithmetical analysis from its geometrical origins was driven in part by the idea that geometrical reasoning stemmed from experience. Since mathematics was supposed to be the domain of non-experiential thought, geometry's alleged experiential source excluded it from mathematics proper. But, if what I have been arguing is correct, mathematicians need not have feared the intrusion of geometry into their mathematical reasoning. For our geometrical concepts—such as the concept of a continuous curve at work in Euclid I.1—are not in fact derived from experience.

Why, then, was geometry seen as problematically dependent on experience? What I want to suggest is that cases like that with which I began—the case of the carpenter, who takes her geometrical knowledge to apply directly to the objects of her experience—misled theorists into assuming that our geometrical concepts must be *derived* from experience. What such cases do reveal is that there is a tight connection between our experience and our geometrical reasoning. But that connection is not a matter of experience serving as the source of geometrical concepts. Rather, such concepts, though *a priori* in origin, do *feature in* experience: they are constituents of the contents of our perceptual states.

What I mean by this is the following: When we perceive an object as a right triangle, we represent it as instantiating the very property about which we reason when we perform the proof of the Pythagorean theorem. This concept is not derived from such experiences – like the concept of a continuous curve needed to ground Euclid I.1, the geometrical concept of a (continuous) straight line involved in the proof goes beyond anything our experience could supply. But our experience is one way for the

intuitive conception of the "completeness, absence of gaps, or continuity" of lines; it is the *real* numbers, \mathbb{R} , as defined by Dedekind. And so it seems plausible that it is a notion of genuinely *continuous* lines that is at work in Euclid's proofs. Thus, I find myself in (partial) disagreement with Friedman (1992, p. 60), who writes that "the intuitive notion of 'continuity' figuring in [Euclid's] Postulate 2 is not our notion of continuity: in particular, it is not explicitly distinguished from mere denseness." Euclid certainly does not *explicitly* distinguish his notion of continuity from mere denseness; but Dedekind's explicit definition is simply a way of capturing what is contained in the intuitive notion of continuity—of "completeness" that goes beyond mere denseness—already at work in Euclid. So I think it fair to say that, in an important sense, Euclid's intuitive notion of continuity *is* our notion.

a priori concepts we utilize in Euclidean proof to be deployed. These concepts are read *into* our experiences, not read *off* of them.

This mirrors the role that our concepts play in the practice of Euclidean proof. As emphasized above, we cannot get onto the idea that a construction of a circle results in a continuous curve from observing (or imagining) the process of construction. Instead, we *conceive of* the constructive procedure *as* producing a continuous curve. This requires utilizing our concept of continuity – it requires building into the practice certain *rules* of construction, which utilize *a priori* conceptual resources. Euclidean proof, then, is a system of geometrical construction *guided by* our concepts of continuous lines and curves – concepts that we do not derive from experience. Recent work on geometrical reasoning has rightly emphasized that we misunderstand the practice of Euclidean proof if we try to see it as a modern formal system—a version of Einstein’s “axiomatics”—devoid of any non-logical content. But such work has misidentified the source of the extra-logical content involved in the practice. The concepts at work in Euclidean proof are not visual (or imaginative, or more broadly experiential). They are *a priori*.

In closing, I want to summarize what I’ve argued, and where my investigation has left us. I began with the parable of the carpenter, which illustrates that there are important connections between spatial experience and Euclidean proof. I went on to consider how we should understand Euclidean proof itself – what kind of cognitive practice is it? In answering this question, I pointed out that Euclid’s proof of I.1, when interpreted purely formalistically, contains a gap. I then suggested that this gap cannot be plugged by anything in the visual realm: the needed distinction between continuity and mere denseness cannot be a picturable meaning, nor supplied by a diagram, nor derived from the ways things look. Finally, I considered whether we could derive the concept of a continuous curve as the *limit* of a visual concept, and I argued that, in the realm of experience, such a limit procedure would get us only to denseness, not full continuity.

From these observations, I concluded, first, that we have *substantive* geometrical concepts – ones whose content is more than merely formal; and, second, that the content in question is *a priori* – that is, the concepts involved are not derived from experience. The question this leaves us with is how these contentful, *a priori* concepts are connected to experience, in cases like that of the carpenter. The answer I have suggested is that our *a priori* concepts partly constitute the contents of our perceptual states. In the next chapter, I will elaborate on and defend this proposal.

CHAPTER 3

A PRIORI CONCEPTS IN SPATIAL EXPERIENCE

In Chapter 1, I proposed a solution to the PROBLEM OF INTERNAL ACQUAINTANCE—a way of accounting for the grasp we have of the specific contents of our sensory experience—which relied on the claim that we have *a priori* spatial concepts that feature in the contents of perception. Defending this claim required two things: establishing that we do indeed *have* spatial concepts that are *a priori*; and arguing that those very concepts feature in the contents of perceptual experience. The previous chapter addressed the first of these tasks: I argued that a close analysis of the practice of Euclidean proof reveals that our geometrical concepts must be *a priori*. In this chapter, I turn my attention to the second task.

Although the concepts we employ in Euclidean proof are not derived from experience, there is nonetheless an important connection between spatial experience and these *a priori* concepts. We automatically apply our *a priori* spatial concepts to the objects we perceive, as in the case of the carpenter from the Introduction. This is the phenomenon of TRANSFER. In the first half of this chapter, I argue that in order to account for TRANSFER, we must acknowledge that the very concepts we use in Euclidean proof—our *a priori* concepts of spatial properties—feature in the contents of our perceptual experiences. This is the thesis of COMMONALITY.

In the second half of the chapter, I provide further support for COMMONALITY by offering an account of spatial experience that explains how the very same set of *a priori* concepts could show up both in Euclidean proof and in perceptual experience. I argue that we have an innate, primitive grasp of basic spatial structure—a set of “proto-concepts”—that features in our cognitive lives in two ways. On the one hand, these proto-concepts are deployed in our experience of the world around us, representing, in a qualitatively rich format, the presence of particular spatial properties. On the other hand, with the acquisition of more sophisticated cognitive capacities, such as linguistic communication, these same spatial proto-concepts are built up into full-blown geometrical concepts, which allows for their explicit articulation in the practice of Euclidean proof—a practice through which we explore in detail the spatial features that our innate proto-concepts represent.

I want to begin by returning to the example of the carpenter who, having worked out a proof of the Pythagorean theorem, immediately applies the theorem to the wooden beams she sees in her workshop. This case illustrates a general feature of our conceptual scheme—a fact of descriptive metaphysics—which we can label TRANSFER:

(TRANSFER) We take the results of Euclidean proof to be applicable to the objects we perceive.

In the previous chapter, I argued against two interpretations of Euclidean proof, each of which would have offered a straightforward explanation of TRANSFER. First, if Euclidean proof were indeed an “axiomatic” system in Einstein’s sense—a system of pure deductive logic—the application of Euclid’s theorems to experience would be explicable in something like the way the logical positivists (and Einstein himself) recommended. On this kind of positivist picture, we find in experience a suitable *interpretation* of the logical relations and primitive terms of the formal language of Euclidean proof.¹ And so, in charting out the logical consequences of the axioms through the (allegedly) deductive proofs of Euclid, we draw conclusions that are applicable to the objects of experience, because those objects serve as one model (perhaps among many) for the formal system.

Alternatively, if the concepts employed in Euclidean proof were themselves empirically-derived, the application of Euclid’s theorems to the empirical world would be entirely unsurprising. On this picture, geometrical reasoning would itself be just a way of thinking about the objects we perceive and their properties, so the theorems, in the first instance, would concern such empirical objects. Applying those theorems in experience, then, would constitute no “extra” step at all.

But, having rejected each of these interpretations of Euclidean proof, I am left with something of a puzzle. Why is it that the carpenter automatically takes the Pythagorean theorem to apply to the physical objects she sees in her workshop, if, as I’ve suggested, the reasoning that she does in proving the theorem utilizes a set of concepts that are neither empirical nor merely formal? On my picture, Euclidean proof would seem to concern a set of *abstract* objects: the geometrical figures constructible in the Euclidean plane, which (unlike physical or “phenomenal” objects) genuinely instantiate the “perfect” geometrical properties (such as full continuity) that I argued were necessary to ground Euclid’s proofs. So why should we take our conclusions about these abstract objects to apply to the distinct category of objects we encounter in perception? On my picture, TRANSFER stands in need of an explanation.

My explanation is this: in experience, the empirical world and its objects are presented as instantiating some of the very geometrical properties we attribute to the abstract figures we reason about in *a priori* Euclidean proof. A collection of wooden beams can be seen *as* instantiating right triangularity – the property about which we reason in utilizing our *a priori* concepts to prove the Pythagorean theorem. In the case of our perceptual representations of spatial properties, unlike in our abstract geometrical reasoning, these geometrical concepts are deployed in virtue of the qualitative character of the mental state in whose content they feature. This does not mean that the “visual image” associated with such a perception specifically picks out “perfect”

¹ This positivist picture is articulated nicely in Strawson (1966).

geometrical properties – continuous, rather than merely dense figures, for example. As I argued above, nothing *imagistic* could do so. Instead, the wooden beams show up in the carpenter’s experience as instantiating right triangularity—a property of which we have a precise, *a priori* geometrical concept—to a certain degree of *approximation*.

I will say more about the details of this picture of perceptual experience in §3.4. But first, I want to defend the claim—which I label COMMONALITY—that experience really does represent the presence of the very features about which we reason in Euclidean proof. For this claim should seem surprising: it suggests that we represent the physical (or phenomenal) objects presented in perceptual experience as literally sharing properties with the abstract geometrical figures of Euclidean proof. And it might seem odd that our experience would represent the concrete objects of the empirical world as instantiating the same kinds of properties that are instantiated by abstract objects. So, we might think, COMMONALITY is an implausible thesis, and we should seek to explain TRANSFER in some other way.

One such alternative explanation begins with a thesis that I will label STRUCTURALISM, according to which we represent the objects of experience not as instantiating the same geometrical properties as abstract Euclidean figures, but merely as *structurally isomorphic* to such abstract entities. According to the STRUCTURALIST, it is this isomorphism that explains TRANSFER: we apply the theorems of Euclidean proof in the domain of experience because the objects in the two domains have a certain structural similarity, not because they share a set of substantive properties.

Below, I will argue that STRUCTURALISM is inadequate to explain TRANSFER. For on the STRUCTURALIST picture, objects would show up in experience only as mapping onto, or representable by, or analogous to, the geometrical figures of Euclidean proof. But objects like the carpenter’s wooden beams don’t show up in experience as merely isomorphic to the geometrical figures of Euclidean proof; they do not show up as possessing spatial properties in a merely metaphorical sense. In experience, objects are represented as literal *instances* of the geometrical figures of Euclidean proof – as lines, triangles, and circles. It is only by acknowledging this more direct connection between experience and *a priori* geometrical reasoning that we can explain how TRANSFER occurs.

3.1. LITERAL AND METAPHORICAL SPACES

To begin, it will be helpful to make a distinction between what we might call literal (or geometrical) and metaphorical spaces. The significance of this distinction has been highlighted by Tim Maudlin:

In the right context, almost any collection of objects can be considered to form a “space”. For example, if one is studying Newtonian mechanics... it is natural to speak of “the space of solutions” of Newton’s

equations of motion. Each “point” in this space, each individual element, describes the motions of a set of particles governed by Newtonian gravity. There is an intuitive sense—which can be made technically precise—in which the various solutions can be “closer” or “farther” from one another, and hence an intuitive sense in which the whole set of solutions can be thought of as having a “geometry”. But this sort of talk of a “space” is evidently not literal. This “space” is, in an obvious sense, a metaphorical space; it is just a way of talking about the solutions and a measure of similarity among them....

In contrast, consider Euclidean space, the subject matter of Euclidean geometry. Euclidian space is an abstract object in the way that all mathematical objects are abstract. But Euclidian space is not just metaphorically a space. When we say that one point in Euclidian space is “closer” to another than it is to a third, we are not suggesting that the first point is more similar to the second than to the third in any way. Indeed, intrinsically the points of Euclidian space are all exactly alike: they are all, in themselves, perfectly identical. The points of Euclidian space, unlike the “points” of the space of solutions to Newton’s equations, really are points: they have no internal structure. The “points” of the space of solutions form a (metaphorical) “space” only because they are highly structured and different from one another. The points of Euclidean space, being all intrinsically identical, form a space only because of structure that is not a function of their intrinsic features. Euclidian space is therefore an instance of what I mean by a geometrical space.²

As Maudlin notes (and laments), the term “space” has come to be used very broadly: we can, for example, talk about “color space,” “the space of possible worlds,” or the “phase space” of a dynamic system. Used in this way, the term “space” simply denotes a way of describing a set of objects and their relations. This way of talking can be useful because there may be structural similarities between, say, the set of state’s a system can be in and a Euclidean plane; or between a region of Euclidean three-space and the degrees of brightness, saturation, and hue of various colors. Such structural similarities allow us to use spatial structures as useful *representations* of various non-spatial features. But it is important not to read too much into such representations. It is important to distinguish merely metaphorical from genuinely geometrical spaces.

In his discussion, Maudlin at first seems content to let his examples of geometrical and metaphorical spaces speak for themselves, since, he notes, it is often counter-productive to try to characterize such intuitively clear distinctions more explicitly. He writes:

² Maudlin (2014, pp. 6-7).

Attempting to give explicit definitions is a dangerous business. So I implore the reader first to reflect on the particular examples I have just given to understand how I mean to use “geometrical” and “metaphorical” when characterizing spaces. There is an evident difference between Euclidean space and the “space” of solutions to Newton’s equations, and it is this difference I mean to mark.³

I share Maudlin’s wariness about offering explicit definitions, and I agree that there is an intuitively obvious sense in which the “space” of solutions to Newton’s equations is not literally a space, while Euclidean space is. But it would be helpful to have a somewhat more concrete characterization of the distinction between geometrical and metaphorical spaces. And Maudlin, in spite of his wariness about explicit definitions, does give some indication of how the two kinds of “space” differ. In the case of metaphorical spaces, he writes, the structure of the “space”—the set of spatial relations, like “distance,” that the elements bear to one another—is determined by the intrinsic features of the elements (e.g., the individual solutions to Newton’s equations) that make up the space, and how those features differ from element to element. By contrast, in the case of geometrical spaces, the individual elements (the points) lack any distinguishing intrinsic features at all (they are all intrinsically identical), so the structure is not a function of any such intrinsic features. This suggests a possible way to define geometrical, as opposed to metaphorical, spaces:

(Def₁) A space is a geometrical space iff its elements are all intrinsically alike.

But Maudlin denies that Def₁ is a good way to mark the distinction. Here is his reasoning:

[Def₁] fits the mathematical spaces commonly studied under the rubric “geometry”. But in addition to these mathematical spaces, we also want it to turn out that physical space—the space (or more properly spacetime) that we actually inhabit, the space we walk around in—counts as a geometrical space. But the points of that space are not all intrinsically alike. For example, some points may be occupied by matter, or by certain fields.⁴

Maudlin’s reason for rejecting Def₁ is somewhat idiosyncratic. He rightly suggests that physical spacetime should (at least potentially) count as a geometrical space. But he thinks Def₁ will not give us this result, because the points of physical spacetime “are

³ Maudlin (2014, p. 7).

⁴ Maudlin (2014, p. 7).

not all intrinsically alike,” in that some may be occupied by matter or fields (while others are not). This suggests that Maudlin takes features like “being occupied by matter or fields” to be *intrinsic* features of the points of spacetime. But we might think that, while spacetime points do differ in terms of what (if anything) occupies them, those differences are not *intrinsic* to the points. Whether matter occupies a given point of spacetime would seem, *prima facie*, to be a *relational* feature of that point – something determined by how the point is related to the matter or fields that occupy it, not by how the point is in itself. So, we could count physical spacetime as a geometrical space according to Def₁ after all, if we insist (plausibly) that the kinds of distinguishing features Maudlin points to are *non-intrinsic* features of spacetime points.

But this is a contentious claim about the nature of spacetime points in general relativity; I don’t want to insist that Maudlin is wrong to take those points to have some intrinsic features. And, in any case, I’m not sure that we should rule out the possibility of there being some geometrical space—even if it is not our physical spacetime—whose points differ intrinsically from one another. So it is worth considering Maudlin’s modified proposal about how to distinguish geometrical from merely metaphorical spaces, which is meant to allow for such spaces to count as geometrical. Maudlin writes that in the case of physical spacetime, while the points of the space do (by his lights) differ intrinsically:

those differences of material content do not analytically determine the geometry of the space in which we live. Simply expressed, all the facts about the intrinsic structure of two points in physical spacetime do not determine their geometrical relation to one another. In the argot of philosophy, the geometrical structure of real spacetime does not supervene on the intrinsic features of the points of spacetime. So physical spacetime—the spacetime in which we live—is a geometrical space.⁵

We can extract from this passage a modified proposal about what it takes to be a geometrical space:

(Def₂) A space is a geometrical space iff its structure does not supervene on (is not analytically determined by) the intrinsic features of its elements.⁶

According to Def₂, the “space” of solutions to Newton’s equations (where the “distance” between any two “points”—i.e., any two particular solutions—is a function of

⁵ Maudlin (2014, p. 7).

⁶ While Maudlin, perhaps owing to his wariness about giving explicit definitions, never endorses Def₂ as a definition of a geometrical space, what he says seems consistent with taking it as such.

the internal structure of the two solutions) and color “space” (where “distance” is determined by the intrinsic hues, saturations, and brightnesses of the colors themselves) are not geometrical spaces, while both abstract mathematical spaces like Euclidean space (where the points have no internal differences at all on which the structure might depend) and physical spacetime (where the points (arguably) have intrinsic features, but where those features don’t themselves determine the structure of the space) do count as genuinely geometrical.

It seems to me that this way of drawing the distinction between geometrical and metaphorical spaces is on the right track; but I think Def₂ doesn’t quite work as it stands.

To see why, we can first note that any particular space, whether geometrical or metaphorical, has two components. First, there is a collection of points, or basic elements; second, there is a spatial *structure*, consisting in a set of spatial relations (such as distance) that hold among the points. The question of whether a given space is geometrical or metaphorical will, in some fashion, turn on the nature of these two components, and how they are related. Def₁ attempts to cash out the distinction purely in terms of the nature of the basic elements; Def₂ defines the distinction in terms of how the intrinsic features of the basic elements are connected to the spatial relations. But, I want to suggest, neither of these definitions gets us to the heart of the distinction, because it is the nature of the structure-constituting *relations* of the space, not anything in particular about the *intrinsic* features of the points, that separates literal geometrical from merely metaphorical spaces. The question is whether a relation like “distance,” in the context of a given space, is a genuinely geometrical relation, or if it is instead merely a way of representing some other kind of relation. By seeing where Def₂ goes wrong, we can formulate a better definition of geometrical spaces, which turns on this crucial question.

Def₂, I claim, gives us a way to distinguish a *subset* of the metaphorical spaces, which we might call “intrinsic similarity spaces,” from geometrical spaces. An intrinsic similarity space is a metaphorical space whose structure is determined by the intrinsic features of its individual elements. In the case of such metaphorical spaces, “spatial” relations like distance are used to represent relations of similarity among the elements that obtain in virtue of those elements’ intrinsic features.

But intrinsic similarity spaces are not the only metaphorical spaces. There are other “spaces” that are clearly metaphorical, but which have structures that are not determined by the *intrinsic* features of their elements. An example of the kind of “space” I have in mind is what we might call “kinship space.” Kinship space tracks the kinship relations among a set of biological individuals – a family of humans, say. Charting out kinship space is what we do when we draw family trees. In this “space,” a given person is one unit of “vertical distance” away from each of her parents and each of her children, and one unit of “horizontal distance” away from each of her siblings. The total “distance” between any two individuals in kinship space is determined by how many relations of each kind (“vertical” links of parenthood, “horizontal” links of

siblinghood) stand between those individuals. Thus, I am two vertical units (and no horizontal units) away from my maternal grandmother, one horizontal unit (and no vertical units) away from my brother, and three vertical units and one horizontal unit away from my first cousin once removed (two units “up” to my grandmother, one unit “over” to her sister, and one more vertical unit “down” to *her* daughter).

Kinship space may be useful for representing the degree of genetic relatedness between individuals, or for other purposes. But it seems clear that kinship space is not a geometrical space. As with color “space” and the “space” of solutions to Newton’s equations, kinship “space” is only metaphorically a space. Saying that I am “closer” in kinship space to my brother than to my cousin is just a way of talking about the kinds of kinship relations I bear to each – it is a way to use spatial concepts to describe or represent non-spatial features of a set of individuals. And yet, according to Def₂, kinship space counts as a geometrical space: the intrinsic features of its “points” (the individual people) do not analytically determine its structure.

To see this, suppose that I have two relatives, an uncle named Jon (the brother of my mother) and a first cousin twice removed named Steve (the son of my great grandmother’s sister). Now, it could turn out, through some bizarre accident of genetics and upbringing, that Jon and Steve are intrinsically identical – they could be molecule for molecule duplicates. But that would not mean that they were the same “distance” from me in kinship space: Jon would still be one vertical and one horizontal unit from me, while Steve would be three vertical and two horizontal units from me. Kinship relations don’t supervene on the intrinsic features of the individuals who stand in those relations, so the structure of “kinship space”—which simply charts kinship relations—isn’t determined by the intrinsic features of its “elements.”

There are many other examples of metaphorical spaces whose structure is not determined by the intrinsic features of their elements. To cite just one: in decision theory, we can describe a subject’s “preference space,” in which the individual elements are various possible states of the world and the “spatial” relations between the elements are determined by the subject’s preferences over those states (this amounts to mapping a subject’s utility function to an oriented line). A subject’s preferences over various states of the world need not track any intrinsic similarities between those states: for instance, a subject might prefer A to B and B to C (making A “closer” to B than to C in preference “space”), even though A and C are more intrinsically similar than A and B. So the structure of this space will not supervene on the intrinsic features of its elements. Still, such a “preference space” would clearly be only metaphorically a space.

What these examples are bringing out is that some metaphorical spaces—call them “relation spaces”—use a set of spatial relations (like distance and direction) to represent a set of non-spatial, but at the same time irreducibly *relational* features (i.e., relations that don’t themselves supervene on the intrinsic properties of their relata, such as kinship relations or preference relations). What makes the spaces so constructed merely metaphorical is that the relations that determine the structure of a

given “relation space” are not, in the first instance, spatial: relations of kinship and preference simply do not in themselves have anything to do with spatial relations like distance or direction. The information encoded by the structure of kinship “space”—just like the information encoded in an intrinsic similarity space, like the “space” of solutions to Newton’s equations—can be completely specified without making use of any spatial notions at all: someone who knew, of all the people in the “space,” whose parents were whose, would thereby know all the information represented by the “distance” and “direction” relations. If such a person had no capacity for metaphorical thought, and so could not make sense of the idea that I am “closer” to my brother than to my third cousin, that person could still know everything there is to know about the relevant kinship relations (she could, for example, know that more instances of the “parent of” relation need to be invoked to explain the familial relation I bear to my second cousin than are needed to explain the familial relation I bear to my brother).

With Euclidean space, by contrast, there is no set of non-spatial relations among the elements of the space, nor any intrinsic features of the elements themselves, from which the structure of the space can be determined: aside from spatial relations like distance and betweenness, the points of Euclidean space don’t stand in any relations at all. Even in the case of geometrical spaces whose points *do* stand in various non-spatial relations (the way the points of space-time in general relativity stand in relations like “is occupied by greater mass-energy than”), the structure of the space does not *supervene on* these non-spatial relations.⁷ What marks off genuinely geometrical spaces from merely metaphorical ones, then, can be defined as follows:

(Def₃) A space is a geometrical space iff its structure does not supervene on (is not analytically determined by) *any* set of non-spatial features of its elements, including both intrinsic features of the elements and irreducible, non-spatial relations among them.

As Maudlin says, attempting to give explicit definitions is a dangerous business. The way of distinguishing metaphorical from geometrical spaces I have just suggested is, in an obvious sense, circular: it defines what counts as a genuine geometrical space in terms of whether any non-*spatial* properties or relations analytically determine the structure of the space. So we would seem to need an antecedent grasp of what counts as a non-spatial feature (and thus, a grasp of what counts as a spatial feature) in order to understand this definition.

⁷ As Maudlin (2014, p. 7, fn. 4) notes, in general relativity, the mass-distribution in the neighborhood of a point does *causally* determine some of the spatial features of that point (namely, the degree of curvature in its neighborhood). But the spatial features are not *analytically determined* by the mass distribution, and, in any case, many other spatial features—like the distance between two points—are not even causally fixed by the mass distribution.

But this circularity is, it seems to me, quite harmless. My aim was to fill out a bit the difference—a difference that should in any case be quite intuitive—between genuine geometrical spaces and cases where we use geometrical terms to represent or map out certain non-spatial features of a set of objects. My discussion has suggested that spaces can be divided into three kinds, one literal and two metaphorical: there are (metaphorical) intrinsic similarity spaces, (metaphorical) relation spaces, and (literal) geometrical spaces. To categorize a given space, we can consider one of its basic spatial relations, like distance, and ask: In virtue of what do the distance relations among the elements hold? In the case of (metaphorical) intrinsic similarity spaces, the “distance” relation is a measure of the intrinsic similarity of the elements; thus, distances in such spaces are determined by the intrinsic features of the elements. In the case of (metaphorical) relation spaces, the “distance” relation is a way to represent some set of non-spatial relations (like kinship or preference) that do not supervene on the intrinsic features of the elements; thus, distance in such spaces is not determined by the elements’ intrinsic features, but it is determined by non-spatial relations among its elements. Finally, in the case of (literal) geometrical spaces, the distance relation is not a function either of the intrinsic features of the elements, or of any set of non-spatial relations the elements stand in; in such spaces, the distance relation is a *sui generis*, essentially *geometrical* relation that itself determines the structure of the space.

3.2. COMMONALITY

The reason I have gone on at such length about the distinction between (literally) geometrical and (merely) metaphorical spaces is that having the distinction firmly in mind will help clarify how I intend COMMONALITY to be understood, how COMMONALITY explains TRANSFER, and how that explanation differs from the alternative picture provided by STRUCTURALISM.

We might formulate an initial statement of COMMONALITY as follows:

(COMMONALITY) We have a set of basic spatial concepts—concepts of, e.g., straight lines, distance, and triangularity—that feature in the contents of both our *a priori* geometrical reasoning and our perceptual experience.

So, according to COMMONALITY, we represent two things as spaces: the abstract mathematical object—Euclidean space—that is the subject matter of Euclidean proof; and the empirical world around us, in which we perceive objects like wooden beams. Having distinguished literal geometrical spaces from merely metaphorical ones, we can now ask a further question about how COMMONALITY is to be interpreted: do we represent these domains as literal or metaphorical spaces?

It seems clear that we represent Euclidean space as a literal space. Its points lack any intrinsic features and bear no non-spatial relations to one another, so the only

available conception of it we have is as a literally geometrical space. Indeed, Euclidean space is the paradigm of a literally geometrical space.

The more substantive question concerns the way in which we represent physical space when we perceive the world around us. As Maudlin notes, the points of physical space(time), unlike the points of abstract mathematical spaces like Euclidean space, do have many non-spatial features. They are, for example, occupied by various kinds of matter. And, perhaps more relevantly (as I will discuss below), they have various relational features, such as being causally connected to each other (or, more precisely, being such that matter occupying one point has causal bearing on matter occupying another). So it is not out of the question that we might represent physical space as a metaphorical space: perhaps we represent its elements as having various *non-spatial* features, in virtue of which those elements can be said to form a “space.”

But in endorsing COMMONALITY, I mean to be ruling out just this possibility: I claim that physical space is represented in experience as a literal geometrical space – a manifold of points that bear various spatial relations to one another, where those relations do not supervene on any *non-spatial* features the points are represented as having. So, to be more precise, we can reformulate COMMONALITY as follows:

(COMMONALITY) We have a set of basic, *literal* spatial concepts—concepts of *literal geometrical* features, such as straight lines, distance, and triangularity—that feature in the contents of both our *a priori* geometrical reasoning and our perceptual experience.

It is this formulation that I will defend.

A few notes. First, COMMONALITY is, crucially, not a *metaphysical* claim about the nature of the world that we live in, the world we encounter in experience. It is a claim about the *way that we perceptually represent* that world, or, equivalently, a claim about how the world *would* have to be, in order for our experience to be veridical. On its own, COMMONALITY does not amount to the metaphysical claim that the world really is a geometrical space. Such a metaphysical claim is denied by some views in contemporary philosophy of physics: on some interpretations of quantum mechanics, for example, the physical world is held to lack any genuine spatial features.⁸ These interpretations of contemporary physics are certainly controversial. But what I want to emphasize is that COMMONALITY says nothing one way or the other about whether the world we live in *actually is* a geometrical space. Instead, COMMONALITY is a claim about how we *represent* that world in experience – it is the claim that we perceive the empirical world *as* a literal geometrical space, with objects that have literally geometrical properties of the kind we attribute to abstract figures in doing Euclidean proof.⁹

⁸ See, e.g., Ney (2012).

⁹ Combining COMMONALITY with certain *additional* theses would yield a claim about the nature of the world itself. On certain externalist views of perception, for instance, the representational

Second, in formulating COMMONALITY in this way, I intend to draw a contrast with the STRUCTURALIST picture I mentioned in the Introduction. For STRUCTURALISM is compatible with—indeed, is almost interchangeable with—a different, weaker reading of COMMONALITY, on which the “basic spatial concepts” posited by COMMONALITY are taken to be merely *metaphorical* spatial concepts—concepts applicable to anything that counts as a metaphorical space. On such a reading, what COMMONALITY tells us is just that, in experience, we represent a set of elements as standing in various relations that can be usefully described using spatial terms, where those relations are not themselves genuinely spatial. The idea would thus be that we do represent the abstract figures of Euclidean proof and the concrete objects of the perceptible world as sharing properties, but only in the weak sense that they share the kinds of *structural* properties that allow us to speak of, say, the colors, or a family of organisms, as forming a “space.” So, interpreted in this weaker way, COMMONALITY would just be a version of STRUCTURALISM: it would amount to the claim that experience presents the world as structurally isomorphic to the literal geometrical space we reason about in Euclidean proof, while denying that experience represents the world as itself literally spatial.

I will argue that such a STRUCTURALIST thesis should be rejected, by showing that the kind of explanation it offers of cases like the carpenter’s is inadequate. Thus, my argument for COMMONALITY (formulated in the more demanding, literal way) will be an inference to the best explanation: I will argue that COMMONALITY provides a satisfactory explanation of TRANSFER, while STRUCTURALISM does not.

This argumentative strategy might seem rather limited, as it can (at best) establish that COMMONALITY offers a better explanation of TRANSFER than one particular competing explanation – namely, STRUCTURALISM. Surely, one might think, there are other alternatives, and arguing that STRUCTURALISM fails does nothing to show that we should accept COMMONALITY, rather than some third account of TRANSFER. But I think that undermining STRUCTURALISM takes us a bit further. For, given what was shown in the previous chapter, TRANSFER stands in need of a certain kind of explanation. We take the results obtained in Euclidean proof—a practice that, as I argued, makes use of a set of substantive spatial concepts, of the kind that characterize literal geometrical spaces—to be applicable to the objects we encounter in experience. This application of our abstract mathematical reasoning to the objects of experience is a

content of experience depends on the actual character of the world represented. On such a view, our experience could only represent the world as literally spatial if the world were indeed literally spatial. And, more broadly, combining COMMONALITY with a claim of experiential *veridicality*—a claim to the effect that our spatial experience accurately represents the external world—yields the result that the world is indeed literally spatial. I will have much more to say about how COMMONALITY relates to these additional theses in later chapters. But for now, I want to remain neutral on such questions: my arguments here are intended to show that our experience represents the world as a literal geometrical space, whether or not such experiential representations are veridical.

striking phenomenon.¹⁰ And so we need something we can point to in our conceptual scheme—and in particular, something about the way in which we represent the objects of experience—that *explains* how TRANSFER occurs.

Furthermore, the explanation in question must explain facts of the following sort: we apply the Pythagorean theorem only to a specific subset of the objects we encounter in experience—those we would describe as right triangles—while instead applying the formula for calculating the area of a rectangle to other empirical objects. What these facts suggest is that we *systematically map* the various substantive geometrical concepts of Euclidean proof onto the objects we perceive. This could occur, it seems to me, in one of two ways. On the one hand, TRANSFER could occur because, as COMMONALITY holds, we represent the objects of experience as instantiating the same literally geometrical properties that we reason about in Euclidean proof. The “systematic mapping” would then be explained as simply a case of applying a concept to an object represented as an instance of that concept. Or, alternatively, the systematic mapping could occur in a less direct way: we might make use of some shared *structural* features in mapping our conclusions about literal Euclidean space onto a set of not-literally-spatial features our experience represents.

I cannot definitively rule out that there might be some third way to explain TRANSFER. But it seems to me that these really are the two options on the table: either COMMONALITY is true, and we achieve TRANSFER directly, by representing the objects of experience as themselves literally spatial; or STRUCTURALISM is true, and we represent the objects of experience as not *literally* spatial, but as having some set of features that shares enough structure with the literal space of Euclidean proof to allow for a systematic mapping from one domain to the other. So, in arguing that this weaker, less direct kind of link is inadequate to explain TRANSFER, I will take myself to be offering very strong reason to accept COMMONALITY.

3.2.1. *Explaining Transfer via Commonality*

The argument I will be making in support of COMMONALITY is, as I’ve said, a kind of argument to the best explanation. So I begin by giving a rough sketch of the explanation of TRANSFER that COMMONALITY provides, focusing on the specific case of the carpenter.

First, the carpenter does some abstract mathematical reasoning and reaches the conclusion:

- (p) The area of the square on the hypotenuse of any right triangle is equal to the sum of the areas of the squares on each of its legs.

¹⁰ In the quotation with which I began this dissertation, it is this phenomenon that Einstein finds so remarkable.

Next, the carpenter enters her workshop and sees an object composed of wooden beams (call it T). The content of her visual experience includes the following proposition:

(q) T is a right triangle.

From p and q , the carpenter concludes that:

(c) The area of the square on the hypotenuse of T is equal to the sum of the areas of the squares on each of its legs.

Given her knowledge of the lengths of the two leg beams and some basic arithmetical proficiency, the carpenter is then able to calculate that the hypotenuse is five feet long.

In the course of her reasoning, the carpenter moves seamlessly from her belief in a proposition (p) arrived at through *a priori* Euclidean proof, to a conclusion (c), via a premise (q) that is the content of a perceptual experience. The reasoning here need not be carried out *explicitly* by the carpenter; indeed, the transition to c from p and q is quite automatic. But the set of moves described above does seem like a natural way to capture how TRANSFER occurs. And a key feature of this explanation is that the carpenter's two initial mental states—her mathematically-derived belief that p and her perception that q —have contents that involve the same concept: it is because a single concept, RIGHT TRIANGLE, is a constituent of the contents of both p and q that the carpenter takes herself to be licensed to *combine* them (via a simple application of universal instantiation) in drawing an inference about the length of the third wooden beam. Again, it seems clear that we represent the subject matter of Euclidean proof as a literal geometrical space. So, in order for the concept RIGHT TRIANGLE to feature in the contents of the conclusions we draw in doing Euclidean geometrical reasoning, that concept must be a concept of a literal spatial property. What COMMONALITY adds to this picture is that the very same literal spatial concept also features in the content of the carpenter's perceptual experience q , thereby explaining the transition from p and q to c . This is how COMMONALITY allows us to make sense of TRANSFER.

3.3. STRUCTURALISM: AN INADEQUATE EXPLANATION OF TRANSFER

STRUCTURALISM denies that perceptual experience represents the presence of literal spatial features like right triangularity. In particular, STRUCTURALISM denies that T is represented as instantiating literal right triangularity. So the explanation of TRANSFER given above is unavailable to the STRUCTURALIST, since the instance of universal generalization involved in that explanation relies on T 's being represented as a literal right triangle – that is, it relies on COMMONALITY.

The STRUCTURALIST, then, must offer a different explanation of how the belief the carpenter forms on the basis of her abstract geometrical reasoning—the belief that p —allows her to reach the conclusion, c , about the particular object she perceives, T (an object that, according to STRUCTURALISM, the carpenter does not represent as a literal spatial object). To begin with, this explanation will need to interpret the content of c as involving a *non-literal* use of spatial terminology: when the carpenter comes to believe that “the area of the square on the hypotenuse of T is equal to the sum of the areas of the squares on each of its legs,” the spatial terms involved (“area,” “square,” etc.) must be the kinds of terms we use in describing *metaphorical* spaces, when, for example, we say that one solution to Newton’s equations (or one color, or one family member) is “closer” to a second than to a third.

Furthermore, our application of such metaphorical spatial concepts in experience must be based on our representing the elements of the “space” as having certain *non-spatial* features. In general, we can represent objects as forming a metaphorical space only if we represent them as having some other set of features. For example, we can represent the colors as forming a “space” only because we can represent their intrinsic features (hue, saturation, and brightness), and we can represent kinship “space” only because we can represent one organism as being the parent of another. In general, in order to represent a set of elements as forming a metaphorical space, we need to pick out some set of non-spatial features of those elements, in virtue of which we can metaphorically attribute spatial relations to them. On the STRUCTURALIST picture, then, there must be some set of non-spatial features we represent objects like the carpenter’s wooden beams as having, on the basis of which we represent those beams as standing in various metaphorically spatial relations (such as having “equal area” or “forming a right angle”). So the STRUCTURALIST must give an account of what non-spatial features objects are represented in experience as having, in virtue of which we represent them as metaphorically spatial, and thereby achieve TRANSFER.

I will argue that no such STRUCTURALIST explanation of TRANSFER is plausible, for two reasons. First, there is no suitable set of non-spatial features that we perceptually represent objects as having, in virtue of which we represent those objects as forming a metaphorical space – a “space” whose structure is determined by the structure of these more basic non-spatial relations. Second, even when our experience *does* represent a set of features isomorphic to Euclidean space, such structural isomorphism is not *sufficient* to induce TRANSFER. It is only when we perceive objects as having genuinely geometrical features like shapes—not when we represent them as instantiating the properties of a metaphorical “space,” like color space—that TRANSFER occurs.

At this stage, I want to set out more precisely the thesis of STRUCTURALISM, and the kind of explanation of TRANSFER it offers. First, I will clarify what the two components of the space in question—the elements and the relations—are meant to be.

Some discussions of spatial perception focus on the spatial relations that obtain among macroscopic objects, like chessboards and clock faces. And we do see such objects as standing in spatial relations to each other: the chessboard might be seen as

five feet below and slightly to the left of the clock face. So we might think that it is macroscopic objects that serve as the points of the metaphorical space of perception, on the STRUCTURALIST picture.

But the kinds of spatial features I'm concerned with—the ones that are involved in TRANSFER—are often *internal* to the macroscopic objects we perceive. The chessboard itself is seen as having certain spatial features, and our representation of these features often induces TRANSFER. For example, a chessboard is seen as square, and seeing it that way might lead us to apply the Pythagorean theorem in calculating the length of its diagonal. The chessboard's squareness can't be captured by taking the chessboard itself as a point in the space we perceive, since its shape is not determined by the spatial relations it bears to *other* objects, like the clock. So the elements of the space relevant to the phenomenon of TRANSFER are not macroscopic objects, but rather the points (or perhaps, minimal regions) of space that those objects occupy. It is a fact about the spatial structure of the points occupied by the chessboard's surface that we represent when we represent the chessboard as square, and it is this fact that induces TRANSFER.

So STRUCTURALISM should be taken as a thesis about the way we represent the space whose elements are the points occupied by objects like chessboards. The STRUCTURALIST holds that these points are experienced as forming a metaphorical space, which we can call E^* . E^* is not represented as a literal geometrical space, according to the STRUCTURALIST. But we do represent the points of E^* as having a certain structure, which we can call S . And, crucially, S is the very structure exhibited by the literal geometrical space—Euclidean space, E —that we reason about in doing Euclidean proof.

Here, then, is a statement of the thesis:

(STRUCTURALISM) The subject matter of Euclidean proof is a *literal* geometrical space—Euclidean space, E —whose points exhibit a certain structure, S . In experience, we represent the points occupied by material objects as forming a *metaphorical* space, E^* , which also exhibits S .

The STRUCTURALIST holds that it is our perceptual representation of the metaphorical space E^* , which exhibits structure S , rather than our perceptually representing a literal geometrical space of the kind we reason about in Euclidean proof, that leads to TRANSFER. In order to make this plausible, the STRUCTURALIST might note that Euclid's proofs do not rely on the specifically *geometrical* nature of Euclidean space, E , but only on the structure, S , that E exhibits. This can be demonstrated, the STRUCTURALIST claims, by Hilbert's formalization of Euclidean geometry, which shows that Euclid's theorems hold for anything that can serve as a model of the properly formalized axioms of the Euclidean system. As Hilbert points out, such a model need not comprise genuinely geometrical entities (a manifold of points); it could just as easily comprise beer mugs and chimney sweeps, so long as those objects have the right kind of formal structure to serve as a model of the formal axioms. So, the idea is that, in experience,

we represent the points occupied by material objects *not* as forming a literal geometrical space, but as exhibiting the same structure, *S*, as geometrical, Euclidean space; and, since the theorems we prove about Euclidean space hold solely in virtue of Euclidean space's exhibiting *S*, we conclude that those theorems hold for the objects of experience as well.

Recall that what makes a space a metaphorical space is that its “spatial” structure—the set of “spatial” relations, like “distance,” that hold among its points—obtain in virtue of some *non-spatial* features of the basic elements. So, in order for the points occupied by material objects to form metaphorical space E^* , they must exhibit some set of non-spatial features in virtue of which the “spatial” relations of E^* hold. Call this set of features F^* . Whatever features constitute F^* , the STRUCTURALIST's proposal is that those features (a) are not themselves literally spatial features; (b) are structurally isomorphic to the literally geometrical features of E ; and (c) lead to our representing E^* in experience as a metaphorical “space,” which explains the phenomenon of TRANSFER.

As noted above, metaphorical spaces come in two varieties: there are intrinsic similarity spaces, where the metaphorical “spatial” relations like distance are measures of the similarity of the elements in respect of some intrinsic feature; and there are relation spaces, where the “distance” relation is not fixed by any intrinsic features of the elements, but instead analytically depends on some set of irreducibly relational, non-spatial features of the elements (like kinship relations). So F^* could either be a set of intrinsic features of the points occupied by material bodies; or F^* could be a set of non-spatial relations among those points. I want to investigate what these features might be: what set of intrinsic features or non-spatial relations are the points of E^* represented as having, in virtue of which E^* is represented as a metaphorical space?

Before considering some candidates, I want to emphasize, again, that the claim the STRUCTURALIST needs to defend is not merely that the physical world *has* some set of features, F^* , that exhibits structure *S*, but that we actually *represent* F^* in experience. The physical world has any number of features about which experience reveals nothing. Suppose the chemical formulas of the compounds composing the bodies we perceive happened to form a metaphorical “chemistry space” with structure *S*. This would do nothing to explain how TRANSFER occurs, since the structure of “chemistry space” makes no impact on our experience and thus provides no basis for our applying the theorems of Euclidean geometry to the objects we perceive. What is needed for the STRUCTURALIST story to go through is some set of features F^* , *represented in experience*, which exhibits *S* and thereby *explains* our empirical applications of Euclidean theorems.

3.3.1. *First Objection: No Plausible Candidates to Play the Role of F^**

So what might F^* be? What set of features represented in experience allows us to achieve TRANSFER? There are no immediately obvious answers. All kinds of non-spa-

tial features—both intrinsic and relational ones—are plausibly represented in our experience: objects are perceived as having various colors, making various sounds, and so forth. But none of these sets of features seems like a plausible candidate to play the role of F^* . This is my first objection to STRUCTURALISM, as a proposed explanation of TRANSFER.

We can begin by considering whether any of the intrinsic features standardly thought to be represented in experience, such as colors, could serve as F^* . It seems fairly obvious that none could. The reason is not that these features all have the wrong structure (indeed, colors are taken to exhibit some aspects of S , as will be discussed below). Instead, the reason is that the (allegedly metaphorical) spatial properties we represent objects as having in experience—the properties we utilize in cases of TRANSFER, such as the right triangularity of the carpenter’s wooden beams—vary completely independently of the non-spatial intrinsic features, like color, that are represented in experience. Whether an object shows up in experience as triangular simply has nothing to do with whether it shows up as green or red. So our representing objects as having intrinsic, non-spatial features like colors cannot explain our representing them as exhibiting S and thereby inducing TRANSFER.

A more plausible candidate for F^* would be some set of non-spatial, *relational* features we represent the points of E^* as standing in. There is some precedent for the view that, in representing the objects of experience as standing in various “spatial” relations, we are really representing them, in the first instance, as standing in some other kinds of relations, either to each other or to the perceiving subject. Poincaré, for instance, held that seeing an object as at a certain distance in a certain direction amounted to representing the set of “muscular sensations” one would experience in moving one’s body to that location:

To localize an object simply means to represent to oneself the movements that would be necessary to reach it. It is not a question of representing the movements themselves in space, but solely of representing to oneself the muscular sensations which accompany these movements and which do not presuppose the existence of space.¹¹

On such a view, F^* could be the set of possible muscular sensations needed to reach the various points we perceive objects as occupying. The proposed explanation of TRANSFER would then run as follows: In experience, we represent E^* as standing in the set of relations F^* , because each point is represented in terms of the series of muscular sensations the subject would experience in moving to that point. We find that, in virtue of these relational features of its points, E^* bears a certain structural isomorphism to E – both exhibit structure S . We then utilize this isomorphism to “map” E^* to E —we map the literal geometrical relations among the points of E to

¹¹ Poincaré (1958, p. 47), quoted in Evans (1985, p. 371).

the muscular-sensation relations of E^* —which allows us to apply the theorems of Euclidean proof to the objects we perceive.

I noted that the *intrinsic* non-spatial features we represent in experience, like color, seem to vary completely independently of the spatial features we perceive objects to have, and so can't explain how we apply spatial concepts in experience. Poincaré's proposal does not suffer from the same problem: the set of muscle movements needed to reach the various points we perceive—call this set M —is not completely independent of the spatial features, like distance, that we represent those points as having, so we could at least expect to find some correlation between the spatial properties represented in experience and M . But the proposal still seems to me to be a non-starter. Although it was widely endorsed in Poincaré's time, the central idea of the proposal—that seeing an object as at some spatial distance amounts to representing a series of muscular sensations—has long been out of favor, and for good reason.¹² Perhaps the most important is a simple appeal to introspection: it simply isn't plausible that I represent anything about elaborate series of muscle-contraction sensations when I visually perceive the world as a spatial one. Looking out over San Francisco Bay, or up at the stars in the night sky, I might note the various spatial arrangements of the objects I am looking at. But surely I need not represent—or even have any clue about—what kinds of muscle movements would be needed to reach those distant buildings or stars, or what such mysterious movements might feel like.

As an account of what our perceptual experience is like, then, Poincaré's story about the space of possible muscle sensations is simply too phenomenologically off-key. Nothing could be more familiar than seeing an object as triangular; nothing could be more alien than representing the set of muscular sensations we'd feel on a voyage to the stars. To attempt to explain the former in terms of the latter seems almost perverse.

Another proposal about what plays the role of F^* —what set of features we represent the points of E^* as having, in virtue of which we represent them as forming a metaphorical space—can be extracted from certain views in contemporary physics about the nature of space. Many philosophers of science have been drawn to the idea that physical spatial relations hold in virtue of *causal* relations: on this kind of picture, to say that two objects are adjacent to each other is simply to describe them as standing in a relation of immediate (potential) causal connectedness.¹³ Looked at in this way, physical space-time can be seen as a kind of metaphorical space, a way of describing a set of non-spatial relations—what we might call “causal potentiality relations”—that hold among the basic elements of the physical world.

¹² See Evans (1985) for an extended discussion.

¹³ A well-known version of this idea in contemporary philosophy of physics traces back to the work of A. A. Robb, though, in some form, the proposal can be attributed to Leibniz. See Winnie (1977).

Such physics-inspired views of space-time as supervening on causal structure do not, in themselves, amount to a proposal of the kind the STRUCTURALIST needs. For such theses concern the metaphysics of space(time), not how we *represent* spatial features. This contrast—between proposals in contemporary physics about the underlying metaphysical nature of space-time, and proposals about how we represent spatial features in experience—can be brought out by considering some even more radical views from contemporary physics. According to some interpretations of quantum mechanics, the “space” of the universe emerges from the dynamics of the underlying quantum wavefunction; the universe is “spatial” only in the sense that the structure of the states of the wavefunction can be mapped onto geometrical spaces.¹⁴ It is not entirely clear what kind of entity the quantum wavefunction is supposed to be, on such theories. But what does seem clear is that we don’t perceive the physical world *as* a set of states of the quantum wavefunction; we don’t perceptually represent quantum wavefunction states. On this picture, then, physical space(time) is indeed, at the metaphysical level, a merely metaphorical space, since it supervenes on something not literally geometrical (a set of relations among possible states of the quantum wavefunction). But this gives us no plausible proposal about what might play the role of F*: it is not plausible that we represent the world as an evolving quantum wavefunction, so this view fails to give us a candidate set of features *represented in experience* that exhibits structure S, and thereby explains how TRANSFER occurs.

Returning to the proposal that we perceive the world as a metaphorical “space” of causal potentiality relations, we can acknowledge that the proposal is not immediately ruled out in the way the quantum wavefunction picture would be. For causal relations, unlike the quantum wavefunction, are not obviously outside our perceptual ken. Hume famously argued that we do not perceive causal relations; but many have disputed this claim.¹⁵ So the idea that we experience “spatial” relations only because we perceptually represent a set of causal relations cannot be immediately ruled out.

On closer examination, however, I think this proposal should still be rejected. I do not wish to dispute the claim that we perceive causal relations. But it seems to me that the spatial features we perceptually represent do not analytically depend on the causal relations we perceptually represent.

We can bring this out by considering what an actual causal-relation space is like. In his work on the nature of causation, Ned Hall frequently makes use of “neuron diagrams,” which depict metaphorical causal-relation spaces. These diagrams represent the causal connections among a collection of events. Each event is represented by a “node”; if the node is filled in, that means the event actually occurs, while empty nodes represent events that do not occur, but would have in some other possible scenarios. The causal connections between the events are represented by lines between nodes, with arrow-tipped lines representing “excitatory” causal connections—causal relations

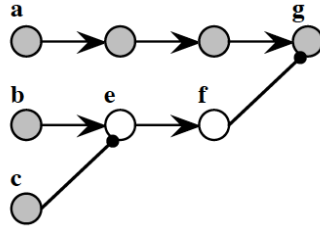
¹⁴ See Ney (2012).

¹⁵ See, e.g., Searle (1983), Siegel (2010).

where one event makes another happen—and circle-tipped lines representing “inhibitory” causal relations—where one event *prevents* another event from happening.

Here is one of Hall’s cases, with its accompanying neuron diagram:

Suzy is piloting a bomber on a mission to blow up an enemy target, and Billy is piloting a fighter as her lone escort. Along comes an enemy fighter plane, piloted by Enemy. Sharp-eyed Billy spots Enemy, zooms in, pulls the trigger, and Enemy’s plane goes down in flames. Suzy’s mission is undisturbed, and the bombing takes place as planned.¹⁶



In the diagram, the top row represents Suzy’s actions, the middle row Enemy’s, the bottom row Billy’s. The node labeled “a” represent Suzy’s initial action (taking off from her airbase, say), the node labeled “g” represents the target’s being bombed, the empty node at “f” represents Enemy’s shooting down Suzy (an event that did not in fact occur, but would have, had Billy not intervened), and so forth.

The spatial arrangement of the diagram gives us the structure of the relevant causal relation space: c is “closer” to f than it is to g, for example, because fewer causal links separate c from f than from g. But suppose you had actually been watching this whole episode from the ground. It is entirely consistent with the neuron diagram’s depiction of the relevant causal-relation space that c—the event of Billy’s shooting down Enemy—would have been represented by your experience as closer *in space* to g than to f. Perhaps the target that was bombed (event g) was quite close to the location where Billy fired at Enemy (event c), while the location where Enemy would have shot down Suzy (event f) was far-off in the sky. In such a case, the causal relation “space” depicted in the neuron diagram would come apart from the spatial relations that show up in perception; the latter simply don’t seem to supervene on the former. But, if the spatial relations we experience were really the metaphorical “spatial” relations of causal-relation space, such a dissociation should be impossible: the proposal was that representing spatial features in experience just is a matter of representing the layout of causal relation space.

There are simpler examples that bring out the relevant dissociation between causal-relation “space” and the space we perceive. When we enter a new house and

¹⁶ Hall (2004, p. 11).

play around with the light switches, we can figure out which switch controls which bulb, and we can thereby come to perceive the switch as causing the bulb to go on and off. But perceiving this kind of causal relation between the switch and the bulb does not change the *spatial* relation the two are perceived to have; the bulb doesn't suddenly appear to be located next to the switch that controls it, once we perceive the causal relation between the two. The spatial features represented in experience—like the distance between a switch and a bulb, or between two fighter planes—simply don't correlate tightly enough with the causal relations represented in experience for the proposal to work.

The discussion above is not meant to constitute a knock-down argument against the idea that there might be *some* set of features F^* that could serve as the basis of the STRUCTURALIST explanation of TRANSFER. I haven't canvassed an exhaustive list of possible features to play the role of F^* , and the STRUCTURALIST can simply insist that there must be some such set of features, even if we have trouble articulating what they might be. But I want to suggest that what my discussion has been pointing towards is a basic fact of descriptive metaphysics: we simply don't represent objects as having spatial features *in virtue of* representing them as having some *other* kinds of features. Plausibly, some features we perceive objects as having are represented in experience only because some other, more basic features are. We might see an object as, say, a coin only because our experience represents it as circular (and shiny, and silvery, etc.). But our experience of spatial features, like shape, does not work this way; spatial features are themselves basic in our perceptual representations. When we see an object as having a spatial feature—when the carpenter sees a set of wooden beams as triangular, for instance—we don't represent the object as having that spatial feature in virtue of representing it in some *other* way. With a non-basic feature our experience represents, like something's being a coin, there is a clear answer to the question "In virtue of what does this object appear to be a coin?" (The obvious answer is "In virtue of its appearing circular, and silvery..."). But there seems to be no comparable answer to a question like "In virtue of what does this object appear to be triangular?"¹⁷ The STRUCTURALIST proposal is untenable because it amounts to the claim that there is an answer to that misplaced question.¹⁸

¹⁷ Except, perhaps, an answer that points to more basic *spatial* features experience represents, like the straightness of the object's edges, or the points on its edges all being co-planar.

¹⁸ It is important to note that the question I claim to be misplaced is "In virtue of what other feature *represented in experience* do objects show up as instantiating spatial features." There may be some sense in which it is in virtue of a subject's brain being a particular state that her experience represents an object as square. Or we might think that it is in virtue of a subject's experiential state having a certain phenomenal character that it has the representational content that it does. (Indeed, I endorsed an idea of this latter kind in Chapter 1.) But what is at issue in evaluating the STRUCTURALIST proposal is whether we represent the world as a merely metaphorical, and that would require that we *represent* it, in the first instance, as instantiating some set of non-geometrical features, in virtue of which it is isomorphic to Euclidean space.

3.3.2. *Second Objection: Mere Isomorphism Does Not Lead to Transfer*

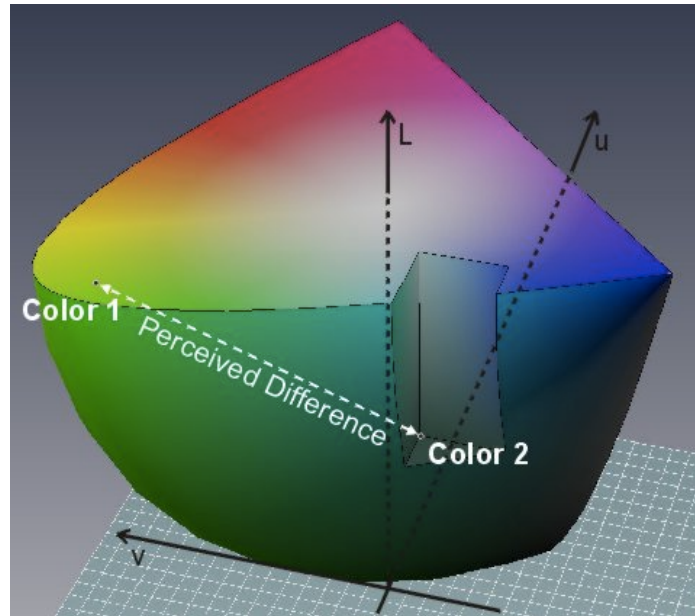
A second reason to reject the STRUCTURALIST explanation of TRANSFER is that, even if the STRUCTURALIST could point to a set of non-spatial features—the elusive F^* —that was both represented in experience and structurally isomorphic to Euclidean space, such an isomorphism is insufficient to induce TRANSFER. For there are indeed certain non-spatial features represented in experience that are isomorphic to E – certain features in virtue of which we could represent E^* as a metaphorical space with structure S . But, in such cases, we do not take the results of Euclidean proof to be applicable to experience in virtue of these features.

Color space, for example, is a metaphorical space whose elements—individual colors—are represented in experience. We obviously see objects as having colors. And the relations among those colors are isomorphic to the spatial relations of a portion of three-dimensional Euclidean space: color space is taken to have three-dimensions, which chart the hue, brightness, and saturation (or three other dimensions that capture the same ranges of variation) of each color, with individual colors placed along those dimensions in such a way that the distance between any two corresponds to the degree of similarity of the colors on that dimension. Within this metaphorical “space,” various spatial relations like betweenness and distance represent these relations of similarity. There are various ways to achieve this mapping and create metaphorical color spaces. In some, the “distance” between any two colors—the degree of perceived color similarity between them—can be calculated using the standard Euclidean distance measure, based on the Pythagorean theorem.

But, crucially, the “spatial” relations of color space do not show up *as such* in our experience. In seeing two objects, or two points of the visual field, as having specific colors, in virtue of which they can metaphorically be described as a certain “distance” apart in color “space,” we do not see those objects *as* having any particular distance relation to each other, such that we would be induced to apply the results of Euclidean proof to them.

To see this, consider the following depiction of color space, in which perceived differences between colors are mapped onto a “solid” such that those differences are proportional to the distances within the solid:

My claim is that no such “ground” of spatial content shows up *in experience*. The spatial properties represented in experience are *representationally* basic: unlike the metaphorical spaces, such as color space, that we can come to represent in virtue of representing non-geometrical features, the space we perceive is not perceived in virtue of our perceiving some more basic set of features.



Arranging colors in this way allows us to see the degree of color similarity between two colors in virtue of seeing the spatial distance between the points with those colors. We can then utilize various geometrical concepts to analyze the colors. For example, the diagram includes a set of axes (labeled V, L, and U) and a triangular prism representing the distance between Color 1 and Color 2 along each axis. Using these orthogonal distances, we can calculate the “straight line distance” between the two colors, by applying the Pythagorean formula (the distance will be the square root of the sums of the squares of the three orthogonal distances).¹⁹

But note that we have to arrange the colors *spatially* in this way in order to be able to apply the Euclidean result—the Pythagorean theorem—to them. Just perceiving the features that determine the *metaphorical* spatial relations among a set of points does not induce TRANSFER. If we rearrange the colors of the diagram, so that each color is still instantiated in exactly one pixel, but the coordination of color differences with spatial arrangement is not maintained—if, say, we randomly arrange the points along a single line—we would continue to see the colors and their different hues, saturations, and brightnesses. But just perceiving this set of features—the features that determine the metaphorical color “space” that is isomorphic to Euclidean space—does *not* lead us to apply our Euclidean geometrical reasoning to what we perceive. We need to arrange the elements of the metaphorical “space” of the colors along *literal* geometrical axes in order to see the relation between Euclidean proof and what we perceive – in order to perceive the points *as* standing in spatial relations. So the kind of isomorphism STRUCTURALISM appeals to between the allegedly metaphorical space of experience

¹⁹ See <http://www.zuschlogin.com/?p=194>, which is the source of the above diagram.

and the literal space of Euclidean geometry cannot be what explains TRANSFER, a phenomenon that occurs only when we perceive objects *as* having spatial properties.

To summarize: Even though metaphorical color space has many of the structural features of Euclidean space, colors simply do not show up in experience as standing in spatial relations, in a way that would allow a subject to transfer her knowledge from the domain of geometry to the domain of perception. The STRUCTURALIST’S explanation of TRANSFER—that it is induced by the structural isomorphism between the literal geometrical space of Euclidean proof and some isomorphic, non-spatial relations, F^* , represented in experience—is simply inadequate, because TRANSFER does not occur when we have mere structural isomorphism between Euclidean space and a set of experienced features. TRANSFER only occurs when objects are represented as literally having geometrical properties – when a set of wooden beams shows up *as a triangle*, for example. That a set of objects might form a “triangle” in the metaphorical “space” of the colors those objects are experienced as having does not lead us to apply spatial concepts to those objects, and so mere structural isomorphism cannot explain TRANSFER.

Again, the point I am making here is quite intuitive, and quite straightforward. While we may speak of many objects metaphorically as forming “spaces,” that is not the sense in which objects show up in experience as having spatial features. Objects can show up in experience *as* triangular, or *as* straight lines, or *as* a certain distance apart. That is the kind of phenomenon that induces TRANSFER, and explains its being so automatic. When objects show up as having *non-spatial* features, like colors, that can be mapped to geometrical spaces—when they show up as forming a *metaphorical* space—experience does not provide the grounds for us to automatically apply spatial reasoning, the way we do in cases of TRANSFER. Mere isomorphism is an insufficient explanation of the phenomenon; only the representation of literal geometrical features in experience—genuine COMMONALITY between Euclidean proof and spatial experience—can explain cases like the carpenter’s.

3.4. AN ACCOUNT OF SPATIAL EXPERIENCE

I have argued that both our *a priori* geometrical reasoning and our perceptual experience of the spatial properties of material objects deploy concepts of literal, geometrical spatial features, and that these concepts are not derived from experience. The sharing of spatial content across these two cognitive domains—the phenomenon of COMMONALITY that links spatial perception and abstract Euclidean proof—explains how TRANSFER occurs, and why it doesn’t occur when we represent non-literal “spaces” (such as color “space”) that are merely structurally isomorphic to Euclidean space. In this section, I will fill out some of the details of this picture, offering an account of spatial experience that makes sense of these facts. I will also tie this account back to the discussion of Chapter 1, showing how the link between our *a priori* geometrical

concepts and our experience of space provides the basis for a satisfying internalist account of spatial perception.

3.4.1. *A Rationalist Account*

The concepts that feature in Euclidean proof are not derived from experience. The very same concepts feature in the contents of perceptual experience. So *a priori* concepts are at play in our experience of the empirical world: sensory experience, in its very nature, depends on a grasp of spatial features that is not itself derived from experience.

This view can be thought of as a particularly strong form of *rationalism*, or anti-empiricism. The core thesis of empiricism is that *all* our concepts are derived from experience. Traditionally, the greatest challenge to this kind of thesis comes from the domain of abstract or mathematical thought: many in the rationalist camp have argued that the concepts we employ in our *abstract* reasoning couldn't be derived from experience. In the previous chapter, I made just this kind of argument: I showed that the geometrical concepts we use in performing Euclidean proof are not derived from experience, and so an all-encompassing empiricist thesis about the origins of our concepts must be false. In this chapter, I have taken the argument a step further, claiming that empiricism is wrong even "on its own turf." Our non-experientially derived geometrical concepts are not confined to a domain of abstract mathematical cognition that sits above experience. Those very *a priori* concepts are implicated in experience itself.

On my picture, then, there is a crucial common thread running through mathematical reasoning and perceptual experience. But it is also important to recognize that experience is quite different from mere thought: it is essentially *qualitative*, in a way that the abstract reasoning of mathematics is not. While mathematical reasoning may be accompanied by various "mental images," those images are not essential to mathematical cognition. Working through a proof of the Pythagorean theorem, different subjects might entertain wildly different images of triangles, or none at all. The existence of blind geometers again brings out quite sharply how inessential mental imagery is to mathematical thought: while blind mathematicians might utilize some kind of "tactual imagination" in the course of their proofs, they seem to be capable of engaging in the very same course of mathematical thought as sighted subjects, without experiencing anything qualitatively similar to visual mental imagery. In the case of sensory experience, by contrast, the qualitative character of our mental states seems essential to them: it is only in virtue of having an experience with a certain qualitative character that I count as *seeing* an object as a right triangle.

So Euclidean proof and perceptual experience both deploy our *a priori* geometrical concepts, but only the latter does so in a way that is essentially tied to the qualitative character of our mental states. The reason, I claim, is that spatial experience—experi-

ence of the spatial features of the objects we perceive—is, in its very nature, a *qualitatively rich* way in which our *a priori* spatial concepts can be deployed, while Euclidean proof is an intellectual way to articulate and elaborate those same concepts. When we have a sensory experience of a certain character—the kind of experience you have when you view a square surface head on, say—our geometrical concepts are exhibited in a particular manner. Once we attain a certain level of cognitive sophistication, those very same concepts can also be exhibited without such rich phenomenology, when we abstractly reason about the nature of the features—triangularity, distance, etc.—that our spatial concepts represent. This is what we do in Euclidean proof.

3.4.2. *Proto-Concepts and Core Cognition*

That is the basic picture. I now want to address an immediate worry. I've suggested that we have a set of *a priori* geometrical concepts that are employed in the domain of abstract Euclidean proof, and that these very concepts show up in perceptual experience. So I might seem to be suggesting that our capacity to perceive the world depends on our possession of concepts we develop through abstract mathematical reasoning. But, one might think, perceptual experience, and in particular, experience of simple spatial features like the shape, is more basic than Euclidean proof. Surely the sophisticated reasoning we do in Euclidean proof can't be implicated in the mere perception of shape. For there are subjects—humans in societies that don't do Euclidean proof, children who haven't yet had training in mathematical reasoning, non-human animals—who can experience spatial features (in seeing an object as triangular, say) but who cannot engage in abstract Euclidean reasoning. So my picture of spatial experience seems to *over-intellectualize* a very basic form of cognition: it makes it seem as though merely seeing the shapes of objects requires a capacity for highly sophisticated mathematical reasoning.

This criticism would be quite devastating if my view implied that we need to first articulate our geometrical concepts mathematically, by doing Euclidean proof, before those concepts could show up in the contents of experience. But I do not mean to suggest that mathematical articulation of our *a priori* geometrical concepts must *precede* spatial experience. Instead, my claim is that we have a primitive set of basic spatial concepts—concepts of features like lines, points, and circles—that we can eventually articulate explicitly, by doing Euclidean proof, but which also, antecedently, are exhibited (in a qualitatively rich form) in perceptual experience. Spatial experience and Euclidean proof have a shared cognitive origin, in our primitive, *a priori* grasp of basic spatial features.

This primitive grasp should not, perhaps, be characterized in terms of full-fledged concepts. Concepts are sometimes taken to be constituents of thoughts, or predicates of possible judgments. On such readings, concepts are fairly cognitively sophisticated – employing them might require language, or at least some quasi-linguistic capacities. But, again, it seems plausible that many non- or pre-linguistic subjects are capable of

spatial perception. So it is implausible that the spatial representations that feature in experience require such sophisticated conceptual capacities.

Instead, then, we can think of our basic grasp of spatial features in terms of our innate possession of a set of *proto-concepts*: representations of basic spatial features that are more primitive than the linguistic capacities of full-blown conceptual thought. We can now summarize my proposed account of spatial experience more fully: We have an innate, primitive grasp of basic spatial structure—a set of proto-concepts—that features in our cognitive lives in two ways. On the one hand, these proto-concepts are deployed in our experience of the world around us, representing the presence of particular spatial properties, in a qualitatively rich format. On the other hand, our proto-concepts are explicitly articulated in the practice of Euclidean proof, where we explore in detail the features of the spatial properties these innate concepts represent.

On this picture, the link between experience and Euclidean proof is as follows: An experience of an object as square is one way that a spatial proto-concept can be active in a subject's cognitive life; it is one way for a subject to represent space. In humans, once we achieve a certain level of sophistication, another way these concepts can be exhibited—another way a subject can have mental states that represent spatial features—is through practices like Euclidean proof. TRANSFER is explained by the fact that both experience and Euclidean proof are ways we utilize the same primitive, *a priori* representation of space.

Elizabeth Spelke's theory of spatial cognition as a system of "core knowledge" is one way of fleshing out this idea. The work of Spelke and her collaborators on spatial cognition is part of an emerging research program in developmental psychology, which pushes back against the long empiricist tradition of seeing human infants as "blank slates" whose cognitive resources must be derived from the "blooming buzzing confusion" of unconceptualized sensory experience. Against this traditional empiricist picture, Spelke argues that:

Cognition develops from its own foundations, rather than from a foundation of perception and action. Initial cognitive capacities give rise, moreover, to conceptions that are largely appropriate to the experience of children and (nonscientist) adults. Finally, initial conceptions form the core of many later conceptions; they are enriched and refined as knowledge grows, but they are rarely overturned.²⁰

These "initial cognitive capacities," or core cognition systems,

²⁰ Spelke et al. (1992, p. 605).

emerge over evolutionary time in a species, such that they are likely to be shared not only among all humans but also with other species. Such representations can be called our “innate” conceptual resources.²¹

Spelke has argued that these innate conceptual resources include core knowledge of geometry.

Much of Spelke’s research is still at a fairly preliminary stage; in describing her theory, I do not intend to endorse all of its details. Instead, what I want to do is to highlight certain aspects of the core cognition research program, and of Spelke’s application of the general thesis to the domain of spatial cognition, in order to illustrate one empirically-grounded way that my picture of *a priori* spatial proto-concepts might be filled out.

Before discussing Spelke’s claims about geometrical cognition in particular, I want to highlight some of the ways that Susan Carey, one of the pioneers of the core cognition research program, describes the basic concept of core cognition, in order to flag the relevant common ground between this developmental theory and my philosophical account of spatial cognition. First, Carey writes that “the representations in core cognition cannot be reduced to perceptual or sensory-motor primitives.” As I argued in the previous chapter, this is true of our spatial concepts: my analysis of the practice of Euclidean proof shows that the concepts we have of geometrical objects like lines and triangles cannot be derived from experience. Second, “some core cognition... is shared by other animals. At least some early developing cognitive systems in humans have a long evolutionary history.” This seems plausible for spatial representations, and the idea that a core cognition system for spatial representations could be shared with non-human animals is in line with my suggestion that we think of the spatial representations that show up in perception as a set of primitive proto-concepts, rather than as full-fledged concepts (which would make the representations in question implausibly intellectually-demanding). Finally, Carey notes that, although the basic systems of core cognition are innate and are expressed in the cognitive activities of very young children, “[c]ore cognition is elaborated during development... but it is never rendered irrelevant. It is never overturned or lost.”²² This again fits with my claim that our *a priori* spatial proto-concepts, which are first exhibited in experience, are made more explicit and precise as we acquire linguistic capacities, eventually leading to the full specification of the contents of our innate grasp of space in the practice of Euclidean proof.

Spelke carries over the basic outlines of Carey’s picture to the domain of geometrical cognition, and she argues that each of the features described above applies to our basic spatial representations. Spelke writes that core cognition of geometry “has a long evolutionary history” and is not “unique to humans.” Furthermore, our system of basic

²¹ Barner and Baron (2016, p. 3).

²² Carey (2009, pp. 66-67).

geometrical representations “emerges early in development, largely independently of any specific experiences with the entities to which it applies.” Finally, Spelke argues that core geometrical cognition is one of several “innate systems with mathematical content,” and, in particular, that core geometrical cognition has “genuine spatial content on which we draw when we learn and perform symbolic mathematics,” such as Euclid’s system of geometrical proof.²³

In support of the first claim—that non-human animals share our core geometrical cognition systems—Spelke cites evidence that there is a dedicated “mechanism of navigation” that is “widespread across animals, including humans, and centrally focused on environmental geometry.” She writes:

As humans and other animals navigate, they represent both distances and directional relationships on the extended surfaces that bound the navigable layout.... The existence and properties of these representations are revealed when animals lose their orientation and must draw on memory for the positions of these surfaces to reorient themselves.... [R]ats recover their orientation by analyzing the shape of their surroundings. When rats explored a rectangular room in which food was buried and then were disoriented by slow turning in the dark, they used the lengths and relative directions of the room’s walls to reorient themselves and therefore searched for the food at the two locations that were congruent with the room’s geometry (e.g., at a corner to the left of a long wall). Subsequent research revealed that the capacity to reorient by the shape of the borders of the environment is found in animals as distant as humans... and ants.... [S]ensitivity to geometry is shown across a wide variety of navigation tasks, in oriented as well as disoriented animals or humans tested by a diverse set of behavioral and neurophysiological methods.... It is encoded automatically, independently of processes for encoding other features of the environment such as landmark objects... by a distinct neuronal network that includes the hippocampus.²⁴

In addition to being relatively primitive, and thus plausibly implicated in sensory experience, this geometrical navigation system seems to be innate:

Importantly, the ability develops in animals independently of experience in a geometrically structured layout: Chicks and fish who were raised since hatching in a geometrically uninformative, circular environment reoriented by the shape of a rectangular environment the first

²³ Spelke (2011, p. 288).

²⁴ Spelke et al. (2010, pp. 867-868).

time they encountered it, and they did this as reliably as chicks or fish who were experienced at navigating by geometry.... In contrast, disoriented animals' use of nongeometric properties such as surface brightness or texture is highly influenced by experience, both in controlled-reared fish... and in mice who are trained to use surface features to locate objects.²⁵

Spelke thus argues that a core system of geometrical cognition is *a priori*, in that it can develop independent of experiential contact with the entities to which its representations apply. This system is also quite primitive, in that it shows up across the animal kingdom, including in rats and chicks. Such a cognitive system seems like a good candidate to serve as a locus of *a priori* spatial proto-concepts, of the kind I have suggested underlie our spatial experience and cognition.

Experiments involving non-human animals can take us only so far, however, if we want to fill out the picture of *human* spatial cognition that I have been developing, which is intended to encompass not only spatial perception, but also reasoning as sophisticated as Euclidean proof. Happily, some of Spelke's other research helps fill this gap. The relevant studies focused on subjects from an indigenous group of the Amazon region, the Mundurucu, whose "language does not have terms for basic Euclidean concepts such as parallelism or right angle" and who "had no instruction in geometry." By focusing on the Mundurucu, Spelke and her colleagues hoped to test whether the basic intuitions that underlie Euclidean proof—our sense that an axiom like Euclid's fifth postulate is "self-evident", for instance—were dependent on particular cultural contexts, deriving from cultural institutions like language and mathematical training.

Spelke found that this was not the case, summarizing her results thus:

Our experiments... provide evidence that geometrical knowledge arises in humans independently of instruction, experience with maps or measurement devices, or mastery of a sophisticated geometrical language.... the spontaneous understanding of geometrical concepts and maps by [Mundurucu subjects] provides evidence that core geometrical knowledge... is a universal constituent of the human mind.²⁶

Spelke's experiments required subjects to sort images based on a grasp of basic Euclidean features like parallelism and right angles. The experimenters found that

across a variety of trials targeting different geometric properties, adult and children Mundurucu used principally the abstract geometric properties of the figures and performed well above chance. Furthermore,

²⁵ Spelke et al. (2010, p. 867).

²⁶ Dahaene (2006, p. 286).

their performance across trials correlated tightly with the performance of adult and children control participants from the US. Despite dramatic differences in geometric education between these two groups, the trials that were harder for the Mundurucu were also harder for the US participants. This test therefore reveals a signature of geometric intuitions.²⁷

The way that the images were perceived by all subjects seemed sensitive to the presence of the basic geometric features on which Euclidean proof is based, suggesting that concepts of these features show up in the experience of human subjects, whether or not they have received formal training in geometry. Furthermore, the concepts thus revealed seemed to go beyond anything that could itself be derived from experience, as the judgments of Mundurucu subjects with respect to geometrical figures that outstrip any possible sensory experience—the infinite lines that are central to Euclidean proof—also accorded with the axioms of Euclid:

Participants were introduced to an ideal shape, either an infinite plane or a sphere. The experimenter narrated the properties of the shape (“it is very, very flat and goes on forever and ever” or “it is very round, like a ball”), straight, infinite lines, as well as dots. Following this introduction phase, participants were given a list of questions pertaining to the properties of straight lines. Impressively, Mundurucu adults and children performed extremely accurately at the test, especially on the plane. Most of them agreed that a new straight line may always be placed in such way that “it would never cross” a first straight line.... Beyond categorization, this last test argues for elaborate, non-perceivable concepts being universal.²⁸

This test essentially probes the subjects’ intuitions about Euclid’s famous fifth postulate, which (in one of its forms) is the proposition that through any point not on a given line, exactly one line can be drawn parallel to the original line. This postulate differentiates Euclidean geometry from other two-dimensional geometries (elliptical or hyperbolic geometries), in virtue of features that go beyond what is accessible in perception (e.g., infinitely extended lines on a plane). Thus, the Mundurucu’s responses on the tests suggest that they, even in the absence of formal geometrical training, intuitively pick out specifically Euclidean features when presented with tasks that concern entities that could never be present in experience.

²⁷ Izard et al. (2011, p 321).

²⁸ Izard et al. (2011, p. 328).

I have not gone into many of the details of Spelke’s empirical research or her overall theory of geometrical cognition. In particular, one central idea of Spelke’s theory is that there are in fact two distinct, innate systems of core geometrical cognition—one for large-scale features of the subject’s environment, used for navigation; and one for smaller objects and two-dimensional displays, used for sorting—which are integrated at a somewhat later stage of development. It is only once this integration (which may turn on the acquisition of linguistic capacities) has occurred that full Euclidean geometrical cognition can be achieved. So there is a clear sense in which the cognitive resources available at an early stage of development—the resources on which spatial experience can draw—are more primitive than the full-blown concepts of Euclidean geometry. Still, according to Spelke and her colleagues, there is a direct link between the two kinds of cognitive capacities:

Our research provides evidence that the basic principles of Euclidean geometry are reflected in intuitions of space that develop progressively throughout childhood, but still appear universal. From early childhood, the perception of shapes provides an intuitive ground corresponding to the geometry of non-oriented solid objects.... These early intuitions are enriched over development.... Although the acquisition of the relevant geometric lexicon seems to play a role in the formation of some discrete geometric categories for US children, this role may be reduced to that of a category-maker and catalyst. Indeed, in a population that did not receive any education or share any relevant lexicon in geometry, a brief description of ideal shapes sufficed to elicit elaborate thoughts about the fundamental ideal concepts of Euclidean geometry, such as infinite lines or parallelism.²⁹

The hypothesis that Spelke endorses as the best explanation of these results takes

geometry to be grounded on “core knowledge,” i.e. representations of abstract content that were selected by evolution, and provide useful guidelines to interpret the environment and learn. Representations of ideal straight lines or planes may be present in the architecture of the perceptual system to serve as anchors for perception.... Under this hypothesis, during childhood these implicit anchors would need to be progressively reformatted into explicit representations, to be able to enter thought processes and be manipulated directly.³⁰

²⁹ Izard et al. (2011, p. 329).

³⁰ Izard et al. (2011, p. 330).

Again, this “core knowledge” hypothesis mirrors quite closely the account I have argued for, on which we have an *a priori* grasp of space in virtue of possessing a set of innate geometrical proto-concepts. These proto-concepts, I have suggested, show up in experience, providing the determinate, geometrical content of our spatial perception; and, at the same time, they are the foundation of the more abstract reasoning about geometry we do when we perform Euclidean proof.

3.4.2. *The Role of Phenomenology in Spatial Experience*

I now want to connect this picture of spatial experience to the discussion of Chapter 1, which focused on the problem of how an internally-determined perceptual state could come to represent a mind-independent world. The problem turned on the link between phenomenology and perceptual content. The former, I suggested, is internally-determined – what it’s like to be in a given experiential state is independent of how that state is connected to the external world. But perceptual states also seem to give us a determinate conception of the external world, to represent the world as being some particular way, just in virtue of having the character that they do. This latter fact seemed incompatible with the internalist picture, since, on internalism, the only conception of the external world available from the subject’s own perspective would seem to be as the unknown cause of her internally-determined phenomenal states.

I suggested that this is in fact the situation we are in with regard to our experience of color: since the character of an experience of redness can vary independently of the particular color it is connected to in a given context, the phenomenal properties of a color experience can’t give us insight into which particular color the experience represents. A color experience can only represent the presence of that property, *whatever* it is, that typically causes experiences of that type. But I think we now have a way to explain why not every aspect of our experience has this kind of representational flexibility, and how experience can represent spatial features in a way that confers acquaintance.

The argument of the last chapter showed that the geometrical concepts we employ in the practice of Euclidean proof are genuinely contentful: they represent specifically geometrical objects—the figures of Euclidean space—rather than merely indicating a set of formal features, which might be exhibited by colors or by beer mugs. Thus, in the domain of geometry, we have a set of representations that do not exhibit the representational flexibility that I argued characterizes our color experience. The argument of this chapter, in turn, showed that it is these very representations—representations of specifically geometrical, and not merely metaphorically spatial, features—that constitute the contents of our spatial experience.

But we might still wonder what the relation is between the non-flexible representational content of our spatial experiences and their phenomenal character. Spatial experiences, after all, do have rich qualitative character. This distinguishes them from mere abstract reasoning about space. And the character of our experiential states does

not seem to be merely contingently connected to those states' representational content; an experience of the kind we have when we see an object as square does not have the same representational flexibility as a color experience. The reason there is a non-contingent connection between the character of a spatial experience and a particular, determinate geometrical content, I want to suggest, is that having an experience of a particular qualitative character just is a way for our *a priori* proto-concepts of spatial features to be exhibited. An experience of shape "pulls in" an *a priori* spatial proto-concept, and thereby represents the determinate property that that proto-concept represents.

By contrast, when we experience an object as red, there is no *a priori* concept (or proto-concept) of redness for that experience to "pull in," no non-experiential grasp of color the experience could be expressing. That is why color experiences are representationally flexible: they don't, on their own, hook up to any specific content. Instead, all they can do is represent a property indirectly: in having an experience that represents redness, we are representing the property, whatever it is, that causes such experiences, a property with which we do not have acquaintance.

I have flagged one way in which the representation of spatial features in experience differs from the abstract reasoning of Euclidean proof: the former, unlike the latter, involves qualitatively rich phenomenal states. Another key difference is that experiential representation of spatial features, unlike abstract geometrical reasoning about those features, involves *approximation*. Indeed, our experiential representation of the world is merely approximate in just about *every* respect: no matter what domain of perception we consider—color vision, hearing, touch—our discriminatory capacities have an upper bound. So too with our representations of spatial features in experience: a key claim from the previous chapter was that nothing experiential—no kind of image or diagram—could be precise enough to pick out "perfect" geometrical properties, like continuity (as opposed to mere denseness). But, if spatial experience is merely approximate, how could it involve the same concepts (or proto-concepts) as Euclidean proof, which depends essentially on the exactness of the geometrical notions it utilizes?

When a chessboard is represented as square in a visual experience, it is represented as having the property we reason about in Euclidean proof. But, in the domain of experience, its being represented as a square means only that it is represented as within a certain *range* of geometrical variation of our non-experientially derived concept of a *perfect* square. The geometrical concept shows up as a kind of anchoring for the merely approximate content of experience.³¹ This concept distinguishes, in a way that no purely imagistic cognitive state could, between continuity and mere denseness, a feature made explicit in the practice of Euclidean proof. But, in a visual experience, where the (proto-)concept of squareness is exhibited in virtue of that experience's qualitative character, the content of the experience is indeterminate, in a way that reflects the

³¹ Here, I echo Izard et al. (2011, p. 330): "Representations of ideal straight lines or planes may be present in the architecture of the perceptual system to serve as anchors for perception."

limited precision of experience as a representational faculty. Because perceptual states, unlike abstract mathematical thought, have their representational content determined by the imprecise medium of phenomenology, they can only ever represent to a certain degree of approximation.

3.5. CONCLUSION

Our spatial experiences, in virtue of their character, pull in our *a priori* geometrical proto-concepts. This relationship is not contingent: having an experience with the qualitative character of our experience of a square object just is a way to deploy a particular *a priori* spatial representation. So such experiences do not exhibit the kind of representational flexibility characteristic of color experience. Our spatial proto-concepts, which are innate, play two roles in our cognitive lives. They are exhibited in our spatial experience. And they are elaborated into full-blown concepts as we acquire linguistic capacities, eventually allowing us to articulate the full geometrical content of our *a priori* grasp of space through the practice of Euclidean proof. That is how we can explain the link between our *a priori* geometrical reasoning and our experience of the spatial features of the empirical world.

CHAPTER 4

THE DISTINCTION BETWEEN PRIMARY AND SECONDARY QUALITIES

On the account of spatial cognition I have developed in the preceding chapters, primary and secondary qualities are represented in very different ways in experience. Our *a priori* grasp of spatial properties allows for such properties to be represented in experience in a way that reveals their nature. But, because we lack such *a priori* concepts of colors, our experience of a color property fails to reveal that property's nature – experience can represent a color only by way of a placeholder.

In this chapter, I address an externalist challenge to this account of color experience, which insists that our experience presents colors, just as much as shapes, in a nature-revealing way. On this externalist picture, the character of an experience of color is itself constituted by the nature of the color perceived; so such an experience necessarily reveals the nature of that property. I reject this account because it fails to make sense of various *spectrum inversion* scenarios, which make clear that the conception of the secondary qualities we get from perceptual experience does *not* reveal the intrinsic natures of those properties. When different subjects' color experiences diverge, disagreement eventually gives out – we are left with no conception of what the disagreement is a disagreement *about*. This is because, being derived from experience, our concepts of color properties give us no non-experiential grip on what those properties are. I contrast such cases of spectrum inversion and disagreement about color with parallel cases involving shape, in order to bring out the special role our *a priori* concepts play in giving our perceptions of spatial properties the nature-revealing content they have.

4.1. THE SIMPLE VIEW OF COLOR

In “A Simple View of Colour,” John Campbell presents a theory of the colors of objects on which

redness, for example, is not a disposition to produce experiences in us. It is, rather, the ground of such a disposition. But that is not because redness is a microphysical property—the real nature of the property is, rather, transparent to us. This view of colours would be available even to someone who rejected the atomic theory of matter: someone who held that matter is continuous and that there are no microphysical properties. (Campbell 1993, p. 258)

On this picture, the traditional distinction between primary and secondary qualities is something of a mistake: colors, no less than shapes, are independent of anyone's experience, and their nature, like that of shapes, is fully graspable even by someone who is ignorant of the scientific theories that have, since Locke's time, led philosophers and scientists to question Campbell's Simple View.

In laying out his account of color, Campbell cites Barry Stroud's John Locke Lectures as a source of inspiration for his thinking. Stroud, in *The Quest for Reality* (the book that grew out of those lectures), likewise poses a challenge to the traditional distinction between primary and secondary qualities, as applied to the colors of objects. In responding to commentary on that book, he writes:

It is true that something that is yellow... is such as to be *seen* to be yellow in such [optimal] conditions. What is seen to be so in optimal conditions is the same as what is so. That link holds of the shapes of things as well, and of their number, and of many other properties. Something that is rectangular is such as to be seen to be rectangular in conditions optimal for telling the shapes of things by looking at them. A collection of seventeen objects is such as to be counted by us as seventeen in number in conditions optimal for us for telling the number of things by counting. These remarks, "rigidly" interpreted, link the idea of something's being yellow, or rectangular, or seventeen in number, with something that is in fact true about us. They mention a disposition involving us that is possessed by objects that have those properties. But they do not express a dispositional view of those properties. (Stroud 2004, p. 429)

Here Stroud is denying that there is anything "special" about the way in which color properties are linked to our experience of them. Shape and number properties are paradigms of categorical properties: no one takes shape or number properties to be "secondary" or "subjective" properties; no one claims they are merely dispositions to cause certain perceptual experiences in us.¹ And yet, Stroud notes, there *is* a connection between such properties and our experiences of them: in optimal conditions, we perceive the presence of these categorical properties. Stroud's suggestion is that the connection between color properties and our experiences of them likewise amounts to no more than this: just as with shape and number, we are able to perceive color properties under the right conditions. But that does not mean there is any sense in which the

¹No one, that is, who is not an "all out" phenomenalist. The point is that, if *any* properties are real, categorical properties of objects, shape and number properties would be such properties. In this chapter, I will not be directly considering the full phenomenalist view on which all properties—shape and number properties as much as color properties—are taken to be no more than dispositions to cause experiences in us (though what I say here might be useful in responding to such views).

property itself is subjective or non-categorical. Categorical properties can perfectly well be disposed to appear to us just as they are, without that in any way making the properties that have such dispositions themselves *dispositional properties*.

Stroud's claim can be spelled out in terms of two bi-conditionals:

- (1) X is yellow \leftrightarrow X is such as to look yellow (in normal circumstances to normal perceivers)
- (2) X is rectangular \leftrightarrow X is such as to look rectangular (in normal circumstances to normal perceivers)

Stroud's suggestion is that these two bi-conditionals have approximately the same status: they are both factual statements relating a property to the dispositions that property has to cause certain perceptual experiences in us. But neither one says anything more than that. In particular, we should not view (1) as giving a *definition* or *analysis* of yellowness, in contrast to (2), which (on this picture) *merely* makes a factual claim about rectangularity, without giving any kind of analysis of it. That would be a way of drawing out the distinction between yellowness (a secondary quality) and rectangularity (a primary quality): it would be a way of saying that yellowness, but not rectangularity, *amounts to* a disposition to cause certain experiences in us. This is the picture Stroud is concerned to reject.

In this chapter, I will present a challenge to the alternative picture sketched by Stroud and Campbell, on which color properties are in no way more subjective or dispositional than properties traditionally classified as primary qualities. I will argue that there is a real sense in which color is—and shape is not—linked to our experience, *at the conceptual level*.

It is important to note that it is at the conceptual level that the difference between primary and secondary qualities emerges. Some traditional ways of delineating those categories—in terms of whether the property in question can be perceived by more than one sense, or in terms of whether the property is somehow *dependent* on us for its existence, or in terms of whether the property is “real” in some metaphysically weighty sense—do not reveal the truly significant feature of the distinction. Those ways of spelling out the distinction have what we might call a more “metaphysical slant” than the version of the distinction I will be arguing for here. The metaphysical slant can be seen in the use to which the distinction has typically been put: from Democritus to Locke to modern-day “color eliminativists” like Paul Boghossian and David Velleman, those who have argued for a primary/secondary distinction have usually done so in the context of arguing that those properties falling on the “secondary” side of the

divide are somehow metaphysically unreal.² It is this metaphysical “unmasking” project against which Stroud sets his face in *The Quest for Reality*: his goal is to show that any attempt to demonstrate that color properties are metaphysically unreal will ultimately fail.

I will not be engaged in any such unmasking project; in suggesting that colors are secondary qualities—that there is a way in which they are more subjective than shape—I do not intend to suggest that colors are unreal. Indeed, as will emerge below, the view of colors that I will be sketching actually makes it *less* coherent to posit an error theory for our beliefs about colors than it would be to posit such a theory for our beliefs about the shapes of objects.

Like Stroud, Campbell situates his account of color by contrasting it with alternative *metaphysical* pictures of the place of color in the world. He begins his paper with this sketch of an argument: “Physics tells us what is objectively there. It has no place for the colours of things. So colours are not objectively there.” According to Campbell, this is the “background argument” to a dispute between J. L. Mackie (a color eliminativist) and John McDowell (a dispositionalist about color). It is also precisely the kind of unmasking argument that Stroud finds untenable. And Campbell, like Stroud, refuses to accept the underlying metaphysical assumptions embedded in the argument: he denies the premise that physics tells us what is objectively there (or, more precisely, the premise that physics tells us *everything* that is objectively there), and he goes on to reject the “metaphysical” version of the claim that colors are secondary qualities. Campbell says that colors *are* “objectively there,” and that they are not dispositions, but are rather categorical properties, just as shapes are. These are, I take it, metaphysical claims; and, again, I do not intend to reject or “unmask” any such metaphysical claims when I say that colors are secondary properties at the conceptual level. The way in which color and shape differ, on the account I will develop below, is a matter of the types of *concepts* we have of each kind of property. This difference does have implications for certain questions that might be termed “metaphysical.” But it will not immediately imply anything about the metaphysical *nature* of colors, or whether they are “objectively there.”

In order to bring out the distinction between primary and secondary qualities³ I will be considering some scenarios (some real, some hypothetical – some fancifully so) in which there is a question about how we would apply our color and shape concepts. My intention is to bring out features of the nature of our *grasp* of these properties; I will show that there is an important sense in which our color concepts are more tightly tied to perceptual experience than are our concepts of shape. The only content we can

² McDowell, who contends that colors are indeed “secondary properties,” is a notable and explicit exception to this pattern; he argues, as I will, that the sense in which colors are secondary properties should not lead us to the conclusion that they are metaphysically unreal. See, e.g., McDowell (2011).

³Or, more precisely, between color and shape properties; I will not be taking a stand on how far the point generalizes to other traditionally-defined primary and secondary qualities.

attach to color concepts is *experiential*, in a sense to be made clear in what follows; in the case of shape, the content of our concepts is not tied to experience in the same way.

The upshot of my argument will be that Campbell's Simple View of color is not sustainable, and that Stroud is wrong to insist that the connection between experience and color is no more fundamental or essential than the corresponding connection between experience and shape. More importantly, seeing where Campbell's account fails—and why Stroud is wrong to draw a parallel between the two bi-conditionals above—will help us understand the deep motivations behind the primary/secondary quality distinction. My hope is that in coming to appreciate that feature, we find a way to navigate the contemporary debate between internalists and externalists about perceptual content.

4.2. A CASE OF DISAGREEMENT

I begin with a true story. When I was growing up, my family used a common brand of garbage bags in the bin in our kitchen. I believed these bags to be black; so far as I could tell, they lacked any discernible hue whatsoever. My brother agreed. But, strangely, both of my parents said that the bags were green. At first I thought this could be explained by my aging parents' foggy sense of the present; perhaps they had used green bags when they were young, and now they just assumed that garbage bag colors were much as they had been in the 1950s. But, much to my surprise, when we brought the garbage bags out into the clear, direct light of the kitchen and examined them, standing side by side, my parents continued to insist that the bags—black as ever, to my eye—were in fact green.

At that age, I carried myself with a certain self-importance, and I had a tendency to discount the opinions of others (in many contexts). I was *positive* that my parents were being daft; I could *see* that the bags were black. In insisting they were green, my parents were merely revealing their emerging senility.

Today, I am far less confident about the color of those bags. In my old age, I feel I have gained a certain wisdom, an appreciation for the opinions of others, a modesty about the relative value of my own. So I am forced to consider the possibility that *I* may have been the one who was mistaken; perhaps the bags were green after all.

But what is striking to consider in thinking about this case, to my mind, is the question of *what*, precisely, I am now unsure *about*. The color of the bags, yes; but the further question is: what *sort of thing* is the color of the bags? What property is in question here? What possibility is it that I initially dismissed, but now consider to be still open?

My claim is that there is no way to spell out Campbell's "Simple View" of colors that will yield a satisfying answer to these questions. Colors are, on Campbell's view, simply transparent to us in perception. But there are cases where that transparency is called into question, where we are not sure if the link between color properties and

our experience of them can be relied upon. And in such cases, we need a way of *conceptualizing* just what it is we are questioning. What would it *be* for those bags to have been green, rather than black, despite my experience's having given no indication of their greenness? I have said that I now acknowledge that it is a genuine possibility that the bags were indeed green. The question I want to ask is: What must my concept of green be like in order to make that possibility genuinely intelligible, in the way it intuitively seems to be?

4.3. ERROR THEORIES

One way to respond to this question would be to deny that there really is any such intelligible possibility to worry about. An error theory about colors—of a certain stripe—gives this kind of response; but it should be noted that another type of error theory does not escape the demand for an answer. Here, again, we must distinguish between metaphysical and conceptual claims about color.

On the one hand, it could be argued that, though there is a coherent possibility alluded to in my thinking that the bags might really have been green, that possibility is metaphysically unrealizable, given the way the world in fact is. This type of theory might have been the kind held by Locke, who (in some of his moods) suggested that our attributions of colors to material objects mistakenly projected a property of our own “ideas” onto the world. On such a view, there could have been a world in which objects were indeed colored. But our world is not like that. In our world, objects have no colors at all; they merely have microstructural properties that tend to cause certain perceptual experiences in us (and these microstructural properties are not themselves to be identified with colors).

According to this error theory, which I will (somewhat hesitantly) call the Lockean View, our concept of colors *is* one that makes the possibility of objects' being colored *intelligible*. Objects might have been green, and yellow, and red. But (science has revealed) they are not. What I want to note here is that the Lockean View, despite its denial that objects in our world really are colored, still owes us an answer to my question from above: What *would it be* for the bags to have been green? The Lockean View says that no objects are really green, but, in so doing, it is saying that the claim that the bags were green—like any other claim about an object's being colored—is *false*. For a claim to be false, it must be meaningful; for it to be meaningful, it must be intelligible. So we must have some grasp of what it would be for the bags really to have been green, even if we go along with the view in denying that they (or any objects in our world) actually are. The Lockean error theory needs a way of spelling out what property it is saying the bags *lack*.⁴ The theory is subject to this demand because it is not

⁴ This line of thought closely mirrors some of Stroud's arguments against unmasking views. Stroud likewise demands that the error theorist have a coherent way of spelling out the possibility that she is rejecting, in denying that objects are colored. It is this challenge that Stroud ultimately finds fatal for the unmasking project.

challenging the commonsense idea that, at the *conceptual level*, we understand color properties to be ones that objects could conceivably possess. It is only a metaphysical error (not a conceptual one) that the Lockean View attributes to us: we mistakenly—but coherently—believe that objects really do have color properties.

We can contrast this metaphysical Lockean error theory with a conceptual version. On this second version, the entire question about the “true” color of the bags is badly formed. The Lockean View denies that the bags are green. This second version of the error theory does not so much deny the bags’ greenness; at a more basic level, it challenges the conceptual coherence of the possibility allegedly under consideration. On such a view, there are, perhaps, facts about *judgments* of color, and maybe also about *perceptions* of color, but there are no further facts about *colors themselves*, understood as properties of objects. It is not that the bags might really have been green, but are not; nor even that objects in general might have been colored, but are not. On this view, the very idea of a colored object is a conceptual confusion. No coherent possibility is described by suggesting that an object might really have been green.

This “conceptual” error theory—which I will (again with some trepidation) label the “Berkeleyan View”—can simply reject the question I want to consider about our concept of green. That question asked what our concept of green must be like in order for it to be coherent to suppose that the bags might really have been green; the Berkeleyan View says that there is no such coherent supposition to be made. According to such a view, my parents and I were not really having a substantive dispute about the bags’ color at all (not even one in which we both were wrong). We were simply in the grip of a conceptual confusion in supposing that it made sense to talk of the colors of objects.

The Berkeleyan View faces difficulties of its own.⁵ But for now I will simply note that the views put forth by Campbell and Stroud are not error theories of this (or any) kind; they are views on which it is not only coherent, but very often correct, to think that an object is green. So it would seem that on Campbell’s Simple View, we should indeed be able to make sense of the possibility that the bags were really green. Our *concept* of green needs to make intelligible for us what it would be for the bags to have been green.

4.4. ALTERNATIVES TO THE SIMPLE VIEW

So let us return to the question: What would it be for the bags to have been green? My charge was that Campbell’s Simple View could not provide an adequate answer to this question. In order to explain why, I first want to point out some of the possible answers that are not available to an advocate of the Simple View. As Campbell notes, the

⁵ These are spelled out at length in Stroud (2000); the basic problem is to say what a *judgment* about the color of an object—something the error theory needs to acknowledge the existence of, given its ambition of “unmasking” such judgments—might be, if the very concept of a colored object is a confusion.

Simple View maintains that greenness “is not a disposition to produce experiences in us”; it also denies that greenness is “a microphysical property.” By looking at the answers that each of these alternative views would give to the question about the bags’ greenness, we can achieve two goals: first, we can see what an answer to the question might look like, even if the particular answers given are unsatisfactory for one reason or another; and second, we can keep firmly in mind the tools that are *not* available to the Simple View in trying to answer the same question, so that we do not “sneak them in” to an answer on the Simple View’s behalf.

First, let us consider the idea that colors are particular microphysical properties of (the surfaces of) objects. In some of his moods, Locke actually could be seen as advocating this type of view. He suggests that colors are in fact no more than certain “textures” of objects, which reflect light in different ways and thereby cause various perceptual experiences in us. On this type of view, when confronted with the disagreement between me and my parents about the color of the bags, we could settle the dispute by analyzing the bags’ microphysical properties. If they turned out to have the “green texture” (or whatever microphysical property we would substitute in a modern version of the theory), then they would indeed be green.

This “microphysical” view of colors, whatever its flaws (and we will get to some below), would provide a very straightforward answer to the question I have been pressing. Green, on this account, is a microphysical property; my experience of green is generally caused by objects with this property. In this particular case, however, the normal causal relation does not hold for me (for whatever reason), though it does still hold for my parents. I can *conceive* of this possibility perfectly well because I know what it is for an object to have a particular microphysical property, and I can imagine that property occurring in an object that fails to cause a perceptual experience of green in me. That is how this account makes intelligible the possibility that I was wrong about the color of the bags.

Now this account may be unpalatable for any number of reasons, but I want to emphasize just one here. It relates to Campbell’s point about the availability of his Simple View “even to someone who rejected the atomic theory of matter.” Campbell’s reasons for stressing this point are connected to his goal of undermining the “background argument” that assumes a metaphysical picture on which “physics tells us what is objectively there.” But for my purposes, what is important to draw out from Campbell’s point is that, if we are attempting a *conceptual analysis* of color, the view we put forth really must be one that is available to a “naïve anti-microphysicalist.” For consider one of Democritus’s interlocutors, or someone in modern times who has not yet learned any particle physics. We do not want to rob such people of their color concepts; surely, they have color concepts just as we do. And yet, if we equate greenness with a microphysical property *at the conceptual level*, we will make it utterly mysterious how any but the most scientifically informed could ever possess color concepts.

It may be objected that an account that equates colors with microphysical properties is not offering a conceptual analysis at all; instead, it is simply putting forth as an

“empirical identity” something to the effect of “greenness = microphysical property X.” And with such an empirical identity, there is no need to assume that someone who possesses the concept employed on the left-hand side need also possess the concept employed on the right-hand side. The fact that someone can possess the concept WATER⁶ without possessing the concept H₂O is no reason for denying the truth of the identity “water = H₂O.” Similarly, the identification of greenness with some particular microphysical property is not threatened by the fact that someone can possess the concept GREEN without possessing any concepts of microphysical properties.

The objection is fair enough, as far as it goes. But if the account being put forth takes the microphysical property to be identical with greenness only in this “empirical” way, it still leaves unsettled just what greenness is *at the conceptual level*, as the term appears on the left-hand side of the identity claim. And, crucially, it does not help us provide an answer to the question about my disagreement with my parents about the garbage bags’ color. For it seems entirely possible that I could have gone through my reasoning about the situation, and could have come to wonder whether I was wrong about the bags’ color, without invoking any sophisticated concepts of modern atomic physics. The possibility that the bags were really green seems to be an intelligible one, just in virtue of the way the concept GREEN functions, without any need for empirical identities that might be unintelligible to many of us.

So let us stipulate that the story we tell about our concept GREEN must not make it mysterious how a scientifically naïve subject could possess such a concept. This does not rule out the possibility of *empirically* identifying greenness with some esoteric property of sub-atomic physics. And it does not prevent us from attributing *hidden structure* to the concept of greenness; it may in fact be quite obscure to the average possessor of the concept exactly what is “built in” to it. But what we cannot do is pretend that what is built in to the concept of greenness includes concepts that would be *utterly inaccessible* to many of the concept’s fully competent users. We want to be able to understand how I could have come to wonder about the real color of the bags, without having to assume that I happen to have a wealth of extremely complex microphysical concepts.

The dispositional account of color that Campbell mentions (and rejects) does not run afoul of this stipulation, and it does (on the face of things) provide the tools for answering our question about the color of the bags. On such an account, we would equate greenness with the tendency to produce perceptual experiences of a certain type in normal observers, in normal circumstances. This view would hold that the question about the bags’ color is a matter of what kinds of experiences normal human observers, in normal lighting conditions, would tend to have when looking at the bags.

⁶ I here adopt the convention of denoting a concept, as opposed to the object or property the concept is a concept *of*, by writing the relevant word in capital letters.

Like the microphysical account, the dispositional view would allow me to conceptualize the possibility of the bags' *really* being green, even though they did not appear that way to me. For we could bring in additional "test subjects," and if all (or most) of the color-normal subjects had perceptual experiences of the bags as green, that would *amount to* the bags' really being green (we would have to conclude, in such a scenario, that I was somewhat "abnormal" in terms of seeing green, despite my general ability to pass color-vision tests). Or perhaps everyone would agree with me, and my parents would be found to be the ones unable to detect the bags' true color. Or we might find that there was no clear consensus among normal perceivers; perhaps there would be a fifty-fifty split in people's experiences of the bags. The point is that each of these possibilities is intelligible. There might be some amount of *vagueness* in the question about the bags' color, on this view (if 51% of the normal perceivers experienced the bags as green, would that make them green? 52%?); but, however we decided to deal with the sticky question of boundaries and with looming sorites paradoxes, we would still know what *sorts* of possibilities were being considered when we considered whether the bags were really green.

My aim at this stage is not to advocate this type of dispositional view. My point is merely that it does satisfy the two desiderata we have formulated thus far. First, it provides an answer to the question of what it would be for the bags really to have been green. And second, it does not bring in any concepts that a scientifically-naïve subject would lack. The idea that an object has a tendency to produce a certain type of experience in normal observers is one that our naïve anti-microphysicalist could easily understand; so analyzing the concept GREEN in this way does not violate the stipulation we made above.

4.5. THE CHALLENGE: WHAT IS GREENNESS?

Let us now turn to the issue of how Campbell's Simple View could spell out what it would be for the garbage bags to have been green. I first want to acknowledge that the Simple View does not straightforwardly make the question about the bags harder to *gather evidence about* than any other view. The microphysicalist view described above would attempt to answer the question by analyzing the microphysical structure of the bags' surface. The dispositional account would attempt to answer it by bringing in more subjects to report on their own experiences of the bags' color. Nothing about the Simple View rules out using such methods to help us investigate questions about the colors of things (though it might be a bit mysterious *why* such tests—especially the microphysical ones—would be relevant to the question).

But these tests, though they might in some way provide *evidence* about the color of the bags on Campbell's view, could not, in the end, give us the real answer to our question. The question I am pursuing asks what it would *be* for the bags to have been green. For Campbell, greenness is *not* a microphysical property. Similarly, greenness is

not a dispositional property; it is not simply a matter of causing certain kinds of experiences in normal human observers. What it *is* for objects to be green is not captured by describing the various tests we might conduct in order to determine the bags' true color. And, since the basic test we have for greenness—our perceptual experience, to which greenness is, according to Campbell, generally “transparent”—issued an ambiguous verdict in this case (within my own family, there were two experiences of greenness, two of blackness), we would seem to need some conception of what it *is* for an object to be green, independent of those experiential tests. We need to know what question the evidence we are gathering is supposed to help settle; we need a conception of what it is that our respective perceptual judgments are in disagreement *about*. It is this conception that I think Campbell's view fails to provide.

I want to emphasize, once again, that the question I am asking is about our concepts, and our understanding, of colors. It is not a verificationist demand: I am not complaining that Campbell's view makes the question of the bags' color “undecidable.” That, on its own, would pose no special problems for the view, by my lights. If there is a problem of decidability in the case I've described, it will exist for just about any view: the dispositional view I described above, for example, makes reference to the experiences of subjects, and it seems plausible that such experiences would be “private” in a way that does not admit of strict “verification.” The challenge I am posing for Campbell's view is not to spell out how we would *know* or *decide* whether the bags were green. The challenge is to spell out what it would *mean* for an object to be green, what it is that we would be testing for in analyzing surface structures or conducting other investigations. A verificationist answer to this question is possible: one could choose to identify greenness with the disposition to register as “green” on a given test (perhaps showing up in a certain bandwidth on a spectrometer). An anti-verificationist answer is also possible: greenness could be identified with the property of being disposed to produce certain “private experiences” that are constitutionally unverifiable to outside observers. Perhaps neither of these answers is particularly appealing, but they do both address the challenge I am posing: They say what it would *be* for an object to be green, by offering an account of our *concept* of the property. On either account, we would have a perfectly good grasp of what is in dispute between me and my parents when we disagree about the bags' color. My claim is that no such understanding is forthcoming on Campbell's picture.

Let me pause briefly here to say something about the dialectical situation. The points made thus far do not, it must be acknowledged, amount to a true *argument* against Campbell's Simple View. At this stage, it is still open to Campbell simply to insist that there is no need to offer an account of what it would be for the bags to have been green, beyond saying simply that they would have *been green*. That is, Campbell can resist the demand for a *further* account of what greenness amounts to. His view is a *simple* view. It says that greenness is, simply, *greenness*. That is the property my parents and I were disagreeing about. What it would be for the bags to have been green is for them to have had *that very* property. There is no more to say than that, and demanding

a further answer is—in asking for an account of greenness as anything other than an irreducible, unanalyzable, basic property—merely begging the question against a view that is explicitly one on which colors are *simple*.

That is all fair enough. The case discussed above is not intended to be a knock-down argument against Campbell's view. What it *is* supposed to do is to (begin to) bring out a tension I see in the Simple View, by questioning whether we really do have a grip on colors as simple properties in the way Campbell describes. It does so by trying to drive apart what would, on Campbell's view, be the *color itself* (the simple, categorical property of greenness) from other properties typically associated with color properties—their dispositions to cause certain experiences in us, and their connection to certain microphysical properties—which on some theories factor crucially into an analysis of color. My goal was to show that if we strip away these associated properties from the color property itself, as Campbell's view suggests we should be able to, it can begin to look as though the “residue”—the simple color property—is not something of which we have any concrete concept. Without microphysical properties and without the connection to experiences in play, we don't have a contentful idea of “greenness itself” at all.

4.6. ANOTHER CASE OF DISAGREEMENT

In order to bring this out more clearly, I want to contrast the type of scenario described above—one in which subjects disagree about the color of an object—with one where there is a disagreement in perceptual judgments about *shape*. I will show that there is an important difference in our grasp of what it is that is in dispute in the two types of scenarios – a difference that helps reveal the different types of concepts we have of primary and secondary properties.

Consider the following scenario. Russian scientists discover an object in a lake beneath the ice of Antarctica. It is dried out and brought to a museum, where it is displayed in a room flooded with natural light. As people walk by the object, a strange phenomenon occurs. An elderly couple, stopping to look at the object, have the following conversation:

S₁: My, what a beautiful circular artifact! And such a striking shade of green!

S₂: Circular? Green? We must get your eyes checked! That object, though beautiful indeed, is perfectly square and red as blood!

This curious dialogue is not an isolated incident. In fact, precisely half of the museum-goers who see the object describe it as green and circular; the other half insist it is red and square. We have before us now two questions, one about shape, one about color:

- 1) Is the object circular, square, or neither?
- 2) Is the object red, green, or neither?⁷

I will show that there is something fundamentally different about these two questions, and that, on Campbell's view, we simply have no way of explaining that difference. The issue is a deep one. There are some ways of presenting the primary/secondary distinction that would suggest the difference is merely a matter of *how we would investigate* the respective questions. On such views, the difference amounts to facts like these: we can use a secondary sense modality (touch) to settle (1), but not (2); there are more wide-ranging effects hinging on the answer to (1) than there are for (2) (e.g., whether the object could be used as the wheel of a bicycle). But these are not the kinds of facts that I will appeal to. Instead, I want to consider a case in which *no* test takes us any further than the initial situation, for either question: Whatever test we think to run on the object, half the population insists the results indicate squareness (redness), half that they indicate circularity (greenness). My claim is that, when all such tests fail to answer the question, there is still a clear sense in which (1) is a *genuine* question, while (2) begins to look more and more unreal—a question which, if we follow Campbell in taking a simple view of color properties, lacks any real content.

I begin by noting that, on Campbell's view, we can imagine this scenario taking place in a world that actually is as a naïve anti-microphysicalist takes it to be. Imagine a world that has colors, as ours does, and inhabitants who perceive them, as we do. But, unlike in our world, where colors have some close relationship with the atoms that make up matter, in this world there are no atoms at all. The inhabitants still see red objects, and know they are red, and grasp the full nature of the property of redness; on Campbell's view (and this is one aspect of the view I think is quite right), the further difference in microphysical properties (that our world has such properties, while their world lacks them) has nothing to do with that understanding. Thinking about this type of world will help us focus in on the crucial conceptual issue, without being distracted by issues connected to the microphysical “supervenience base” of colors that are irrelevant to the Simple View itself.

So suppose we find ourselves in such an atom-less world, with the strange Antarctic object on display. In such a world, it is hard to know what we could do to investigate question (2). Perhaps we would observe the object in different kinds of lighting conditions, or display it in front of animals that are known to react differently to red and green objects. But suppose these tests remained equally unhelpful: whatever the lighting, the object continues to look obviously green to half the population (and half the animals), and obviously red to the other half. The question I want to press in this case is not simply “What else could we do to determine the true color of the object?”

⁷ I'm assuming that “both” is not an option in either case, for simple conceptual reasons: both a round square and a greenly red object are conceptual contradictions.

Again, the crucial question to ask is “What do we even *mean* by ‘the true color of the object’? What would it *be* for this object truly to be green rather than red?”

My suggestion is that, *if we restrict ourselves to the resources allowed under Campbell’s Simple View of colors*, we will simply run up against a conceptual wall here. We know that half the population experiences the object as red, half as green. But the further question of which half of the population is *correct* has begun to seem empty. In order for it to be meaningful, we would have to be able to conceive of two genuinely different possible worlds, both exactly alike in all the facts about people’s perceptual judgments about the object’s color, but which differ with respect to the fact of the object’s color itself. This further fact cannot, on Campbell’s view, be a matter of any facts about the object’s microphysical structure (since the worlds we are considering here lack any microphysical structure). The problem is that, even from “a God’s-eye view,” there doesn’t seem to be any sense in which the two worlds in question really are *different* worlds at all. There would seem to be nothing to genuinely distinguish them, nothing to make it the case that we were in the “red” world rather than the “green” one (or vice versa). When separated from both facts about microphysical structure and facts about perceptual experiences, the concept of color seems to be simply empty.⁸

We can contrast this question about our understanding of what it would be for the object really to have one color rather than the other with the analogous question about shape. There would, presumably, be more tests we could imagine doing to determine the *true* shape of the object, having already established the (non-decisive) facts about people’s perceptual judgments. We could let the museum-goers *feel* the object, in addition to seeing it; we could attach it to a bicycle and see if it spun on a road.⁹ But it is always conceivable that these tests would fail to settle the question: the square-seers could all be square-feelers, the circle-seers all circle-feelers. It is somewhat harder to imagine the more elaborate tests failing to settle the question, but there is no real conceptual impossibility here: Suppose that whatever test is run, half the population insists that the outcome shows circularity (“The bicycle rolled down the street just fine!”), half that it shows squareness (“The bicycle didn’t move an inch!”).¹⁰

⁸ It is important to note here that I am not denying that there is, in fact, a further question about an object’s real color, when perceptual judgments disagree. What I am claiming is that, when we follow Campbell in taking a Simple View of color, denying that there is a conceptual connection between color and either color experience or microphysical properties, we are unable to make sense of there being such a further question. In this section, I am investigating what happens when we do take Campbell’s Simple View on board. It is only when we accept such a view that the further question about the object’s true color seems empty. My claim, in the end, is not that this shows there really is no further question; rather, it is that this result gives us reason to reject Campbell’s Simple View.

⁹ Compare Campbell’s suggestion along similar lines in his (1993), and Bennett’s discussion in his (1971).

¹⁰ Alternatively, we could cut off the tests by adding a further feature to the mysterious object: if it is moved from its display case, it will immediately combust. So seeing it and touching it

Crucially, however, there would never come a point when the question “Is the object *truly* square or circular?” ceased to *make sense*. If the tests have all failed—if no empirical test we could even conceive of could settle the question—we would still know what we *meant* by “the true shape of the object.” And that is because we can spell out what we mean by the question in completely experience-independent terms: For the object to be truly circular is for all the points on its edge to be equidistant from its center. I have a robust conception of what that means, and, however unverifiable the *truth* of the claim that this particular object is circular might be, I am in no danger whatever of losing my grasp on the *meaning* of the claim itself. I know what it *would be* for the claim to be true: In contrast to the case above concerning color, I have a clear understanding of two genuinely different possible worlds, alike in all facts about perceptual judgments, but differing in the further fact of the object’s shape itself. These two worlds differ in *how the object occupies space* within them;¹¹ and that notion is fully intelligible even when, as in this case, the object’s manner of occupying space has been severed (at least in part of the population) from the perceptual experiences typically caused by such objects.

This, I think, goes to the heart of the primary/secondary distinction. That distinction is not a matter of multiple sensory paths to knowledge (in the case of the primary qualities) versus a single sensory path (in the case of the secondary qualities). It is not a matter of a property’s playing an appropriately wide role in the causal structure of the world; nor is it, as is sometimes thought to follow from the question of “playing an appropriately wide causal role,” a matter of whether a given property is “metaphysically real.” It is a matter of the nature of our *grasp of the meaning* of the terms. For primary qualities, that grasp allows us to understand which property is in question no matter what may be going on with our (or anyone else’s) perceptual access to it. For secondary qualities, our grasp of which property is in question goes only as far as our grasp of the perceptual effects of that property. This crucial difference is a difference in the nature of our *concepts* of greenness and of circularity.

I have suggested that we can see this difference in our ability to give, from a pre-scientific perspective, a contentful *definition* of the primary qualities that makes no reference to perceptual effects of that property (for CIRCULAR, that definition is “having all points on its edge equidistant from its center”),¹² while no such definition is possible for the secondary qualities. But it may be objected that we *can* give such a definition for the secondary qualities. Blue, for example, can be defined as the color

are the only tests that can be done. The point is that the difference in question seems to me to be one that transcends any facts about the verifiability of claims about color, on the one hand, and shape, on the other. If we imagine a case where the two questions are equally undecidable, the contrast I want to draw out will be easier to bring into view.

¹¹ This way of formulating the notion of shape in experience-independent terms echoes McDowell (2011).

¹² Here again we could also say something along the lines of McDowell’s formulation of the concept of circularity as that of an object’s “occupying space in a particular way.”

that is opposite yellow in color “space.” This definition is not independent of other color terms, to be sure, but neither is the definition given above of “circular” independent of other *spatial* terms (it requires terms like “point” and “equidistant”).

I think there is something to this objection, in a limited sense. We might indeed be able to give interdependent definitions of color terms; and it might seem that that is all we are doing with the shape terms, as well. But in considering the contents of these different definitions, it seems to me that we cannot “latch on” to the color definitions, without appealing to experiences, in a way that lets us understand them as genuinely contentful. That is, we could know that yellow, whatever it is, is opposite blue, whatever *it* is, in some metaphorical “space.” But there is nothing more to be added (except the positions of red and green, and perhaps some facts about the “asymmetry” of humans’ color “space”).¹³ This “color space,” described in these terms, is indistinguishable from many other sets of features that similarly map to an abstract space of two pairs of “opposites.” So we need something else to give these definitions real content; and, unlike in the case of shape concepts, that content can come only by making reference to experience itself.¹⁴

Campbell might suggest, here, that it is precisely the *intrinsic, categorical* properties that colors *are* that supply the needed content. But what my example of the mysterious Antarctic object is supposed to bring out (building on the idea initially hinted at in the example of the garbage bags) is that we don’t have as firm a grip on those intrinsic properties as he supposes, if we think of them as properties of objects independent from their effects on human perceivers. When we cannot appeal to effects on perceivers in explaining the intrinsic nature of the property—because we have a case where those effects offer no answer to the question of which property the object has—we are left with no firm sense of what the property *is* at all. This suggests that our conceptual grasp of color properties is fundamentally tied to perceptual notions; without those perceptual notions in play, we simply have *no* conceptual grasp of color.

In the case of shape, we are not so badly off: even in a scenario where perceptual notions give us no clue about the true shape of an object. The content of the claim “That object is circular” can be spelled out in a way we understand, even when we acknowledge the impossibility of verifying its truth by means of our senses. That is because the “space” we use to understand the concept CIRCULAR just is our notion of *literal space*. Our notions of the structures we are describing when we say that the points on a circle’s edge are equidistant from its center are some of our most basic modes of understanding; they spring from the *a priori* spatial proto-concepts I described in Chapter 3. We do not understand spatial features like points and distance by reference to our experience of the relevant types of features (edges, equidistance,

¹³ For detailed discussion of some of these potential complications, see Hardin (1998).

¹⁴ In terms of the discussion of Chapter 3, color space is a merely metaphorical space, and so, without having a grip on the intrinsic features of its “points”—the hues, brightnesses, and saturations of the individuals colors—we cannot get a grip on the “space itself,” since that space is merely a way of representing the intrinsic features.

circles, points); rather, we can conceptualize our experience as instantiating those features because we already *have* a grasp of space.

The key point to take away from the discussion at this stage is that our understanding of primary qualities like shape is *a priori*, while our understanding of secondary qualities like color is not. The former we can define and understand without reference to experience; with the latter, we cannot. The reason this is significant is that, on Campbell's view, there is supposed to be an equivalent conceptual "distance" between, on the one hand, green's dispositions to produce experiences in us and the property of greenness itself, and, on the other, traditional "primary" qualities, like circularity, and the dispositions of *those* properties to cause certain experiences in us. In the case of a shape property like circularity, it seems quite right to say that the experiences the property is disposed to produce are not to be *identified* with the property itself – and, correspondingly, that facts about what experiences the object is disposed to produce would not *constitute* an answer to the question of an object's circularity. Such facts might in practice generally give us more than enough reason to draw a firm conclusion about an object's shape; but, as we can see in thinking through the example of the mysterious possibly-circular artifact, we maintain our understanding of *what it would be* for an object to be circular quite independently of any facts about the experiences the object produces. In thinking through my dispute with my parents about the color of those garbage bags, by contrast, we can begin to see that the relation between a secondary property like greenness and the experiences it is disposed to produce differs in a fundamental way from the *non-constitutive* relation between our grip on a primary property like circularity and our experience of that property. In the case of greenness, the conceptual distance between the experiences produced and the property itself is simply lacking – there is a fundamental connection between our concept of greenness and our experience of it.

We could attempt to spell out this tighter connection in terms of the dispositional account of color I sketched above. On that view, we can understand a dispute about the *real* color of an object, when people disagree in their perceptual judgments, as amounting to a disagreement that itself is to be cashed out in terms of further facts about perceptual experience (specifically, facts about whether normal observers would have a certain type of experience when faced with the object in normal circumstances). That is why the fact about the color of the object cannot completely come apart from the facts about perceptual experiences of the object: facts of the first kind simply *are* facts of the second type. So if the strange Antarctic object really does strike exactly half the color-normal human population as green, half as red, this view would be forced to conclude that it is in fact neither green nor red. And, while I do not want to advocate such a view at this stage, I think it is worth acknowledging that there is at least something intuitively plausible in this verdict. In a situation such as this, to insist that there must be some *real* answer about the object's color, over and above the known facts about its striking half the population as red and half as green, would seem

unmotivated. What could this further fact amount to?¹⁵ This is the conceptual wall we ran up against in trying to describe two distinct possible worlds as the “red world” and the “green world” above.

And again, we should keep in view the contrasting case of shape. Knowing that half the population judges the object to be circular, the other half square, does nothing to lessen the impression that there is some concrete fact of the matter about the shape of the object itself. The points on the object’s edge are all equidistant from its center, or they are not. There is no temptation here to think that the facts about our perceptual experiences exhaust the facts about the object’s shape properties; there is simply too obvious a conceptual distance between the categorical property of shape—a property on which we have an *a priori* grasp—and the perceptual experiences we have of shape properties.

4.7. REVERSING THE CONCEPTUAL CONNECTION

In the previous section, I attempted to spell out a conceptual difference between shape and color properties, in terms of those properties’ relation to the experiences we have of them. In doing so, I was posing a challenge to the view that there is no important difference between properties traditionally classified as primary and those classified as secondary. That view is expressed by Stroud when he dismisses the notion that the secondary qualities—in contrast to the primary qualities—should be thought of as *dispositional*. In the introduction, I spelled out Stroud’s claims in terms of two biconditionals, one involving color and the other shape:

- (1) X is yellow \leftrightarrow X is such as to look yellow (in normal circumstances, to normal perceivers)
- (2) X is rectangular \leftrightarrow X is such as to look rectangular (in normal circumstances, to normal perceivers)

An account of secondary qualities that maintains that such properties are significantly different from primary qualities will claim that there *is* an important difference between the biconditional expressed in (1) and that expressed in (2). This difference is not a difference in the truth of the biconditionals; surely, it is perfectly true not only that yellow objects are such as to look yellow to normal perceivers, but that rectangular

¹⁵ It might be thought that what the further fact amounts to is some fact about the physical character of the object’s surface, or the physical character of the light reflecting off of it. While I think there is something to this line of thought, it is not one that the Simple View can avail itself of. Remember that on Campbell’s view, colors are explicitly *not* to be thought of as physical properties of object’s surface; and, in the version of the scenario we are considering here, the object exists in a world that simply *has* no microphysical surface properties.

objects are such as to look rectangular. The difference the advocate of the primary/secondary distinction has in mind is rather a difference in the *status* of the truth expressed in (1), as compared to the truth expressed in (2). This difference in status is somewhat hard to articulate. Stroud glosses the idea—which he himself rejects, in rejecting the distinction between (1) and (2)—by saying that the advocate of the primary/secondary distinction takes (1), but not (2), to be in some sense an *analysis* of the property on the left-hand side of the biconditional:

[Biconditional (1)] is meant to tell us (as it is sometimes expressed) ‘what it is’ for an object to be yellow, or what is ‘essential’ to its being yellow, or what a thing’s being yellow ‘consists in’, in the sense of what ‘x is yellow’ means when said of an object or what we believe when we believe that an object is yellow. These different expressions of the point of the “analysis” are perhaps not all equivalent, but each is to be understood as involving some claim of necessity between the colour of an object and its disposition.¹⁶

So Stroud’s claim, in his rejection of the “dispositional” account of color, amounts to something like this: (1) does not express an *analysis* of yellowness, any more than (2) expresses an analysis of rectangularity. Both biconditionals are true, and in approximately the same way. There is nothing particularly *conceptual* or *necessary* about the truth expressed by (1) to differentiate it from the truth expressed by (2).

Given such a lack of any conceptual or necessary connection between a property and our experiences of it, it should always in principle be possible to conceive of a situation in which the two come apart. And it should always be clear that a question about whether an object really has the property is a *further* question, one whose answer is not itself determined by the facts about the object’s dispositions to cause experiences. If, by contrast, we think that there *is* a necessary or conceptual connection, situations in which the property and its dispositions come apart will *not* be genuinely conceivable. If being a certain color simply amounts to having a disposition to cause experiences of a particular type in us, then there is no sense to the idea of an object that is that color but fails to have the relevant disposition. Likewise, if being a certain color amounts to having a disposition to cause experiences of a certain type, then there is no *further* question to be asked about an object’s true color once all the facts about its dispositions are known. There would be no way to conceive of two separate possible worlds in which the answer to the question about the object’s color was different, even though the facts about the object’s dispositions were the same.

I have been suggesting that something like this barrier to conceivability in fact holds for colors – or, more precisely, that it would hold, if we cut ourselves off from microphysical or dispositionalist accounts of color in the way Campbell’s Simple View

¹⁶ Stroud (2004, p. 426).

does. I imagined scenarios where the facts about an object's dispositions to cause perceptual experiences did not neatly conform to either the "green" or the "not green" pattern, and I suggested that, in such scenarios, it becomes mysterious what it would be for the *true* color of the object to be green (rather than red or black). I also suggested that an account of color that *analyzes the concept* of greenness in terms of its dispositions to produce experiences in us can explain why there would be no further question about the object's true color, once the facts about its dispositions to produce experiences are given.¹⁷

In suggesting that the biconditional expressed in (1) is no more a conceptual analysis than that expressed in (2), Stroud would seem to be required to deny that there is any necessary or conceptual link between an object's color and its having the relevant disposition. That is, it would seem that on Stroud's view, there should indeed be a further question about the object's *true* color, even given all the facts about what dispositions it has to cause experiences. This idea of the further question is what I have been pressing in the cases described above. In each case, I gave facts about an object's dispositions to cause experiences, and then wondered whether we could—on a view like Campbell's—make sense of the fact of the object's true color as a *further* fact beyond those dispositional facts. If there is no necessary or conceptual connection between an object's colors and its dispositions to cause experiences, then it seems there should indeed be a further question about the true color of an object, independent of the facts about the perceptual experiences caused by that object. It was the difficulty of making this further question intelligible on a view like Stroud's or Campbell's that I tried to bring out above.

So the problem for the Simple View is that it seems to be subject to two incompatible pressures. On the one hand, its insistence that colors are not to be *analyzed* in terms of dispositions to produce experiences, nor in microphysical terms, leaves us with something of a mystery about what the remaining "residue" of the simple property of greenness is supposed to be, once those associated facts are removed from the picture. On the other hand, the very fact that colors are not to be equated with dispositions to produce experiences means that there should be a real conceptual possibility of colors coming apart from color experience, and that questions about the true color of an object are not exhausted by facts about its dispositions. So the view both acknowledges the possibility of certain scenarios where color and color experience

¹⁷ Again, an important caveat: I actually *do* think there is a further question to be asked in the types of scenarios I've described. It is just that I don't think we can bring that question into view using the resources that Campbell's account allows itself. (See fn. 8 above) That is, the way in which I think there is a further question depends crucially on accepting that color is not "simple" in the way Campbell's account suggests, and instead is connected both to dispositions to cause experiences in us and to physical facts about the surfaces of objects, in a way to be spelled out below.

come apart, and at the same time makes it hard to spell out what those possible scenarios amount to (and, thus, what it is that we are asking about when we ask about the true color of the object itself).

One way to get out of this difficulty would be to argue that the alleged possibilities are not real possibilities at all. If color and color experience cannot come apart in the way described in the scenarios above, then there would be no need for the Simple View to offer an account of what possibilities are in question in those scenarios. Both Stroud and Campbell, in the end, take something like this route. They deny that there is a real possibility of colors and their perceptual effects coming fully apart. Instead, they maintain that there *is* a sort of conceptual connection between green objects and the particular type of experiences those objects tend to produce. But, whereas the dispositionalist thinks this conceptual connection is a matter of the very concept of greenness being spelled out in terms of the tendency to produce a particular type of experience (which is what taking biconditional (1) as a *conceptual analysis* of yellowness amounts to), Campbell and Stroud suggest that the conceptual connection runs in the reverse direction. It is not that the very concept of greenness is the concept of an object disposed to cause “green experiences”; rather, it is that the very concept of a “green experience” is the concept of an experience *brought about by green objects*. The only way for us to conceptualize certain experiences *as* experiences of the particular type we call “green experiences” is for those experiences really to be experiences of the particular color that is greenness (that is, episodes of a subject’s being perceptually aware of a genuinely green object). That is what makes it the case that they *are* experiences of that type. We cannot independently identify some class of experiences as “green experiences”—perhaps picking them out by their “intrinsic phenomenal character”—and then wonder whether, in a particular imagined scenario, the objects that cause such experiences are green. Rather, we are only entitled to think of those experiences as “green experiences” at all if we already conceive of the objects causing the experiences as *green objects*.

In the next two sections, I want to investigate this proposal in some detail. Before I do, I first want to quickly point out two important features of this way of spelling out why there *is*, after all, something of a conceptual connection between colors and our experiences of them.

The first is that, even if we accept something like Campbell’s and Stroud’s accounts of how it is that “green experiences” can be identified, and we acknowledge that this puts a limit on how far greenness and experiences of that type can come apart, it is hard to see how such an account deals with the types of scenarios I described above. In those scenarios, the relevant background for identifying an experience of a certain type as a “green experience” is already in place: both the real situation with my family’s disputed garbage bags and the imagined one involving the mysterious Antarctic object occur in worlds in which there are many green objects that everyone experiences as green, with no inter-subjective disputes. So it is plausible to think that even on Campbell’s or Stroud’s account, we’d have a way of picking out the general class

of experiences we label “green experiences” in those worlds: we’d identify that class of experiences as the type people have when viewing uncontroversially green objects.¹⁸ When we then have a dispute about the greenness of a *particular* object because some people are having that type of experience and some are not, we still seem to require some way of conceptualizing what further fact is in dispute (beyond the acknowledged mismatch in the types of experiences had by the various subjects) in the dispute about the object’s *true* color. It’s not clear that an account like Stroud’s or Campbell’s can rule out the problematic scenarios arising on particular occasions, even if their accounts do rule out a totally general possibility of greenness and “green experiences” coming apart. So it doesn’t seem as though positing a conceptual connection between color and color experience in *this* way would allow Stroud or Campbell to reject the scenarios I described above—scenarios in which color and color experience come apart—as simply conceptually impossible.

The second important feature of the account of the conceptual link between colors and experiences of color given by Stroud and Campbell is that the account is not in any way specific to color properties, nor even to secondary properties generally. The idea of the account is that the type or character of an experience or set of experiences cannot be identified independently of what the experience is caused by. But this should be equally true for primary qualities: if there is a problem about categorizing and conceptualizing experiences by their character—independent of facts about what causes those experiences—then that problem should be just as sharp when it comes to primary qualities like shape. And so, on Campbell’s and Stroud’s account, whatever conceptual barriers we run into in trying to pull apart color and experience of color should be equally forceful in trying to pull apart shape and experience of shape.¹⁹ I have already suggested above that that is not the case—that we do in fact seem to be able to conceive of scenarios where the question about an object’s true *shape* is entirely independent of the facts about its dispositions to cause experiences. This divergence between color and shape properties—in terms of how well we can conceive of the properties in question coming apart from their perceptual effects—is not explained by the account Stroud and Campbell give of the conceptual connection between properties and our experiences of them. If there is a conceptual connection between color and color experience along the lines Stroud and Campbell propose, then the conceptual barrier we run into in trying to pull apart color and color experiences should be every

¹⁸ Stroud at one point suggests that identifying experiences as “green experiences” is indeed possible in this way, once we acknowledge that there are experiences of genuinely green objects (see Stroud (2000, p. 178). It is unclear whether this acknowledgement is really Stroud’s considered view, or just a concession he makes for the sake of argument in his discussion of dispositional accounts of color.

¹⁹ Stroud and Campbell would both, I think, be happy to accept this verdict: they both think it is not possible to identify or characterize perceptual experiences as belonging to some particular type without making reference to the external world objects (or facts) those experiences are experiences *of*.

bit as firm when we instead try to pull apart shape and shape experience. The fact that this seems not to be the case (i.e., the fact that we can quite intelligibly pull apart questions about objects' true shapes from facts about our experience of those shapes) suggests that there is no conceptual connection in the case of shape. Below (in §4.9), I will bring this out further by exploring Stroud's claims about the parallel between biconditional (1) and biconditional (2), in light of his account of the way that experiences are conceptually linked to the properties they are experiences of.

4.8. ROLOC

In the previous section, I noted that denying that biconditional (1) is an *analysis* of yellowness, as Stroud and Campbell do, makes it seem initially that color and color experience should be able to diverge in a way that they do not, in fact, seem able to diverge. Campbell is actually explicitly sensitive to this kind of worry, acknowledging that an objection could be made to his view by claiming that it creates an unnatural distance between colors and our experiences of them. And, as mentioned above, Campbell does maintain that there is an important connection between a color and our experience of it. He spells this out by saying of a color property like greenness that “the real nature of the property is... transparent to us” in perceptual experiences of green. Green objects look green to us, and in so looking, reveal the nature of that particular color property. That is how we have a conceptual grasp of color properties.

It might be thought that an “inverted spectrum” scenario would put pressure on this view. For if my experiences of green objects were of a different kind from yours—if, say, green objects looked to me the way red objects looked to you—it could hardly be the case that the very nature of greenness was revealed to *both* of us in our respective color perceptions. For at least one of us, transparency would have failed, and the conceptual distance between color properties and our experience would make it unclear whether we had any conceptual grip on greenness at all.

But Campbell has a response to this line of objection: he denies that there really is any possibility of green objects' looking different to me and you—at least not as a general rule. The inversion hypothesis implicitly assumes that we can latch onto various “color experiences”—or “the way objects look” to various subjects²⁰—independently of the colors in the subjects' environment that cause those experiences. Instead, Campbell suggests,

²⁰ I have spelled out the inversion scenario in terms of green things' looking to me the way red things look to you, which is the most intuitive way to describe it, by my lights. In doing so, I slipped back and forth between speaking of “how things look to us” and “the types of experiences we have.” This move is certainly not uncontroversial: There are many competing theories of the nature of the relation between X's looking F to S and S's having an experience of type G when viewing X. But, for now, I think that the move between the two will not create any substantive problems and will allow us to frame the crucial issues more clearly.

what constitutes experiences being experiences of the particular colours they are is their being responses to just those features of the environment. Of course, it is not that illusion is impossible. It is rather that an experience's being an experience of a particular colour depends upon the subject's being able to use his colour vision to track that particular colour. So there is no possibility of setting up alternative causal hypotheses to explain colour vision [i.e., hypotheses on which, say, green things cause "red experiences" rather than "green experiences" (for some, or all, subjects)]: they simply bring with them changes in the characterization of the experiences to be explained.²¹

So Campbell's reason for maintaining that the nature of greenness can be revealed to us in perceptual experience, even though colors themselves are categorical properties independent of experience, is that the character of experiences depends on the categorical properties causing them. Occasional illusions are possible, but on the whole, green objects must *look green* (they must cause "green experiences"). This is because what it *is* for an experience to be a "green experience"—what it is for it to have the particular character that, say, my experience of looking at a cucumber has—is for it to be the type of experience caused by greenness. We need not worry that the nature of greenness might fail to be revealed in our experience because our experiences might be "inverted" relative to the actual colors of things. There is no possibility of our having "green experiences" when faced with red objects, no worry that green objects look to me the way red objects look to you. All of us, if we perceive the color of green objects, have experiences of the same type. The character of those experiences is tied, at the conceptual level, to the character of the green objects they are experiences of.

This is the "reverse" conceptual connection between color and color experience that I spoke of in §4.7. Instead of claiming that a given color is conceptually dependent on color experiences of a certain type, as the dispositional view of color does, Campbell claims that the character of a given experience type depends, at the conceptual level, on the particular colors the experiences are caused by.

It is crucial to note that this line of thought puts a lot of weight on there being an unshakeable connection between an object's being a particular color and its looking a certain way (not always, but as a general rule). This connection is not merely a contingent causal one; it stems from the way we analyze the concept of a "green experience" on Campbell's view. There is only one type of visual experience that could conceivably be caused by greenness, one particular character that such experiences can have; any experience that is an experience of greenness necessarily has this character because that is what it is for it to be an experience of greenness – it is for the experience to have the character of *greenness itself* presented to one. It is this connection that explains how we can have a grasp of the nature of color properties on the Simple View.

²¹ Campbell (1993, p. 262).

I do not think that there can really be this kind of tight connection between the character of perceptual experiences of greenness and greenness itself. In order to bring this out, I want to develop one further thought experiment in which we can dissociate looking green from greenness. If that is indeed possible, it will pose a problem for Campbell's view, because it will show that there is no guarantee at the conceptual level that the character of our experiences of green gives us access to the nature of greenness. The conceptual connection Campbell proposes between greenness and the character of our "green experiences" will have been shown not to hold.

We can see the possibility of dissociating looking green from greenness itself by considering a fictional planet, Roloc, that is a version of Hilary Putnam's Twin Earth. Although all the objects on Roloc are exact molecular matches of ours, the light rays produced by Roloc's sun are very different. Because of this difference (which also affects all of Roloc's artificial light sources), the cucumbers on Roloc, though composed of the same stuff as ours, reflect light in a bandwidth identical to that of the light reflected by ripe tomatoes on Earth; similarly, all other object-reflectance relations are inverted on Roloc relative to Earth.^{22 23}

We might be tempted to say that what this means is that green objects on Roloc look to the Rolocians the way red objects look to us on Earth, and vice versa. But that would be a bit hasty in this context, where part of what is in question is what it *is* for an object to look red to a given person (or population). So for now, I will refrain from such speculations about how to make comparisons concerning how things look to Earthlings and to Rolocians (though I do, in the end, think that the initial claim—that red objects look to the Rolocians the way green objects look to us—is how the situation should be described). Instead, I will consider the following scenario: We Earthlings one day in the distant future discover Roloc and land a team of scientists on its surface. These scientists look around and see what seem to them to be blue lemons, yellow #2 billiard balls, green tomatoes, and red cucumbers. They meet the native Rolocians and discover that their language is much like our English, except that they

²² Roloc bears some similarity to Ned Block's Inverted Earth (see Block (1990)), but it is importantly different. On Block's planet, light sources and object reflectances are not different from those on Earth. Instead, objects themselves have different physical features: lemons have a surface-structure that reflects *normal* light in the "blue bandwidth," and are therefore (according to Block, at least) themselves blue lemons. The lemons on Roloc, by contrast, are exact molecular duplicates of Earth lemons, and so are (by Block's lights, and by mine) yellow in color.

²³ This scenario, as described, would run into some difficulties due to the particularities of our visual system, and its "asymmetries" with respect to various colors. While these empirical issues may present some problems in spelling out the details of the case, it seems those problems are likely not insurmountable (though red/green is not a perfect symmetry in human color "space," there are other, less easily-articulable symmetries, which could be utilized in constructing suitable inversion scenarios); and, in any case, it is entirely conceivable that the mechanisms of color perception might not have created such difficulties.

seem to have inverted their color words: The “red-looking”²⁴ cucumbers are called “green” by the Rolocians, the “green-looking” tomatoes “red.”

The scientists do a bit of investigation, however, and discover that the Rolocian tomatoes are, physically, just like ours; they are made up of just the same chemicals and so forth. And when a Rolocian “green” tomato is brought back to display on Earth, the scientists are wildly disappointed: the exotic tomato gets mixed in with a few of our Earth tomatoes, and it is immediately indistinguishable from them. On Earth, the Rolocian tomato looks, to all the disappointed Earthlings, every bit as red as ours.

I’ve just said that the Rolocian tomato *looks* every bit as red as ours, when it’s brought to Earth. But I also think we can say that it *is* every bit as red as ours. There are a few ways to defend this claim. Perhaps the most direct is to note that Earthly and Rolocian tomatoes are physically identical, and, on Campbell’s own view, we can assume a type of supervenience relation between physical properties and color properties: no difference in color without a difference in physical properties.

Now we can ask: granting that the Rolocian tomato is red here on Earth, what color was it back on its native Roloc? Viewed in Rolocian sunlight, the tomato always would have been described by any English-speaking Earthling who saw it as “green” (supposing the Earthling was as yet unaware of the tomato’s molecular match with Earth tomatoes). So there might seem to be some initial temptation to say that the tomato *was* green on Roloc; it only *became* red when it was taken to Earth and its Earthly lighting sources. But this is a mistake. We can stipulate that no changes were made in the surface of the tomato during its space voyage. And, since color is, on Campbell’s view, an intrinsic, categorical property of (the surfaces of) objects, we therefore know that no change in color has occurred.²⁵ The Rolocian tomatoes, like our Earth tomatoes, are and *always were* red.

So, there have always been red tomatoes both on Earth and on Roloc. I now want to ask a crucial question: What color do red tomatoes *look* to be on Roloc? In order to answer this question, I first want to note something about the Rolocians’ language, and, specifically, their color vocabulary. Remember that, upon arrival on Roloc, our scientists initially thought that the Rolocians’ color vocabulary was inverted relative to our own. The Rolocians called their tomatoes—which *looked green*, to the visiting Earth scientists—“red.” But, having now determined that their tomatoes, as much as ours, are indeed red, we must conclude that actually, their language and ours match precisely

²⁴ The scare quotes here are meant to indicate that this “looks talk” is just how things seem to the Earthlings; the question of how the cucumbers and tomatoes look to the Rolocians is addressed below.

²⁵ We could also simply appeal to the idea that we do not take differences in illumination to actually change the colors of objects. A white sheet of paper under a red lamp may look red (or pink), but it is straightforwardly still white. I take this fact to be a basic datum in our investigation of our color concepts (and it is certainly something that anyone who, like Campbell, views colors as *categorical* properties of objects must accept).

(even when it comes to which color a particular terms designates): “red” in Rolocese is used to pick out red objects like tomatoes (whether of Earthly or Rolocian origin). So we can, it seems, abandon the idea that we have two separate languages here at all: our color terms, just like the rest of our languages, match up perfectly.

But now suppose I take a trip to Roloc and see some tomatoes. They look to me to be pure green, but, being familiar with the way that Rolocian lighting is different from our own, I know that they are actually red. The Rolocians, however, insist that, in addition to actually *being* red (a fact that we all agree on), the tomatoes on Roloc also *look* red. These red tomatoes look to them the way red tomatoes have always looked to them, when viewed on a sunny afternoon or in the direct light of the kitchen; so of course the Rolocians will say they “look red.”

We’ve just seen that “red” means, for the Rolocians, the same thing it does for us – it refers to the color possessed by all the tomatoes on both our planets. So it seems I am having a genuine disagreement with the Rolocians, not a misunderstanding due to a difference in the meaning of our respective languages’ color terms. Suppose I then fly a Rolocian back to Earth with me. She sees some tomatoes and says they look pure green to her, though, of course, they look red to me. Both we and the Rolocians acknowledge that the tomatoes on both planets *are* red (in *both* our respective languages). And, we both agree, the way tomatoes *look* on Earth is very different from the way they look on Roloc. But we Earthlings say they *look* red on Earth, green on Roloc, while the Rolocians say the reverse.

Remember that on Campbell’s view, red things must *look the way they are*, as a general rule (there may be occasional illusions, but there cannot be totally general illusions). But this must hold as much for the Rolocians as for us (for, given the way I’ve told the story, we can imagine that we ourselves are “actually” the Rolocians—perhaps I carelessly reversed the names of the planets in my telling of the story). So, which way of looking red reveals the true nature of the way red things are? On which planet are the colors transparent to perceptual experience?

We and the Rolocians will not agree on this: We think it is the way red things look on Earth that reveals their nature, the Rolocians that it is the way red things look on Roloc. Both cannot be right. If the way tomatoes look on Roloc revealed the true nature of redness, then all of us Earthlings would have been mistaken about redness all along: we would never really have known *which* property redness *was* at all, since its nature turns out to be revealed only when it is seen in the way we call “looking green.” And similarly for the Rolocians, if it turns out that redness’ true nature is in fact revealed only when red objects look the way they do on Earth. Worst of all, it would seem completely arbitrary for us Earthlings to insist that we were right and the Rolocians mistaken; we simply now have no idea whether we know what red *really* looks like – and therefore, no idea what content our concept of redness has.

In the previous paragraph, I used the phrase “the way red things look on Earth [or: on Roloc].” Implicit in this was an assumption (which I had earlier said I wanted to be careful about) that there is a *single* way that red objects look *on a given planet*, to

observers *from either planet*. But Campbell could object to this assumption; without it, we would not be forced to conclude that either *the* way red things look on Earth or *the* way they look on Roloc was the way that revealed red's nature. If there is no *single* way a red object looks on a given planet, then the worry of the above paragraph would disappear. For suppose there is no single way that red things look on a given planet. Suppose that, on Earth, red objects look red to us, but green to the Rolocians; on Roloc, red objects look green to us, red to the Rolocians. Both of us are correct in our claims about which planet is the one on which red objects look red—we simply need to add that we each mean “look red *to our kind*.” And, with that qualification, we can see that we all agree about what it is for an object to look red. In such a scenario, there would no longer be any worry that the entire population of one of the planets was badly mistaken about the very nature of redness.

There are a few interrelated problems with this response. First, we might ask *in virtue of what* the tomatoes look different to us than they do to the Rolocians, when we both view them on a given planet. Being a Twin Earth planet, Roloc is inhabited by beings that are genetically identical to us (they could even, after our communities begin to mingle, interbreed with us!). So it would be mysterious *why* a red tomato would look different to a Rolocian, when we are both standing in front of it here on Earth (or there on Roloc), with the same light illuminating it for both of us, and the same physiology allowing us to see it. Second, and more problematic, we would now have a bit of a mystery for Campbell's view. We are acknowledging that red objects could look totally different to two different observers in the same observational situation. Campbell's view is that the way red objects look, in a situation where one is veridically perceiving their color, is determined by the way they categorically *are* (namely, red), perhaps combined with facts about the observer's point of view (where this picks out her location, the direction of her gaze, the lighting conditions, and so forth). But here we have a case where the way the object is and the facts about the observer's point of view (and even facts about her perceptual mechanisms) do *not* determine the way the object looks. On the suggestion we are now considering, red tomatoes on Earth look one way to Earthlings, another to Rolocians. The problem with this kind of scenario, for Campbell, is to explain what makes it the case that the tomatoes look different to us than to the Rolocians. Given his other claims about the nature of color and of our perceptual relation to it, it seems utterly mysterious how this could be accomplished.²⁶

²⁶ A view on which facts about a subject's *history* partially determine the way things look to a subject might help here. On such a view, tomatoes might look red to the Rolocians on Roloc because of the historical fact that they had grown up seeing red tomatoes in Rolocian light; those same red tomatoes, in that same lighting, would look green to us Earthlings because we had grown up seeing red objects in different lighting. This would allow us to reject the idea that either we or the Rolocians were mistaken in thinking tomatoes looked red on our native planets, green on the alien planet, because it would imply that tomatoes simply look different to each race—due to the races' differing *histories*—when those tomatoes are viewed on a given planet. I am not convinced such an account ultimately works (for one thing, after the planets

So I think we really are forced to conclude that red things look the same to us and to the Rolocians, when we stand side by side on a given planet. And with that conclusion comes the question I have been pressing: do red things look red on Earth, or on Roloc? Roloc is a fantasy, to be sure. But what I think it brings out is a crucial feature of our grasp of the nature of color properties. Campbell's view requires that red things, in general, look red here on Earth; that is how the very nature of redness can be revealed to us in our perceptions of red things. But, if our choice—about which way of looking was the one that revealed red's true nature—was arbitrary when we were directly arguing with the Rolocians in our imaginations, it seems no less arbitrary when we are simply considering our own situation and the nature of our grasp of what redness is. Given that red objects *could* (if we were unfortunate enough to be on Roloc) look the way green objects in fact look to us (and vice versa), in what sense can we know that we have a grasp of the very nature of redness in seeing it the way we happen to see it in our circumstances? Why couldn't we be mistakenly attributing *greenness's* nature to red objects, in seeing them the way we do? Presumably that is what we would have to say the poor Rolocians are doing, if their encounters with red objects are all instances of those objects' looking green. And again, we have no way of saying why it is their perceptions, rather than ours, where the mismatch between experience and color occurs.

In responding to these worries, we would seem to need some way to grasp just what redness's nature *is*. Then, we could know that it was that very nature we were encountering in seeing red objects, that very nature that was revealed by a particular way of looking (and not by other ways of looking). But such a grasp is precisely what is lacking here. We have no more claim to it than the Rolocians. In trying to give it content, we can each do no more than gesture at our respective experiences of red objects, which, we have seen, would lead us to incompatible understandings of what that content is.

I began this section by noting that Campbell's view requires a tight conceptual connection between an object's being green and the way it looks – the character of an experience of green. On Campbell's view, an object's looking green is a matter of its *being* green, and a subject's perceiving that greenness for what it is. There should, on this view, be no concept of looking green that can be pulled apart from the concept of greenness itself. And yet, what Roloc seems to show is that we do have a concept of looking green that comes apart from greenness itself (and from experiences *caused by* greenness). We can see this dissociation displayed quite clearly in the *language* of the Rolocians: Their term "green," used to denote a color property of objects, picks out the same property as our term "green." But their term "looks green" does *not* pick out the type of experience that our term "looks green" picks out. They deploy the term

had discovered one another, we would have to consider what to say about inter-racial children who grow up spending time on both planets), but in any case, this is not a direction Campbell seems to want to go (though Stroud might be more sympathetic to such a move).

“looks green” when they perceive cucumbers on Roloc—cucumbers that look *red* (if we are speaking Earth English). Rolocese provides a linguistic proof of the fact that greenness and looking green are dissociable, in just the way that Campbell’s view implies they should not be. And so, we have strong reason to reject the account of the conceptual connection between color and color experience that Campbell gives.

4.9. THE DISTANCE BETWEEN SHAPE CONCEPTS AND EXPERIENCE

The case of Roloc suggests that we cannot spell out a conceptual connection between color and experience of color in the way Campbell posits. There does not appear to be a conceptual connection—running from a particular color property to experiences of that property—that makes the very concept of an experience with a given character depend on the color that in fact causes that experience. In §4.7, I suggested that there was another problem for the view that the connection between color properties and experience of them ran in this direction: namely, that the account would apply equally to shape, even though there seems to be no conceptual connection between a particular shape and the experiences caused by objects with that shape. In this section, I want to press that objection further, in relation to Stroud’s account of the biconditionals (mentioned in §4.7) that link properties with their dispositions to cause experiences of particular types.

Stroud claims that a connection exists between experiences of a property and the property itself, just as much for the “primary” qualities as for the “secondary.” My claim throughout has been that, though there may in fact be some connection between shape properties and the experiences caused by particular shapes, that connection is not a *conceptual* one: our *concept* of a given shape property is not a *concept* of something that is disposed to produce certain sorts of experiences. This is in contrast with color properties where, I have suggested, the connection between a particular color and the tendency to produce experiences of a particular type *is* at the conceptual level. This difference may not emerge in our day-to-day encounters with shapes and colors, but it does come out when we assess the intelligibility of various counterfactual scenarios: the non-conceptual nature of the connection in the case of shape allows us to easily grasp the possibility of a divergence between shape and experience of shape, while the tighter connection in the case of color puts pressure on our ability to make sense of analogous scenarios involving color experience (again: at least if we reject dispositionalist and microphysicalist accounts of color). In §4.6, I brought this out by describing a hypothetical scenario in which all the facts about people’s experience of shape left unanswered the (still perfectly intelligible) further question about an object’s shape itself. But I think we can see the same point emerge without needing to consider far-fetched imaginary worlds. It seems to me there is a sense in which the divergence between shape and our perceptual experience of shape is, in certain kinds of cases, quite *actual*.

We can bring out this real-world divergence simply by changing the particular examples Stroud uses when he suggests that there is a connection between shape and shape experience that parallels the connection between color and color experience. Stroud says that an object “that is rectangular is such as to be seen to be rectangular in conditions optimal for telling the shapes of things by looking at them,” and, likewise (in the case of another primary quality, number), that a “collection of seventeen objects is such as to be counted by us as seventeen in number in conditions optimal for us for telling the number of things by counting.” These remarks are true enough, and they do suggest a *prima facie* analogy between primary and secondary properties, in terms of their dispositions to cause certain experiences in us. But Stroud’s claims, applied to different shape and number properties, would simply be untrue. Consider a regular googolplex-sided polygon. We can have a perfectly concrete, contentful concept of such an object: it is simply the concept of an object with a googolplex sides of equal lengths, all connected at angles of an equal number of degrees.²⁷ But if we try Stroud’s formulation here in spelling out a connection between such an object and our experience of it, we get something that is manifestly false: It is in fact *not* the case that a regular googolplex-sided polygon is such as to look (or be seen to be) googolplex-sided in conditions optimal for telling the shapes of things by looking at them. Such an object would, in ideal circumstances, look to be *circular*. No matter how good the viewing conditions, a human observer could never discriminate the individual sides of such an object; the object would be entirely indistinguishable from a perfect circle in terms of the experiences it would produce (indeed, it would be closer to *being* a perfect circle than any object any of us has ever actually seen). Similarly, a collection of a googolplex objects would not be such as to be counted by us as a googolplex in number in conditions optimal for counting; in ideal conditions for counting, we’d all be dead before we finished the counting.²⁸

²⁷ There is an obvious parallel here to Descartes’s observations about conceiving of (as opposed to *imagining*) chiliagons and myriagons in the Sixth Meditation. While I think some of what Descartes says there is congenial to the point I am making, the examples here are deployed for somewhat different purposes.

²⁸ We could perhaps finesse the notion of “conditions optimal for counting” here so that we could get around facts about how we would surely perish well before counting the whole collection (and the further worry that a collection of a googolplex objects of discernable size would likely not fit within the physical universe (see <http://en.wikipedia.org/wiki/googolplex>)). But such machinations seem harder to defend in the case of the googolplex-sided object. What idealization of “optimality,” precisely, would we be abstracting to, in imagining a scenario in which the googolplex-sided object would indeed be seen to be googolplex-sided? Conditions in which we had shrunk to a size that allowed us to see the individual sides? This begins to look like Berkeley’s challenge to specify the “true” colors of objects, given that they look to be different colors at different levels of magnification. The search for “truly optimal conditions” in these kinds of cases—where optimal conditions are taken to be something other than typical conditions in which we actually view objects—seems hopeless.

As noted above, Campbell and Stroud propose a conceptual connection between a given type of color experience and the color that causes that type of experience. Their claim is one about the nature of perceptual experience as such, and is not in any way specific to color experience. For that reason, the conceptual connection they propose—which claims that experiences are identified, at the conceptual level, only in relation to the properties they are experiences of—should apply equally to shape experience. But what both the example of the possibly-circular Antarctic object and our consideration of a googolplex-sided object show is that we *can* pull apart shape and experience of shape, at the conceptual level. We have a concrete conception of the possible world in which half the population experiences the Antarctic object as circular, even though the object's true shape is not circular. Similarly, we have a concrete conception of a googolplex-sided object, and we know that—even in the actual world—the experiences that object causes would *not* be experiences of its true shape. The concepts we have of shape and shape experience are not essentially connected in the way that Stroud and Campbell suggest they should be: the biconditional connecting circularity with experience of circularity does not describe a connection at the *conceptual* level; and when the particular shape in question is changed, the connection evaporates entirely, and the biconditional is simply false.

Let us step back for a moment to consider the dialectical situation. When I initially introduced the two biconditionals in §4.7, I noted that Stroud claims that *neither* biconditional should be regarded as offering an analysis of the property in question. In this section, my claim has been that a certain type of conceptual connection does not hold between shape properties and experience of them. So it might seem that my claims are actually quite congenial to Stroud's view: he and I both deny that a biconditional connecting a shape property and experience of that property expresses a conceptual analysis of the shape property.

Stroud does not want to interpret biconditional (2) as an analysis of rectangularity. Still, I argued above that Stroud's account does, in a subtler way, entail that there is a conceptual connection between a shape property like circularity and experience of that property. In §4.1-6, I had developed a line of argument to suggest that there is a certain tension in the kind of view held by Stroud and Campbell. On the one hand, the denial that a color property can be analyzed in terms of the experiences it typically produces suggests that we should always be able to conceive of the experiences and the color itself coming apart. On the other hand, the refusal to identify color properties with either microphysical properties or dispositions to cause experiences of a certain type makes it hard to contentfully specify what property a particular color property *is*, in those very scenarios where the color and experience of it come apart. I suggested, in §4.7, that this tension could be resolved by denying that the scenarios in question were genuinely possible. One way to do this would be to give a dispositionalist account on which color properties are indeed analyzed in terms of biconditionals like (1). But, rejecting that option, Stroud and Campbell both suggest another way in which we might try to close the gap in conceptual space between colors and color experience,

and thereby rule out the problematic scenarios as not genuinely conceivable. They claim that experiences of a particular type can only be specified as such, at the conceptual level, by reference to the properties those experiences are experiences of. If there is no possibility of independently identifying experiences as “green experiences,” then we need not worry about scenarios in which green experiences come apart from greenness. There simply would not be conceptual space for such scenarios, and thus there would be no need to meet the demand for a specification of what is going on in them.

This move had a further consequence, however, unrelated to color properties. Due to the nature of the connection proposed by Stroud, that connection would apply equally to shape properties, and so, to the extent the account ruled out the problematic color scenarios, it would also rule out scenarios in which shape and shape experience diverge. That is why it seems problematic for Stroud’s view that biconditional (2) does not, in fact, capture any real conceptual connection between shape properties and experience of them. Even if biconditionals (1) and (2) are not to be taken as analyses of the properties mentioned, if there is a “reverse” conceptual connection between experiences of a certain type and the property disposed to produce them—if these biconditionals are meant to be analyses of the concepts of the relevant *experiences*, instead of analyses of the properties experienced—then the biconditionals ought to be expressing some sort of conceptual truth. And that is what seems not to be the case with (2): if we change the particular shape mentioned in it, we can see that the biconditional not only fails to express a conceptual truth, but also simply fails to *be* true.

At this stage, we might wonder whether the lesson to draw is that we should abandon the idea of *any* type of conceptual connection between a given property and the experiences it causes. We seem to be pushed in this direction for shape, as the considerations in this section show. So why not simply regard biconditionals (1) and (2) as equally non-conceptual claims, expressions of simple matters of fact about the relations between particular properties and experiences?

Here we come again to the crucial difference between primary and secondary qualities, in terms of the relation between those qualities and experiences of them. For consider whether sense can be made of a version of biconditional (1) analogous to the “googolplex version” of (2). That is, consider whether we can alter the color property mentioned in (1) in such a way that we can bring out the contingent, non-conceptual nature of the connection between color properties and experiences of them, the way we were able to do with shape properties by switching to a googolplex-sided shape in (2). Suppose we suggest that there could be a particular shade of yellow that is not such as to look yellow in normal circumstances, but is instead such as to look, say, red.

If color really were just as conceptually distinct from our experiences as shape proved to be, it should be perfectly possible to find such a case. But it seems to me that it is not possible, and that the reason is not merely a contingent, idiosyncratic feature of color and shape properties. The problem is that in imagining a shade of yellow that is such as to look red, there is no real sense in which we are actually imagining a shade of *yellow*; we are, instead, imagining a shade of red.

In the case of the googolplex-sided object, I can spell out what makes it the case that, in spite of the fact that I am considering an object that is such as to look circular, I am *not* discussing an object that *is* circular. The fact that makes the object in question non-circular is the fact that it is an object with a certain number of straight sides, joined at equal angles. Again, it is the fact that I can spell out a specific *way in which the object occupies space*, and that fact is not at all undermined by the fact that the object would not be experienced as so occupying space. In contrast, there seems to be nothing grounding the claim that a particular shade experienced as red is in fact a shade of yellow. Instead, with nothing beyond the experience of the shade (as red) to ground a claim about its color, it seems as though we are simply mischaracterizing the scenario as involving a shade of yellow, when it really involves a shade of red.²⁹

The crucial difference between shape and color here is that our concept of shape has concrete content, independent of experience. That is what allows us to conceive of experience and shape varying independently, and to identify particular shapes that do not cause experiences of those very shapes. In order for there to be real conceptual distance between a property and our experience of it, we need a concrete, non-experiential conception of the property *itself*; we need to know what it is that our experience has conceptual distance *from*. We can spell out that conceptual content quite well in the case of shape; we cannot do the same in the case of color. That, in the end, is what marks off the crucial divide between the primary and secondary qualities.

4.10. CONCLUSION

In closing, I want to bring to the fore something that I have to this point relegated to the background, in various footnoted and parenthetical caveats. The structure of my argument in this chapter has had a somewhat unusual character. I have, at times, been offering arguments about the impossibility or inconceivability of certain scenarios in which color and color experience come apart. At other times, I have been demanding a way of contentfully specifying those scenarios, and arguing that no such specification is possible. But my arguments have not been intended to establish the absolute (or

²⁹ Once more, a caveat: I do not think the idea of a shade of yellow that is such as to look red in normal circumstances is fully incoherent, as the last few paragraphs may suggest. You might imagine that we *do* have something analogous to “a particular way of occupying space” for the colors, which allows us to specify *in virtue of what* the shade *is* a shade of yellow: namely, its reflecting light of a particular bandwidth. We have a concrete concept of such a property, and we might say that an object that reflects light in a particular portion of the “yellow bandwidth,” but which is nonetheless disposed to look red (perhaps because viewing it disrupts the human brain’s functioning in a particular way), is in fact yellow. It seems to me that something along these lines is in fact correct. But here, the point I am trying to draw out is that, with the conceptual tools Campbell and Stroud bring to bear on an analysis of color—which specifically exclude microphysical facts, or others unavailable to naïve subjects—we cannot invoke *this* way of spelling out how there could be an object that is truly yellow, despite looking red.

genuine) inconceivability or conceptual emptiness of these scenarios. It is only within the context of a view like Stroud's or Campbell's, one on which we have ruled out analyzing color concepts in terms of microphysical properties or dispositions to produce experiences, that the scenarios become impossible to conceive of or to specify. That is to say, my argument is a conditional one: *if* we have an account that rejects the dispositionalist and microphysicalist analyses of color, then we have a real problem in thinking about these scenarios.

The problem can be framed as a sort of internal tension in the accounts of Stroud and Campbell. Those accounts claim that color and color experience are conceptually separable, but they also make it mysterious how to conceive of the scenarios in which the separation occurs. I considered a way of responding to this worry from within the framework of Stroud's and Campbell's account, one that calls into question the conceptual separability that forms one half of the internal tension. I argued that that way of denying the separability could not be made to work: first, as shown in the example of Roloc, because the kind of conceptual connection—running from particular colors to the experiences they cause—that the account posits can be shown not to hold; second, because the connection does not hold even *factually* for shape, and the account suggests that it should hold at the level of our very concept of shape experience. I then went on to claim that there is, nonetheless, reason to deny the conceptual separability of *color* and experience: I suggested that we cannot fully separate our concept of a particular property from our experience of it in the case of color.

I now want to acknowledge that this last claim must be qualified, by taking into account the structure of this chapter's argument, as described two paragraphs above. My denial that there is any sense to the idea of color and experience of color coming apart is only a conditional one—conditional on the rejection of the dispositionalist and physical-reductivist tools that Campbell and Stroud do, in fact, reject. I think that there is, in the end, a way to make sense of such scenarios, that the question we started with back in §4.2 does indeed have an answer: I think there *is* a way to give color concepts the kind of content needed for it to be intelligible what question we are asking when we ask whether my family's garbage bags were *really* green. At the same time, there is a genuine limit on how widely color and color experience can diverge, at the conceptual level. Our concepts of color properties are not, as in the case of shape properties, fully specifiable independently of experience. And yet, color properties themselves are not tied to our experiences with any kind of metaphysical necessity.

We cannot divorce our concepts of the colors from our experience of them. There is, unlike with shape, a conceptual connection; this is what it is for colors to be secondary qualities. But the conceptual connection does not, in fact, rule out as impossible scenarios in which colors and color experience diverge. This is the tightrope that an account of color must walk: it must make intelligible scenarios in which experience of color and color itself come apart, and yet it must acknowledge the role that experience plays in our very concept of color. Campbell and Stroud both attempt to capture

something like this idea, but, I have argued, their accounts fail to accurately portray the way our color concepts work.

Instead, the intelligibility of scenarios in which colors and color experience diverge is, in an important respect, indirect. It is because we can conceive of objects in terms of their *primary* qualities—because we have *a priori* concepts of their spatial features—that we can conceive of them as independent of our experience. Conceiving of an object as something that *occupies space* gives us a way to think of it as independent of our experience; it is this conception that grounds our capacity to conceive of the shapes of objects coming apart from our experience of them. Once we have this conception, the idea that objects' colors are features of the objects can be cashed out in terms of those objects having the property, whatever it might be, that objects actually do have, in virtue of which they tend to cause certain experiences in us.³⁰ Such a property—the categorical ground of objects' actual dispositions to cause experiences of color in us—could, counterfactually, come apart from color experience, on this conception. But, in conceiving of such possibilities, we are still conceiving of the colors—properties that are not, in their natures, dependent on us—by way of our grasp of the experiences those categorical properties cause in us. We are acknowledging that the property of objects that actually does tend to cause certain experiences in us—a property with which we do not have acquaintance—might, in some other possible world, have failed to do so.

That is only the most minimal sketch of an account of color experience, a topic that deserves a much fuller treatment.³¹ But I will have to leave the discussion here, and return, in the following chapter, to my central topic: the nature of our experience of *primary* qualities, and, in particular, our experience of shape.

³⁰ This idea echoes the Ramseyan account of perceptual content I discussed in Chapter 1. Here, I am suggesting (as I did there) that such an account, though unsatisfying if taken to apply to *all* aspects of perceptual content, is indeed plausible when it comes to our experience of color.

³¹ See Evans (1985) for a sustained discussion of a related view of colors and color perception.

CHAPTER 5

SHAPE PERCEPTION IN A RELATIVISTIC UNIVERSE

The feeling that this concept, “the shape of the moving body,” has an immediately obvious meaning is due to the fact that in our day-to-day experience we are accustomed to encountering only such velocities of motion that are practically infinitely small compared with the velocity of light.

Albert Einstein¹

How should we think of the spatial content of perceptual experience? In particular, how should we think of the content of our experience of shape, when, for example, we see an object as square? On what I will call the ‘placeholder’ view, an experience of squareness just represents *whatever* property it is that usually causes such experiences; this reference-fixing content will pick out different determinate properties in different circumstances.² On the ‘presentationalist’ view, by contrast, experiences of squareness are more committal: their content always represents the same determinate shape property—*squareness*—rather than merely providing an open-ended description. One natural way to develop this latter view is as follows: the particular property all experiences of squareness represent is the geometrical property of squareness about which we reason in doing *a priori* Euclidean proof.³

Such a presentationalist view has deep intuitive appeal. But Einsteinian relativity theory—by calling into question the idea that our physical universe instantiates Euclidean spatial properties—seems to confront the presentationalist with a dilemma: either she must give up the thesis that our shape experiences represent Euclidean spatial properties; or she must accept that all of our shape experiences misrepresent the objects we perceive.

In what follows, I will argue for the presentationalist view, defending it against the charge that contemporary science makes it untenable. Einsteinian relativity theory does not force us to accept that, as the placeholder view holds, our shape perceptions provide only an opaque reference-fixing device, leaving the determinate property thereby represented outside our ken; nor does it imply that shape perceptions *misrepresent* objects as instantiating Euclidean shapes. Rather, what we learn from contemporary physics is that our shape experiences directly present determi-

¹ Einstein (1911, p. 348).

² Such a view has been defended by David Chalmers under the label ‘shape functionalism’. See Chalmers (2012, Chapter 7; and forthcoming).

³ I have been defending a presentationalist view of spatial experience in the preceding chapters; the “placeholder view” is a version of the Ramseyan view I outlined in Chapter 1.

nate Euclidean properties, while *implicitly* representing those properties as instantiated in a particular *manner*: namely, relative to our own frame of reference. As I will show, such a presentationalist view does not have the implication that our spatial perceptions are illusory: even a relativistic universe like ours can be veridically represented in this way. Properly understood, Einstein’s theory poses no threat to the commonsense idea that we have genuine perceptual acquaintance with the spatial character of our world.

5.1. INTRODUCTION

During the Early Modern period, many philosophers, impressed by the discoveries of the scientific revolution, argued that colors were unreal. According to these color eliminativists, all of our perceptions of color are illusory, representing the objects we perceive as having properties that no physical object actually has. The eliminativists were swayed by the following principle, emerging from the then-current scientific zeitgeist:

NON-INSTANTIATION: No properties other than geometrical properties (and, in particular, no secondary qualities like colors) are instantiated by any physical objects in our universe.⁴

Recently, though, many philosophers of perception have been swayed by another, more intuitive, idea:

VERIDICALITY: Our color perceptions are not all illusory; that is, the properties attributed to objects by our color perceptions are (at least sometimes) properties the objects causing those experiences actually have.

Some of the advocates of VERIDICALITY have, like the eliminativist faction, been impressed by the evidence in favor of NON-INSTANTIATION.⁵ And so they have been led to adopt an account of the *contents* of our color perceptions (an account of the way in which those perceptions attribute properties to objects) that can reconcile the two propositions. According to (one version of) this view, a perception of an object

⁴ Galileo and Descartes would likely have endorsed such a version of the principle, while Locke—though he sometimes suggests something along these lines—was less committal about which specific properties physical objects instantiate (he sometimes suggests that they may be properties that are simply unknown to us).

⁵ For contemporary theorists, NON-INSTANTIATION will take a modified form, since the list of properties regarded as ‘scientifically-respectable’ has changed (it is no longer limited to geometrical properties). But contemporary scientific theories do—at least on many interpretations—still entail that objects never instantiate irreducible *color* properties. See Mackie (1976, p. 18).

as, say, red, is to be counted as veridical if and only if the object in question instantiates the property—*whatever* property it turns out to be—that typically causes such experiences. NON-INSTANTIATION is no threat to VERIDICALITY on such a view, since the content of a color experience includes a kind of ‘placeholder,’ rather than a determinate specification of a property that science has (allegedly) revealed not to feature in physical reality. That placeholder is then (through elaborate empirical investigation) to be filled in by a scientifically-respectable property, with which we identify the property of redness: for Locke (who, in some of his moods, was an early advocate of this kind of view), this property turns out to be a particular ‘texture’; for modern versions of the account, it turns out to be a particular spectral-reflectance profile.⁶

Early Modern theorists did not, generally speaking, apply this story about the contents of our perceptions of colors—the story involving a placeholder in the specification of the perceptual content—to *primary* qualities like shape. This was because a natural, pre-theoretical account of the contents of shape perceptions—an account that did not include any placeholders—could accommodate both NON-INSTANTIATION and a shape-centered version of VERIDICALITY. Since squareness, unlike redness, is itself a geometrical property—one of the properties the science of the Early Modern period was willing to grant to physical objects—a perception of squareness could attribute that very determinate property to an object *veridically* (that is, the experience’s content could include the determinate, geometrical property *squareness*, without any need to insert a placeholder, and still represent the object accurately).

The science of the Early Modern period, then, challenged the pre-theoretical conception of the colors of objects, but not the conception of their shapes; and, as a result, philosophers proposed revisionary accounts of color—but not shape—*experience*. Later revolutions in science, however, might seem to have moved us beyond this traditional divide. In particular, on some interpretations, the Einsteinian revolution of the twentieth century called into question the scientific respectability of the primary qualities, in much the same way that the science of the Early Modern period called into question the scientific respectability of the secondary qualities. Consider Minkowski’s famous claim that, given Einstein’s special theory of relativity (henceforth, ‘STR’), ‘space by itself, and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality’. On this picture, a *spatial* property like squareness is no more scientifically legitimate than a color property was for the Early Moderns; both are ‘doomed to fade away into mere shadows’.

This line of thought has generated much debate about how our conception of the spatial features of the universe might need to be revised in light of scientific

⁶ For an interpretation of Locke as holding this kind of view, see Campbell (2002); for defenses of the modern version, see Chalmers (2006) and Thompson (2010).

theory.⁷ Until recently, though, there has been little discussion of whether the Einsteinian revolution should also lead us to revise our conception of spatial *experience*.⁸ As described above, the Early Modern challenge to our conception of the colors, combined with a commitment to VERIDICALITY, led many to adopt a placeholder account of color experience. The Einsteinian challenge to our conception of spatial features, then, would seem to augur a parallel revision in accounts of *shape* experience. And, indeed, some philosophers of perception have recently suggested that, in order to preserve VERIDICALITY for our shape experiences, we must include a placeholder, rather than a determinate spatial property, in specifying their contents.⁹ On this account of shape perception, the property attributed to an object by an experience of that object as square is just *whatever* property in fact typically causes such experiences; and, given what we have learned from Einstein, the property in question will not turn out to be a purely spatial property—a kind of property that nothing in our universe actually has—but rather some arcane property of the spatio-temporal manifold (Minkowski’s ‘union’ of space and time, whose independent existence is all that remains on the Einsteinian picture).

In the case of traditional primary qualities like shape, however, any such placeholder account of the contents of our perceptual experiences faces a problem. The problem arises when we try to combine VERIDICALITY and NON-INSTANTIATION—two theses the ‘placeholder’ account of *color* perception was designed to accommodate—with a third feature of our conception of shape, which has no analogue in the case of color. This third feature comes into view when we note that shape properties, unlike color properties, feature in our conceptual scheme in a variety of ways, some of which are not tied to our perceptual experiences at all. In particular, we reason about shape properties in the *a priori* domain of Euclidean geometry, and the property of squareness referenced, for example, in Hippocrates’s proof of the quadrature of the lune *is* a ‘purely spatial’ property (rather than a property of Minkowski’s spacetime manifold). Furthermore, it seems that it is *this very property*—the property about which Hippocrates proved some crucial propositions, which we might call ‘Euclidean

⁷ See, e.g., Reichenbach (1958), Horwich (1978), Friedman (1983). DiSalle (2006) provides a useful historical overview of the scientific and philosophical literature.

⁸ Strawson (1966) and Hopkins (1973) are important exceptions, though both of their discussions are focused on a narrower Kantian conception of experience, and on difficulties raised for such a conception by Einstein’s *general* theory of relativity (GTR). By contrast, my focus here will be on issues raised by *STR*, since, despite philosophers’ tendency to focus on GTR in discussing Einstein’s discoveries, it seems to me that a more fundamental question about the instantiation of spatial properties in our universe arises already in *STR* (on this point, I am in agreement with Tim Maudlin (see his (2012, pp. 127-128)).

⁹ Thompson (2010, p. 170) and Chalmers (2012, p. 334) explicitly acknowledge being motivated by considerations stemming from Einsteinian physics in proposing a placeholder account of spatial experience, though many who advocate this type of view are more centrally motivated by general externalist leanings. The arguments I present here are, in part, supposed to provide a way to resist some of the moves made by proponents of a broadly externalist account of perceptual content.

an squareness’—that shows up in the contents of our shape experiences. It is Euclidean squareness that we attribute to the objects we perceive when we perceive them to be square. The problem, then, is that we can combine these observations with *veridicality* and NON-INSTANTIATION¹⁰ to generate a set of three incompatible theses:

COMMONALITY: The property we reason about in doing *a priori* geometrical proof, Euclidean squareness, is the same property that we attribute to objects in shape perception.

VERIDICALITY: Our perceptions of squareness are not all illusory; that is, the property we attribute to objects in perceptions of squareness is a property the objects causing those perceptions (at least sometimes) actually have.

NON-INSTANTIATION: Euclidean squareness is not instantiated by any physical object in our universe.

We are now faced with a puzzle, since each of these three mutually inconsistent propositions is independently plausible.¹¹

Some might choose to resolve the puzzle by rejecting VERIDICALITY. This would be to see the Einsteinian revolution as an extension of the eliminativist project of the scientific revolution: the science of the Early Modern period revealed the unreality of the secondary qualities, and Einstein’s theory completed the project by revealing the unreality of the primary qualities.¹²

The kind of universal error such a view ascribes to our perceptual experiences is not terribly appealing, however, and it does not cohere very well with our actual practice (those familiar with Einstein’s theories do not generally take themselves to be permanently suffering from spatial illusions). So another option, one that many advocates of a placeholder account of shape perception endorse (at least implicitly), is to reject COMMONALITY. The lesson of the Einsteinian revolution, on this picture, is that we do not—from the perspective of perceptual experience itself—have a direct grasp on the particular contents of our spatial experiences, any more than we

¹⁰ Each suitably modified to apply to our *shape* perceptions and the scientific picture that emerged from the Einsteinian revolution. In what follows, I focus on the particular case of squareness for ease of exposition, but analogous points can be made for any shape property.

¹¹ Note that an analogous puzzle cannot be generated in the case of color because there is nothing analogous to Euclidean squareness—no pre-theoretical, non-experiential conception of redness to which we are committed—that can be used in formulating a version of COMMONALITY for redness. This leaves us free to identify the property of redness that we attribute to objects in perception—the property mentioned in the formulation of VERIDICALITY—with a property that is *not* ruled out by NON-INSTANTIATION (this property might turn out to be a texture, or a spectral-reflectance profile).

¹² This would seem to be Minkowski’s view, judging by the above quotation.

do in the case of color. In both cases, we have only a kind of placeholder, a promissory note that *whatever* determinate property our perceptions attribute to the world around us, that very property is the property that has typically caused those perceptions. We might have hoped that we had a better grip on the contents of *shape* perceptions—in particular, we might have thought that the properties attributed by such experiences were the very properties we reason about in doing *a priori* geometrical proof—but Einstein revealed that that could not be the case if we want to avoid ascribing universal error to our shape experiences. In light of Einstein’s discoveries, this view insists, we need to *bifurcate* our spatial concepts: the property of squareness we attribute to objects in perception is *not* the property of Euclidean squareness about which we reason in doing geometrical proof (instead, it is a property of Minkowski’s space-time manifold).

In recent work, David Chalmers has defended such a placeholder account of shape experience. He argues, in part on the basis of considerations from STR, that the property attributed by our perceptual experiences of squareness is not one on which we have a determinate, *a priori* grip. Instead, squareness is to be picked out in terms of the role it plays in generating our experiences: *whatever* property in fact typically causes our experiences of squareness will be the property represented by those experiences. And, given what we know from STR, that property will not be Euclidean squareness.¹³

Though it allows us to escape the puzzle, the move to a placeholder view in the case of shape experience has heavy costs. The idea that we have an *a priori* grasp of space and spatial structure, and that it is our *a priori* spatial concepts that we apply in experience, is deeply intuitive. More importantly, it is only *because* the spatial concepts we apply in experience are the familiar concepts of Euclidean geometry that we have any real grip on the spatial features of our world. To perceive an object as square just is to perceive it as having a quadrilateral surface with four equal sides joined at four right angles—to perceive it as instantiating Euclidean squareness. If the spatial properties represented in our experience turned out *not* to be these familiar Euclidean properties, we’d have no real understanding of the contents of our own spatial perceptions. And, as Kant famously observed, our representation of the spatial

¹³ In Chalmers’s words, his view is one on which ‘we do not have any “direct” grip on the basic nature of spatial properties’ represented in experience (Chalmers 2010, p. 492). Chalmers does think, however, that we have a determinate concept of a *different* property, which he calls ‘Edenic squareness’, and which seems equivalent to the property I have labeled ‘Euclidean squareness’. But contemporary physics, according to Chalmers, has shown that Edenic squareness is not instantiated in our universe (Chalmers 2006, p. 443; 2012, p. 333; and forthcoming, p. 22); and yet we still take our squarish experiences to be veridical. This suggests that the property represented by our squarish experiences is not Edenic squareness (Chalmers 2012, p. 326 and forthcoming, p. 22). Thus, Chalmers favors precisely the kind of ‘bifurcation’ of our spatial concepts described in the preceding paragraph: in light of contemporary scientific discoveries, Chalmers thinks we should reject COMMONALITY in order to hold on to VERIDICALITY.

features we perceive seems to play a unique role in our capacity to form a conception of the mind-independent objects of the empirical world; thus, severing the connection between our *a priori* spatial concepts and the features represented by experience threatens to leave us without any determinate conception of the world we perceive.¹⁴

I want to suggest that we can avoid these costs, by resolving the puzzle in a different way: we can reject NON-INSTANTIATION, and thereby dissolve the pressure that Einstein's theory supposedly puts on COMMONALITY. In what follows, I will address Chalmers's argument in support of a placeholder account of the contents of our shape perceptions, suggesting that his picture rests on a faulty analogy between the relativity of spatial properties revealed by STR and the variability of experiential content revealed by classic philosophical thought experiments involving colors. I will then show how the challenge that NON-INSTANTIATION poses to VERIDICALITY and COMMONALITY hinges on a key ambiguity in spatial terms like 'Euclidean squareness'. Einstein's theory does show that a certain kind of *absolute* spatial property—what we might call '*absolute* Euclidean squareness'—is not instantiated in our universe. Spatial properties that are absolute in this sense are properties objects instantiate independent of reference frame—independent, that is, of any temporal facts. But there is another way to think about spatial properties that does *not* require them to be absolute: we can delineate a set of genuinely spatial properties by assigning shapes to objects *relative to* a specified frame of reference. The objects we perceive do have spatial properties in this second sense, and those spatial properties are *Euclidean* spatial properties.¹⁵ So our perceptions of shape can attribute the very properties about which we reason in geometry to the objects around us, without misrepresenting those objects. Properly understood, Einstein's theory is compatible with both VERIDICALITY and COMMONALITY, and with the intuitive idea that we do indeed have a direct grasp of the contents of our spatial perceptions.

5.2. CHALMERS'S ARGUMENT AGAINST SHAPE PRESENTATIONALISM

Consider a visual experience of the type you have when you view a square surface head-on. Following Chalmers, we can call this type of experience, specified in terms

¹⁴ Chalmers himself actually embraces something like this consequence: he holds that the contents of our experience (and our associated 'everyday' beliefs about the objects we perceive) do not in themselves differentiate between a world in which we perceive material bodies and a world in which our experiences are generated by a computer hooked up to our envatted brains. (Chalmers does allow that we have 'metaphysical' beliefs 'about the underlying nature of reality' that differentiate between material and computer-generated worlds (Chalmers 2010, p. 459); but these beliefs do not show up at the level of the *perceptual* contents available to the subject on Chalmers's view.) Though Chalmers stresses the alleged epistemological benefits of this picture (Chalmers 2010 and forthcoming), the radical opacity of perceptual contents it embodies seems like a cost we would do well to avoid.

¹⁵ At least as far as STR is concerned. It's true that the spatial properties objects have according to GTR will not be perfectly Euclidean; I briefly address this worry below (fn. 48).

of its intrinsic phenomenal character (that is, in terms of what it's like to have the experience), a 'suarish' experience.¹⁶ We might wonder what the *content* of a suarish experience is; that is, we might ask what property such experiences attribute to the objects perceived, and what form the attribution takes.¹⁷ Chalmers describes two possible views, corresponding to the 'placeholder' and 'presentationalist' options I described above:

Shape functionalism (Chalmers's term for the 'placeholder' view) holds that the property represented by a particular kind of shape experience is 'whatever normally causes the relevant shape experiences'; on this view, 'shapes such as squareness are picked out in virtue of their role in causing our experiences of shape'.

Shape presentationalism holds that 'all suarish experiences represent the same property: squareness'. (Chalmers forthcoming, p. 21)

An important implication of shape functionalism is that *different* determinate properties can be represented by suarish experiences in different contexts, if different properties in fact typically cause those experiences in the respective contexts.¹⁸ This contrasts with shape presentationalism, on which there is a single determinate property attributed to objects by all suarish experiences.¹⁹

¹⁶ Chalmers uses terms of the form 'X-ish' to pick out experiences with a particular type of intrinsic phenomenal character, namely, the character of those experiences we typically have when we perceive Xs. I will follow his usage throughout the chapter. There may be a worry about whether it is even possible to individuate experiences in terms of their intrinsic phenomenal character (independent of the objects that cause those experiences); some versions of 'naïve realism' or 'disjunctivism' (see Campbell (1993), Martin (2004)) and some versions of 'strong representationalism' (see Dretske (1995), Tye (2000), Lycan (2001)) have the implication that it is not. I will not be directly addressing this kind of worry, or the views on which the talk of 'X-ish experiences' is illegitimate, here.

¹⁷ Some question whether experiences *have* contents—whether they attribute properties to the world—at all (see, for example, Travis (2004), Brewer (2006), Papineau (2013)). Siegel (2010) presents a sustained argument in favor of what has been called the 'Content View' of perceptual experience—the view that perceptual experiences do indeed have representational contents. I will not enter into that debate here, but will simply assume for the sake of my argument that Siegel is right, and that it makes sense to ask what content a perceptual experience has.

¹⁸ Or, at least, such a possibility is not *ruled out* by shape functionalism; there may be some facts external to the basic theory of shape perception itself that make such a case impossible.

¹⁹ Shape presentationalism goes very naturally with a commitment to COMMONALITY: if all suarish experiences represent a single, determinate property, it is reasonable to think that that property will be Euclidean squareness, since that is the property with which we seem to be acquainted in having suarish experiences. But shape presentationalism is compatible with a denial of COMMONALITY: one could hold that all suarish experiences represent the

One way to decide between these views, then, is to ask whether we can generate an example of two subjects whose squarish experiences represent different determinate properties. The inverted spectrum has often been used in this way to argue for a placeholder account of *color* perception: subjects who are spectrum inverted relative to each other represent two different color properties²⁰ when having the same type of color experience (or so the argument goes).

In the contemporary literature, such arguments are often framed in terms of ‘Twin Earth’ cases. Putnam’s original Twin Earth example can be seen as providing an argument in favor of a placeholder account of the contents of *water* perceptions.²¹ Since Oscar’s and Twin Oscar’s ‘waterish’ experiences are both veridical, despite the fact that they are caused by different determinate chemical-composition properties (H₂O and XYZ, respectively), there can’t be any *single* determinate property that all waterish experiences attribute to the liquids perceived. Instead, we should think of a given subject’s waterish experiences as representing *whatever* determinate chemical-composition property has in fact typically caused such experiences in that subject’s environment.²²

Chalmers’s strategy in arguing for a placeholder account of shape perception is to construct a series of cases involving shape experience that he takes to be relevantly analogous to Putnam’s Twin Earth.²³ Before getting to the specifics of Chalmers’s

same determinate property, but deny that the property represented is Euclidean squareness. The appeal of such a combination of views is not immediately obvious, but I will address a possible motivation for this type of position at length in §5.4.

²⁰ Where these color properties are to be identified with particular textures or spectral-reflectance profiles.

²¹ In his original paper (Putnam 1973), Putnam himself is focused on the contents of *language*, rather than those of mental states, and he argues only for *semantic* externalism on the basis of his Twin Earth example. Others, such as Tyler Burge (Burge 1979), later extend this type of argument to include the contents of *propositional attitudes*, like belief. The kind of ‘placeholder’ analysis of *perceptual* contents discussed here, and the argument for that analysis on the basis of Twin Earth-style cases, is developed by writers such as Ned Block (Block 1990) and Chalmers (Chalmers 2006), though it is at least suggested by Putnam’s own later work (Putnam 1981). Block’s Inverted Earth case is a version of the inverted spectrum argument that makes use of the contemporary Twin Earth paradigm; Block’s argument can be seen as a defense of a ‘placeholder’ account of color perception.

²² We might question whether perceptual experiences really attribute ‘high-level’ natural kind properties—like the determinate chemical composition H₂O—to the objects of perception *at all* (see (Siegel 2010)). It won’t matter much for my purposes whether they do, since the shape properties I am concerned with are paradigms of *observational* properties, and so they will feature in the contents of perceptual experience on any view that allows that perceptual experiences do indeed have contents.

²³ (Thompson 2010) also constructs such a case—which he, following Susan Hurley (1998), calls ‘El Greco World’—in defending a placeholder account of shape perception. Burge (1986), Davies (1993), and McGinn (1989) also discuss related cases, though to somewhat

argument, it will be helpful first to consider the general structure of Twin Earth arguments.

In order to formulate a Twin Earth argument in support of a placeholder account of some feature X , we need a pair of subjects (the ‘twins’)—one on Earth, one on an alien planet—who are both having X -ish experiences. We then need to identify two distinct properties—the *Earthly property* (EP), and the *alien property* (AP)—such that: (a) due to differences in the twins’ local environments, the typical cause of the Earthly twin’s X -ish experiences has been EP , while the typical cause of the non-Earthly twin’s X -ish experiences has been AP ; and (b) we intuitively judge that each twin’s X -ish experiences are veridical when, and only when, the perceived object instantiates the property that has typically caused X -ish experiences in that twin’s own environment.

What I want to flag is that, in a Twin Earth scenario, features of each twin’s *historical environment* play a crucial role in fixing the target referent for that twin’s X -ish experiences. Our judgment that Twin Oscar perceives veridically when he has a waterish experience of XYZ is driven by the fact that XYZ is the drinking liquid he knows and loves; it is the liquid with which he has a past history. By contrast, if *Oscar* travels to Twin Earth, has a waterish experience of XYZ, and reports that he has found water, he is mistaken—his experience has misled him. Lacking any *previous* causal interaction with XYZ, Oscar does not have perceptual states that represent it; his waterish experiences represent H_2O .²⁴ A Twin Earth argument works by revealing that what counts as a veridical X -ish experience depends on what *has*, in a given subject’s *past*, caused X -ish experiences in that subject.

We can now turn to Chalmers’s Twin Earth argument against shape presentationalism, which he employs in the course of defending a ‘thoroughgoing spatial functionalism’ (Chalmers forthcoming, p. 22). Chalmers begins by imagining that there is a planet just like Earth—where that means the planet contains a ‘doppelganger’ of each Earthly object and person—traveling past us in a uniform direction at 87% the speed of light. Call this planet ‘Lorentz Earth’. One of the surprising results of STR is that objects moving past a given observer at close to light speeds will undergo ‘Lorentz contraction’ relative to the observer’s frame of reference; a meter stick will be less than a meter in length, a circle will be ‘squished’ into an elongated ellipse. So objects on Lorentz Earth, when considered from Earth’s reference frame, will be contracted in the direction of Lorentz Earth’s motion relative to us, by a factor of two. For example, a 2'-by-2' square object on our planet,

different ends. I’ll focus on Chalmers’s argument here, since his version gets most directly at the issues raised by STR.

²⁴ At least, Oscar will not *initially* have states that represent XYZ. It is plausible that if Oscar lives on Twin Earth for an extended period, he will, after suitable ‘adaptation’ to his new planet, have waterish experiences that represent XYZ. This phenomenon is sometimes referred to as ‘slow-switching’. What is crucial here is that, in Twin Earth arguments, historical context sets the veridicality conditions for a subject’s ‘pre-adaptation’ experiences.

if it were placed on Lorentz Earth, would become a 2'-by-1' rectangle relative to our frame.

Now suppose that an Earthling named Albert is in Fenway Park in Boston, looking directly down at third base²⁵ (call this object "Third Base_E"). Albert will be having a squarish experience of the 15"-by-15" object²⁶ he is looking at. His twin on Lorentz Earth, Twin Albert, will also be looking down at *his* third base (call this object "Third Base_L"), in Lorentzian Fenway Park. Though the object Twin Albert is looking at is a 15" by 7.5" rectangle relative to Earth's reference frame, it is square relative to Lorentz Earth's reference frame, so Twin Albert will have a squarish experience of it.

Chalmers suggests that this Lorentz Earth scenario is relevantly similar to Putnam's Twin Earth, and he uses it in formulating an argument against shape presentationalism. We can reconstruct his argument as follows:²⁷

- 1) Albert has a squarish experience of Third Base_E and Twin Albert has a squarish experience of Third Base_L.
- 2) Albert's experience veridically represents the shape of Third Base_E iff Twin Albert's experience veridically represents the shape of Third Base_L.²⁸
- 3) Albert's experience veridically represents the shape of Third Base_E: that is, the shape property represented by Albert's perceptual experience is the shape property Third Base_E actually has.²⁹
- 4) So Twin Albert's experience veridically represents the shape of Third Base_L: that is, the shape property represented by Twin

²⁵ Perhaps he is Albert Pujols, having just uncharacteristically hit a triple in a game against the Red Sox, and he is pausing to enjoy the rare opportunity to observe third base from his unfamiliar vantage point.

²⁶ See Rule 1.06 of Major League Baseball's *Official Baseball Rules* <http://mlb.mlb.com/mlb/downloads/y2014/official_baseball_rules.pdf>.

²⁷ See Chalmers (forthcoming, pp. 19-20). A closely related argument appears in Chalmers (2012, pp. 327-328). Note that Chalmers's full argument in favor of shape functionalism involves extending the conclusion from the Lorentz Earth case to two further cases ('Absolute Lorentz Earth' and 'El Greco World'). Since I will be arguing that Chalmers's initial conclusion based on the Lorentz Earth case should be rejected, I won't explicitly address his consequent arguments about the other cases, which depend on that conclusion.

²⁸ This premise is based on symmetry considerations inherent in STR: since there is no privileged reference frame, it would be arbitrary to attribute illusory perceptions to just *one* of the two observers. Thus, we should count Twin Albert's experience as veridical if (and only if) we are prepared to count Albert's as veridical.

²⁹ This premise is an expression of VERIDICALITY: since this is a typical shape experience of a competent perceiver of shape, we should assume that it is veridical.

- Albert's perceptual experience is the shape property Third Base_L actually has (2,3).
- 5) The shape property instantiated by Third Base_E is **squareness**; the shape property instantiated by Third Base_L is **2:1-rectangularity**.
 - 6) The shape property represented by Albert's experience is **squareness** (3,5).
 - 7) The shape property represented by Twin Albert's experience is **2:1-rectangularity** (4,5).
 - 8) **Squareness** and **2:1-rectangularity** are not the same property.
 - 9) So not all squarish experiences represent the same property (1,6,7,8).
 - 10) If shape presentationalism is true, then all squarish experiences represent the same property.
 - 11) Thus shape presentationalism is false (9,10).³⁰

Note that Chalmers's argument depends on several key uses of shape terms in steps (5) through (8) (the relevant instances are highlighted in bold). These instances of shape terms are supposed to *express* the idea of, as Chalmers puts it, 'ordinary squareness' (and other 'ordinary' shape properties) in Earth English (as opposed to *mentioning* terms of Earth or Lorentz English). Ordinary squareness is thus supposed to be the *referent* of our term 'square' and, relatedly, the property represented by

³⁰ Here, I have stated the conclusion of Chalmers's argument as the denial of shape presentationalism, a thesis about the contents of shape experience. But Chalmers's explicit statement of the argument's conclusion is that the *word* 'square' is 'Twin-Earthable'; he explicitly states a parallel claim about the contents of squarish *experiences* only after his discussion of the 'Absolute Lorentz Earth' case (Chalmers forthcoming, pp. 21-22). Still, the *premises* of Chalmers's argument concern the veridicality of the twins' shape experiences (not the truth of their utterances), and so the argument only makes sense if Chalmers is assuming that perceptual and linguistic contents are to be given a unified analysis. Thus, the conclusion of Chalmers's argument can reasonably be taken as a claim about the contents of squarish experiences (as I have here) and not just as a claim about the 'Twin-Earthability' of the word 'square'. Indeed, given that Chalmers's premises are framed in terms of the veridicality of experiences, to the extent that there is a significant gap between the conclusion about the word 'square' and that about squarish experiences, only the *latter* would be supported by the argument. In what follows, I will be assessing the argument's force as an argument about the contents of shape experience (addressing corresponding issues about the reference of shape terms only in passing). Note that the version of the argument given in (Chalmers 2012) is spelled out entirely in terms of the reference of spatial expressions; it is, however, introduced to address a question about the possibility of shape *illusions* (Chalmers 2012, pp. 328). Again, this move—from a question about the veridicality of shape experiences to an argument about the contents of spatial expressions—only makes sense if Chalmers is assuming that the two kinds of content are to be given a unified analysis.

squarish experiences of Earthly subjects (see step (6)). The key question I now want to press is: What property is ordinary squareness; what, on Chalmers's view, is the property represented by our squarish experiences here on Earth?

One way to resist the argument would be to take ordinary squareness to be a property Chalmers labels 'rest-squareness': the property an object has if and only if it is square relative to its own reference frame.³¹ This would undermine steps (5) through (9) of the argument, since rest-squareness is a property that *both* twins' bases have: Third Base_L is, like Third Base_E, square relative to its own reference frame. So, if we were to take rest-squareness to be ordinary squareness (that is, take it to be the referent of our ordinary term 'square'), the crucial claim—in premise (5) of Chalmers's argument—that Third Base_L is *not* square (but rather a 2:1 rectangle) would be false.

Chalmers considers the proposal that ordinary squareness is rest-squareness, but he rejects it on the grounds that it is inconsistent with the use of spatial vocabulary in STR. In particular, Chalmers points out that we describe objects as 'contracting' when they accelerate in STR, even though their *rest*-shapes remains constant; he takes such statements to be inconsistent with the proposal that the referent of our ordinary term 'square' (and, correspondingly, the property represented by our squarish experiences) is rest-squareness.³²

This leaves us with the question of what, on Chalmers's shape functionalist picture, ordinary squareness—the property represented by our squarish experiences on Earth—turns out to be. Chalmers is not very explicit in answering this question,³³ but we can see what he might have in mind if we recall that his argument is supposed to be a shape analogue of Putnam's original Twin Earth argument. So, as noted above, for the argument to work, we need to find a pair of *distinct* properties, *EP* and *AP*, such that: (a) Albert's squarish experiences have typically been caused by instances of *EP* (and not *AP*), while Twin Albert's have typically been caused by instances of *AP* (and not *EP*); and (b) we intuitively judge that these historical differences fix the contents of the twins' respective squarish experiences, so that Albert's experiences represent *EP*, while Twin Albert's represent *AP*. "Ordinary squareness" is Chalmers's term for the property represented by Albert's squarish experiences. The question of what ordinary squareness amounts to, then, is just the question of what plays the role of *EP* in Chalmers's argument. And, given STR, a

³¹ Standard scientific jargon suggests this property should be labeled 'proper squareness', on analogy with 'proper length' (where 'proper' does not mean 'correct' but rather 'relative to *one's own* frame').

³² See Chalmers (2012, p. 329; forthcoming p. 20). Chalmers's reliance here on the use of spatial vocabulary in STR is a somewhat curious move, given that his *own* account of the reference of ordinary spatial terms also seems to deviate from standard STR usage (see fn. 37 below).

³³ He says only that ordinary shape terms pick out 'relativistic properties in a relativistic universe' like ours (Chalmers 2012, pp. 325-326); but this leaves open *which* relativistic property 'square' picks out.

natural thought is that *EP*—the property that has typically caused squarish experiences on Earth, but not on Lorentz Earth—is the property of *squareness relative to Earth*, while *AP* is the distinct property of *squareness relative to Lorentz Earth*.

Given the argumentative context, however, these expressions—‘squareness relative to Earth’ and ‘squareness relative to Lorentz Earth’—do not provide a particularly helpful way of picking out the properties in question, since they both *use* spatial terms, and the question at issue concerns how the reference of our spatial representations is fixed. Is ‘square’, as it appears in the expression ‘square relative to Earth’, to be given a *presentationalist* analysis, or a *placeholder* analysis? Answering *this* question would seem to require antecedently settling the question at issue. The argumentative context also presents a further difficulty. I am assessing shape functionalism as a way to resolve the puzzle described in §5.1 by accepting *non-instantiation* and rejecting COMMONALITY. But the use of terms like ‘square relative to Earth’ might suggest that we are, contrary to *non-instantiation*, attributing ‘purely spatial’ properties to objects; so using such terms might seem inconsistent with the shape functionalist’s position.

To avoid these difficulties, we can frame Chalmers’s argument in accord with Minkowski’s claim about ‘space by itself’ fading into shadow. On Minkowski’s picture, the only underlying physical properties that objects in our universe have are *spatiotemporal* in nature; thus, the properties generating our squarish experience will not be *spatial* properties, like squareness (relative to Earth), at all. The underlying *spatiotemporal* property that Third Base_E has (call it ‘ST_E’) will be the real referent of our term ‘square’ and the property represented by our squarish experiences (since it is the property causing those experiences). ST_E is, then, the property playing the role of *EP* in Chalmers’s argument.³⁴

The shape-functional picture that now emerges is the following: In Albert’s historical, Earthly environment, squarish experiences have been caused by a particu-

³⁴ What property, exactly, *is* ST_E? It is the underlying property of the Minkowski space-time manifold that ‘grounds’ squareness-relative-to-Earth, in something like the way that H₂O ‘grounds’ water. This property won’t be one that has a neat specification in terms of the Minkowski space-time geometry itself: it will have to be a property instantiated by objects that are square relative to Earth, but not by objects that are square relative to Lorentz Earth, even though there is no *intrinsic* difference between those objects in terms of their Minkowskian features (the two objects’ space-time ‘worms’ or ‘world-volumes’ will occupy congruent regions of Minkowski space-time). Instead, ST_E will have to be a kind of *relational* property of the spatiotemporal manifold: an object will instantiate it in virtue of (a) the space-time region occupied by the object’s world-volume and (b) the angle between the object’s world-volume and the world-volume of Earth (‘angle’ here corresponds to two objects’ velocities relative to each other; for an object at rest on Earth, like Third Base_E, the two world-volumes will be parallel, so the angle will be (by convention) equal to zero). See (Sattig 2015, p. 222) on the relation between objects’ Minkowskian world-volumes and their shapes. The important point is that the analysis in terms of world-volumes can be spelled out purely in terms of the *spatiotemporal* features of the situation; this gives us a way to think about ST_E that does not itself rely on concepts of spatial properties, whose analysis is under dispute.

lar, determinate *spatiotemporal* property, ST_E . In Twin Albert's historical Lorentzian environment, squarish experiences have been caused by a *different* spatiotemporal property (call it ' ST_L '; this is the property that plays the *AP* role in the argument). The particular, determinate spatiotemporal property with which each twin has interacted, in turn, is fixed as the property represented by that twin's squarish experiences: for Earthlings like Albert, squarish experiences represent ST_E ; for Lorentzians like Twin Albert, squarish experiences represent ST_L .³⁵

The problem with this proposal is that, although ST_E and ST_L are indeed properties we can identify as distinct causes of squarish experiences in the two twins' respective historical environments—they satisfy condition (a) in setting up the Twin Earth scenario—it is *not* plausible that those different historical interactions *fix the contents* of the twins' squarish experiences in the manner proposed— ST_E and ST_L don't satisfy condition (b). We do not intuitively judge that an Earthling like Albert has veridical squarish experiences when, and only when, he perceives an object that instantiates ST_E .

In order to bring this out, we can consider the shape perceptions of observers who happen to be moving relative to Earth but who have (unlike Twin Albert) *grown up* on Earth. Suppose that Albert has a normal (non-Putnamian) twin, raised alongside him on Earth, who—instead of pursuing a promising baseball career—elected to enroll in an experimental space-flight program; call him 'Astronaut Albert'. Astronaut Albert is on his first space mission, zooming past Earth in his spaceship at near-light speed (we can imagine that he is moving, relative to Earth, in the same direction and at the same speed as Lorentz Earth). As he zooms along, he is staring down at a base (call it Third Base_A) that he has brought on his spaceship as a reminder of his baseball-playing brother.

Astronaut Albert will be having a squarish experience of Third Base_A; the crucial issue is what the shape-functional account we are now considering implies about the contents and the veridicality of this experience. Since Astronaut Albert's squarish experiences have, until now, occurred on Earth, their content will, on the shape-functional view, be fixed by the property that has typically caused those experiences in his Earthly environment: namely, ST_E . But, since Astronaut Albert is moving relative to Earth, the base he is looking at will instantiate a *different* spatiotemporal property. In fact, given that Astronaut Albert is currently in the same state of motion as Lorentz Earth, his base will instantiate the same spatiotemporal property as *Twin* Albert's base: ST_L . So Astronaut Albert's squarish experience will represent Third Base_A as instantiating ST_E , when it does not—that is, Astronaut Albert will be suffering an illusion.

This verdict is simply not a plausible one, given the way STR is typically understood. In standard STR examples, we have two *Earthling* observers—one on Earth, one flying past Earth at near-light speeds—in precisely the situation of Albert and

³⁵ Though it is a natural way to develop the shape functionalist position, this proposal is not explicitly endorsed by Chalmers (who, as noted above, is simply silent on the question of which specific properties the twins' experiences represent (see fn. 33)).

Astronaut Albert. In such scenarios, each subject will judge that the objects around him have their ‘normal’ shapes, while the objects in the other observer’s frame are ‘compressed’. Albert will have a squarish experience of Third Base_E, but he will not have a squarish experience of Third Base_A; and *vice versa* for Astronaut Albert.³⁶ A key idea in STR is that neither observer in these cases is ‘more right’ in his perceptions than the other. But if we accept the shape-functional claim that the squarish experiences of Earthling subjects represent the particular determinate property ST_E, we would have to conclude that Albert is getting things right, while Astronaut Albert is misled: his squarish experience of the base on his ship and his experiences of objects on Earth as contracted are *illusory*, rather than accurate perceptions of the metaphysical facts.

Those metaphysical facts are strange, to be sure: according to STR, the objects of Earth *actually are contracted* relative to Astronaut Albert’s reference frame. But, in order to capture the truly revolutionary nature of the theory, we need to acknowledge that it indeed concerns the spatial properties of *objects themselves*. If we suppose that Astronaut Albert is simply misperceiving the true shapes of objects, we make it seem as though STR merely describes a certain kind of *perceptual error* that befalls subjects who accelerate out of their native reference frames. But similar perceptual effects already occur in non-relativistic contexts: an astronaut accelerating to merely very fast—but non-relativistic—speeds would misperceive the shapes of objects on Earth, because those objects would look blurry and stretched. STR is not a theory about such perceptual effects; it is a theory about how the spatial properties of objects themselves are different, relative to different reference frames.

One final case brings out just how counterintuitive the consequences can get if we take Earthlings’ historical interactions with ST_E to fix the contents of their squarish experiences. Suppose Astronaut Albert has flown off in his spaceship and then switched off his engines, in order to enjoy a weightless meditation session in space, no longer in view of Earth. Astronaut Albert might be moving at near-light speeds relative to Earth, or he might not; he has lost track of how many times he has fired his engines, and he has not been tracking the Earth’s movements. He now looks at Third Base_A and has a squarish experience. If ST_E is fixed as the property represented by Earthlings’ squarish experiences, whether Astronaut Albert’s experience is veridical will depend not only on the intrinsic properties of the object, nor on

³⁶ Strictly speaking, Astronaut Albert won’t have *experiences* of objects on Earth as contracted, since those objects will be moving past him at far too great a speed for his visual system to discern their shapes at all. As Tim Maudlin emphasizes, talk of subjects’ ‘observations’ of objects in different reference frames in STR is *not* meant to line up with ‘literally what the observer would *see* if she opens her eyes’ (Maudlin 2012, p. 103). But the key point here is that Astronaut Albert’s ‘observations’ of the shapes of objects on Earth (effected via sophisticated measuring devices) will represent those objects as compressed; and, on the view we’re now considering, both those observations and Astronaut Albert’s (literal) visual experiences of objects on his own ship would be ruled inaccurate.

how he is related to it, but on his state of motion relative to a planet halfway across the galaxy.³⁷

Clearly, something has gone wrong here. At base, the shape-functional picture runs off the rails when it loses sight of the *symmetry* that lies at the heart of STR. We do not take one of the two twins in typical STR examples to be suffering from illusions; we emphasize that each has an *equally accurate* take on the spatial properties of the objects in the scenario.³⁸

Chalmers himself emphasizes this symmetry in describing the Lorentz Earth scenario (as noted above in fn. 29), and he even uses spatial terms (presumably in ‘ordinary English’) in the standard way when he writes: ‘From an objective point of view, the situation is completely symmetrical.... Twin Albert is compressed relative to Albert’s reference frame, while Albert is compressed relative to Twin Albert’s reference frame’ (Chalmers forthcoming, p. 19). This is all correct, but it is inconsistent with the way in which (as I’ve just argued) Chalmers’s *own view* requires our ‘ordinary’ spatial terms to be tied to the specific properties that have played the relevant role in our *Earthly* environment. ‘Compressed’ is an ‘ordinary’ spatial term that, on Chalmers’s picture, refers to the determinate property of the spatiotemporal manifold that has generated experiences of compression on Earth. So Albert’s objects are not compressed, while Twin Albert’s are, by Chalmers’s own lights. And, when we consider two *Earthling* observers moving very rapidly relative to each other, the symmetry recedes even further: now, we must convict one of the subjects, but not the other, of widespread *perceptual* and *judgmental* error.

³⁷ Another difficulty for this view comes out at the level of *language*: STR examples often invoke a pair of spaceships moving relative to one another, where *neither* is specified as being at rest relative to Earth. In these cases, if we were to take our historical interactions with objects on Earth to fix the reference of our spatial *terms* along with the contents of our spatial *experiences* (as Chalmers suggests we should), we would always need to know how the objects described were moving *relative to Earth* in order to know how we should apply those terms, and whether one (or both) of the observers (assuming them to be Earthlings) should count as making false claims about objects’ shapes. Similarly, in considering a standard example where a meter stick is accelerated to near-light speeds, we would need to know whether it is accelerating *out of* Earth’s frame, or instead *into* Earth’s frame (that is, whether it is at rest relative to Earth at the *beginning* or at the *end* of its acceleration) in order to know whether we should say that the stick undergoes Lorentz *contraction* (which is what we would say in the former case) or rather Lorentz *expansion* (the verdict we would reach in the latter case). This is out of step with the way that spatial concepts, and spatial terms, are actually employed in STR, and with Chalmers’s own claim that his account of the meaning of ‘ordinary’ spatial terms is supported by the language of ‘Lorentz *contraction*’.

³⁸ Note that the kind of symmetry in question here is not the complete qualitative matching of two subjects’ mental lives typical of Twin Earth cases. It is the more mundane symmetry of inertial reference frames (and of the perspectives of observers within those frames, whether or not the observers are qualitative duplicates) that is invoked when discussing STR in physics.

So it seems the Lorentz Earth scenario cannot be used as a Twin Earth argument in support of Chalmers's shape-functionalism: our judgments about the veridicality of Earthling space-travelers' perceptions—the judgments that correspond to the most natural interpretation of STR—reveal that we do not take the historical environment of a given subject to fix the contents of her squarish experiences. One of the conditions for a successful Twin Earth argument—condition (b)—is not met.³⁹

At this stage, one might object that Chalmers does not require such a tight analogy between Lorentz Earth and Twin Earth for his argument to work. The failure of the analogy stems from the way in which, unlike in the case of natural kinds like water (where we take a subject's *historical* circumstances to fix the contents of her waterish experiences), we take the contents of a subject's shape experiences (given STR) to depend on features of her *current* environment (specifically, on her current state of motion). So perhaps, in the case of shape experience, we can weaken the conditions needed for a standard Twin Earth argument by jettisoning the reference-fixing role of the historical environment. Instead, we might analyze the Lorentz Earth case in light of a modified version of shape functionalism, one on which a given squarish experience represents the property that plays the relevant role in generating squarish experiences in a subject's *current* environment.⁴⁰ For an Earth-bound subject like Albert, this property will be ST_E , and so his squarish experience will be veridical; for Astronaut Albert, on his spaceship, this property will be ST_L (the very property his base instantiates), so his experience will be veridical as well. We can thus recover our intuitive judgments about the symmetry of the situation and the veridicality of *both* twins' experiences.

The reference-fixing role of the historical environment is not, however, an optional extra in developing the kind of placeholder account of perceptual content Chalmers defends. Without it, the placeholder view will fail to set any substantive constraints on the relevant perceptual contents; virtually all experiences will count as veridical. Take the present proposal about shape experience: a given squarish experience represents whatever property plays the role of causing squarish experiences in a subject's *current* environment. Now suppose we have a subject—Fun Albert—who enters a funhouse. He sees a rectangular object that, due to the funhouse's distorting mirrors, causes a squarish experience in him. Given that Fun Albert is *currently* in the funhouse environment, rectangularity is the property that plays the relevant role in generating squarish experiences in his current environment. So his experience will count as a veridical perception of the object's rectangularity. But surely experiences

³⁹ I've argued that Lorentz Earth does not provide Chalmers with a suitable analogue of Twin Earth, and so cannot be used in defending shape-functionalism, but I have yet to say where Chalmers's argument itself (as spelled out on p. 9) goes wrong. That will have to wait until I have developed my own account of shape perception in STR; see below, p. 24.

⁴⁰ This would make spatial content a sort of *indexical* content: the properties picked out by a token state's spatial contents would depend on the circumstances in which *that token* is instantiated. Thanks to an anonymous referee from *Mind* for this suggestion.

in funhouses are paradigms of illusions: the experience in question is not veridical. Even more troublingly, consider how the present proposal would have us assess the squarish experiences of a recently-envatted brain. Another subject—Vat Albert—is walking along on Earth, when he is suddenly snatched away and his brain is suspended in an evil scientist’s vat; he is fed inputs that match those that he used to get from square objects on Earth. Since Vat Albert is *currently* in the vat environment, where a certain kind of electrical signal plays the relevant role in generating squarish experiences, his experience will count as veridically representing that electrical-signal property.⁴¹

What these examples illustrate is that a placeholder view that does not take perceptual contents to be historically fixed is, in an important sense, *empty*. It places no real constraints on what counts as veridicality for a particular experience; and so it does not allow us to recover our intuitive judgments about everyday (or fantastical) instances of illusion. In trying to assess the question raised by STR—the question of how our shape experiences represent the objects we perceive—such a view is not a serious contender.⁴²

So, to sum up: Twin Earth arguments, as employed in defense of placeholder accounts of perceptual experience, require us to see the contents of the relevant experiences as *historically anchored*. Such arguments work by revealing that the content of each twin’s experience is *fixed* by the features present in that twin’s own past environment. But such historical anchoring seems to be absent in STR. In standard Einsteinian relativity cases, reference to shape properties shifts immediately: whatever a subject’s history may be—even if she is an Earthling enjoying her very first space voyage in a non-Earth reference frame—the veridicality of her shape perceptions depends only on the shapes objects have relative to her *current* state of motion.⁴³ Having acknowledged the lack of historical anchoring of shape contents in STR, a proponent of a *placeholder* view cannot hope to salvage the account by jettisoning the historical-anchoring requirement. On a *presentationalist* view, there would be no need to appeal to historical context: according to the presentationalist, squarish experiences *in themselves* represent a single, determinate shape property, so a squarish experi-

⁴¹ As an anonymous referee from *Mind* points out, this version of the placeholder view will also have trouble accommodating the intuitive truth of statements Astronaut Albert, while on his ship, makes about the shapes of objects he observed *in the past*, when he was back on Earth (e.g., ‘That *was* square’, in reference to an earlier encounter with Third Base_E): if Astronaut Albert’s term ‘square’ refers to ST_L when he is on his ship, such statements will have to be regarded as false.

⁴² Note that the placeholder view Chalmers himself endorses is not empty in this way. Chalmers emphasizes that *temporary* illusions are possible on his picture, because the content of a particular experience is fixed by facts about the *historical environment* of the subject. For example, in explaining why a *recently*-envatted brain suffers illusions, even though, on his view, an *always*-envatted brain does not, Chalmers writes, ‘my conception of external reality is anchored to the reality in which I have lived most of my life’ (Chalmers 2010, p. 474).

⁴³ That is, STR cases do not exhibit the ‘slow-switching’ characteristic of genuine Twin Earth cases (see fn. 24).

ence comes with determinate content ‘built in’. But on a placeholder view, a squarish experience has no built-in determinate content. Historical anchoring is the only way for the placeholder view to set substantive veridicality conditions; abandoning the historical-anchoring requirement amounts to stripping the view of any real content.

5.3. SHAPE PRESENTATIONALISM IN A RELATIVISTIC UNIVERSE

What then, should we say about the contents of shape perception in a relativistic universe? Recall the puzzle from the introduction: Euclidean squareness is the property we reason about in doing geometry, but (according to *non-instantiation*) that property is not instantiated by any physical object in our universe. So if Euclidean squareness were the property represented by our squarish experiences (that is, if COMMONALITY were true), all those experiences would be illusory. Chalmers’s shape functionalism was supposed to be a way of assigning contents to our shape perceptions without implying this kind of universal error. Such a view would allow us to maintain VERIDICALITY for our shape perceptions, but only by divorcing their contents from the Euclidean shape concepts we employ in *a priori* geometrical reasoning: the determinate property represented by our squarish experiences would turn out to be the spatiotemporal property ST_E , rather than Euclidean squareness. Having rejected Chalmers’s argument, we are now free to avoid this bifurcation of our empirical shape concepts and our *a priori* ones; that is, we are now free to accept COMMONALITY.⁴⁴ But does holding on to COMMONALITY mean that we will have to give up VERIDICALITY and accept an error theory for shape perception?

I want to suggest that it does not. We can have our COMMONALITY and our VERIDICALITY, too. In order to see how, we need to examine more closely the threat of NON-INSTANTIATION that STR supposedly raises for any such view. The idea was that STR dislodges the separation between space and time, between spatial properties and temporal ones. In Newtonian space-time, the spatial dimensions are in a strong sense separable from the temporal dimension. An object’s being square has nothing to do with its (or anything else’s) state of motion (where motion is a matter of how the object’s spatial position changes over *time*); an object can be said to be square *absolutely*. But in the kind of space-time we have in STR, Minkowski space-time, there is no clean separation of space and time. We cannot specify a set of spatial properties that apply to an object irrespective of which inertial frame—which state of motion,

⁴⁴ Of course there might be *other* reasons to reject COMMONALITY, given STR. Peacocke (1989), for example, presents a kind of *transcendental* argument for an externalist account of spatial perception, on which the geometry of the environment fixes the spatial content of experience. If we were to combine such a view with NON-INSTANTIATION, we would be forced to deny COMMONALITY (though Peacocke himself actually seems to endorse a version of COMMONALITY similar to that advanced below, which suggests that he instead denies NON-INSTANTIATION (Peacocke 2015, p. 382)).

which way of progressing spatially through *time*—is regarded as ‘at rest’.⁴⁵ This forces us to see the spatial properties of the object as tied inextricably to its temporal properties. And that was supposed to rule out the possibility that Euclidean squareness—a purely spatial property—could actually be instantiated by physical objects in our universe.

But we are not, in the end, forced to abandon spatial notions altogether within STR; we do not need to accept *non-instantiation*. We simply need to distinguish two ideas of ‘objective’ space (space that has ‘an independent reality’, in Minkowski’s terms), only one of which is ruled out by STR. STR *does* entail that no objects in our universe instantiate a property that a certain pre-theoretically intuitive picture would ascribe to them. Chalmers calls this property ‘perfect’ or ‘Edenic’ squareness; we might call it ‘absolute Euclidean squareness’. And it is the *absoluteness* of the property that really rules it out as a property that objects in our relativistic universe could have. If an object is square relative to its own reference frame, it will not be square relative to a frame moving with respect to it; so we can’t say that an object is *absolutely* square, in the sense that it is square relative to every reference frame (or square irrespective of reference frame), given STR.⁴⁶

We *can*, however, *specify* a reference frame, and *then* make a claim about an object’s shape: we can say, for example, that, relative to Earth’s reference frame, Third Base_E is square. This claim, though it is explicitly acknowledged to be non-absolute (in that it mentions a particular reference frame that is relevant to its interpretation), is genuinely a claim about a spatial property of the object, in a perfectly *objective* way. That is: the claim is not ‘true’ in a weak, relative sense, where that means the statement could correctly be evaluated as true by one assessor, false by another.⁴⁷ Even within STR, it is *objectively* true that Third Base_E is square, relative to Earth’s reference frame.

⁴⁵ This is just to make the familiar point that there is no unique ‘foliation’ of the Minkowski space-time manifold into a series of spatial simultaneity slices (or ‘three-planes’).

⁴⁶ The sense in which I am here using the term ‘absolute’ corresponds to one half of the contrast that Michael Friedman labels ‘absolute-relative’ (as opposed to ‘absolute-relational’ or ‘absolute-dynamic’). Friedman defines absoluteness, in this sense, in the following way: ‘an absolute element of the spatio-temporal structure is one that is well-defined independently of reference frame or coordinate system’ (Friedman 1983, p. 62). My point is that, given STR, squareness is not a property that is absolute in this sense. (It is a further question whether squareness is absolute in Friedman’s other two senses; it seems to me that it is, in both of these other senses, but I will not go into those issues here.) My use of ‘absolute’ also corresponds to the kind of absoluteness denied to space and time by Paul Horwich’s ‘Relativist Thesis G’: ‘Space and time are relative [i.e., not absolute] in the sense that certain magnitudes such as duration and distance vary from one frame of reference to another’ (Horwich 1978, p. 400).

⁴⁷ See MacFarlane (2014) for a development of this kind of relative notion of truth. The point I am making about the objectivity of non-absolute claims about the shapes of objects in STR is that such claims are *not* ‘assessment sensitive’ in MacFarlane’s sense.

Indeed, it is *only* through the use of these specifications of reference frames that we can make sense of the canonical claims of STR. When we say that STR reveals that a moving rod contracts, we are employing spatial notions, and those notions can only be employed within a framework that has room for the idea of space (as opposed to the indissoluble space-time fusion that is the ‘absolute’ manifold of STR). Our attributions of squareness to objects within STR make use of our spatial concepts in an intelligible way, despite Minkowski’s claim that ‘space by itself’ would fade into the shadows under Einstein’s theory. And, crucially, the *spatial* manifold we pick out by defining a set of coordinates on Minkowski space-time (by specifying what inertial reference frame is to be considered at rest) is a *Euclidean* space.⁴⁸ The concepts of length and shape we use in talking about the theory—for example, the shape concepts we use when we say that a moving sphere will be compressed into a flattened ellipsoid—are our familiar Euclidean spatial concepts. It is this fact that allows us to make sense of the theory in the first place; we describe the results of STR in terms of the Euclidean spatial properties objects have relative to various frames of reference because such properties are the only genuinely *spatial* ones on which we have an intuitive grasp.

I think we now have the tools to defend a version of shape presentationalism that is consistent with both COMMONALITY and VERIDICALITY. Squareness, as it shows up in the contents of our squarish experiences, is the very Euclidean shape property we employ in doing geometrical proofs. As such, it is *not* picked out in virtue of the role it plays in generating our experiences; it is picked out by the determinate, *a priori* geometrical definition ‘having a quadrilateral surface with four equal sides joined at four right angles’. We have a more direct grasp on this property than Chalmers’s placeholder analysis—on which squareness is ‘*whatever* property in fact plays the relevant role in causing our squarish experiences’—would suggest. This

⁴⁸ This is an expression of the fact that Minkowski space-time is *pseudo*-Euclidean. One might worry that once we move to the General Theory of Relativity (GTR), according to which the space-time of our universe is *pseudo-Riemannian*, we will no longer have even frame-relative Euclidean spaces, and so our spatial concepts (if they are indeed Euclidean) will not veridically apply to the objects we perceive. While I do not want to go into too much detail on the implications of GTR, as opposed to STR, for accounts of our spatial cognition, I will say briefly: Even within GTR, we can (and *must*, at the theoretical limit) define *local* areas of space-time that are roughly Minkowskian, and we can determine spatial coordinate systems within those areas. All of our actual experience of the world takes place within such roughly Euclidean spaces, so we can regard our attributions of Euclidean spatial properties to the objects around us as veridical because, relative to a set of coordinates defined within a local pocket of our pseudo-Riemannian universe, and allowing for a degree of indeterminacy (to accommodate the fact that space on these scales is only *roughly* Euclidean) that will have to be acknowledged in any case in order to allow for veridical perception in *any* domain (see Stazicker (2011 and forthcoming)), objects instantiate such Euclidean spatial properties. See Hopkins (1973) for an insightful discussion of how our experiential attributions of shape properties (our ‘visual geometry’) can be seen as consistent with the *slight* divergence from Euclidean geometry that, according to GTR, we encounter in our actual spatial environment.

means that we do not have to wait on the findings of empirical science to reveal what determinate property a squarish experience represents; we know, as COMMONALITY insists, that it is the very property of Euclidean squareness—a property on which we have an *a priori* grasp—that features in the contents of our perceptions of squareness. The one addition we must make to our account (if we are to hang on to VERIDICALITY) is that the intuitive, *a priori* property of Euclidean squareness must be supplemented by an extra ‘parameter’—a specification of a reference frame—if it is to be a property instantiated by objects in our universe.⁴⁹ Once that parameter is filled in—by specifying the current reference frame of the observer having the perceptual experience in which the property of squareness features—the property attributed by the experience will be one the object perceived can genuinely have. This extra parameter, though it is not itself a part of the *presentational* content of squarish experiences (since it does not show up *explicitly* in those experiences themselves), allows our experiences to veridically present Euclidean squareness.

This move away from what might be seen as the *fully* intuitive, pre-theoretical analysis of the contents of our shape perceptions—an analysis in terms of *absolute* Euclidean squareness—does not give us reason to reject shape presentationalism, or to deny that it is Euclidean squareness we attribute to objects, in favor of a placeholder view, on which the property *we* attribute to objects when we have squarish experiences is a different property from that attributed by Lorentzians when *they* have squarish experiences. Albert attributes Euclidean squareness to Third Base_E (relative to his reference frame); Twin Albert attributes *the very same property*, Euclidean squareness, to Third Base_L (relative to *his* reference frame). There is an implicit parameter that varies between the two observers, but the *property* attributed by each—the property that features in the contents of both of the twins’ squarish experiences—is the very property that features in our *a priori* geometrical reasoning.

It might be objected that, on the account just sketched, it is *not* the same property that is represented by Albert’s and Twin Albert’s experiences, since ‘Euclidean squareness (relative to Albert)’ is a *different relational property* from ‘Euclidean squareness (relative to Twin Albert)’. So (the objection goes) the account is not really defending presentationalism (the view that all squarish experiences represent the *same* property) at all.

Suppose, for the moment, that we were to grant that there are indeed two distinct properties here. We can still insist that there is a single *relation* represented by both observers’ squarish experiences—the relation of *being square relative to the observer*.⁵⁰ And, crucially, ‘squareness’, as it appears in the formulation of that relation, does

⁴⁹ How we should understand the role of this reference frame ‘parameter’ will be the subject of §5.4.

⁵⁰ On the view that the twins’ experiences represent the same *relation*, are we compelled to accept that those experiences represent different *properties*? We might take Albert’s experience to represent *squareness-relative-to-Albert*, and Twin Albert’s to represent *squareness-relative-to-Twin-Albert*. Since Albert and Twin Albert are different objects, this would indeed suggest that there are two different relational properties in play. This, in turn, would suggest that

not allow for the possibility that different particular properties might, in different circumstances, be picked out by *that* term. It maintains its standard, Euclidean meaning, and the relation represented by all squarish experiences centrally involves the property of Euclidean squareness.⁵¹

In order to see that it is not particularly problematic to take the content of squarish experiences to include a determinate *relation*, or ‘parameterized property’, we can compare the case of experiences of left and right. Chalmers (forthcoming) considers a view on which a single particular relation, ‘being to the left of’, can be represented by each of two different ‘leftish’ experiences had by two different subjects, even if those subjects are facing each other (and so two different directions in (absolute?) space are being picked out). According to Chalmers, such a view—on which the same relation, but not the same absolute property, is represented by all experiences of a certain type—would count as a presentationalist view. That is essentially the kind of view I am proposing in the case of shape experience. For we

there is a narrow sense in which we might think of squarish experiences as ‘Twin-Earthable’: the squarish experiences of the twins, while representing a single common relation, would be associated with two different relational properties (note, though, that this narrow sense of Twin-Earthability does not in any sense support—and may not even be compatible with—a placeholder view of shape experience (see fn. 51)). But we can also understand the idea of *squareness-relative-to-the-observer* by thinking of the observer as the ‘center’ of a centered world. The idea then would be that the contents of shape perceptions are to be evaluated relative to *centered* possible worlds. On this proposal, Albert’s and Twin Albert’s squarish experiences will both represent the same centered-world property. This proposal seems well-motivated, given that the commonality of perceptual contents across subjects (or within a subject at different times) already requires analyzing those contents in terms of centered worlds. For example, Chalmers notes that on a presentationalist (or, in his earlier terminology, a ‘Russellian’) account of *location* contents, ‘the location property [represented in an experience] must be a relative location property: the property of being a certain distance in front of the perceiver at a certain angle, for example’ (Chalmers 2006, p. 386). In order for this ‘relative location property’ content to be shared across subjects, as Chalmers intends it to be, ‘the perceiver’ will have to be understood as the center of a centered world. Andy Egan develops an account of perceptual content using centered worlds (Egan 2010), though he has a much broader understanding of what can qualify as a ‘centering’ feature (for issues raised by STR, a subject’s path through space-time will suffice).

⁵¹ In §5.4, I argue that the view I’m defending does *not* require us to concede that squarish experiences represent a relation, rather than a property. But for now, the point I’m stressing is that, even if we did make such a concession, the resulting account would be a version of shape *presentationalism*, not a placeholder view. Indeed, a *placeholder* version of this proposal—an account that parses ‘square’ as it appears in the relation ‘square relative to the perceiver’ as ‘whatever normally causes squarish experiences’—is the view that squarish experiences represent whatever normally causes them in a given perceiver, without tying that content to the perceiver’s *historical* context (‘normally’ would instead mean ‘normal in the perceiver’s *current* environment’). So a placeholder version of this view collapses into the *empty* view discussed at the end of §5.2.

could frame the presentationalist view about the contents of leftish experiences in terms of an implicit parameter of the kind I am proposing for our shape experiences: all leftish experiences can be seen as representing the same property, leftness; but that property then needs to be supplemented by filling in an implicit parameter that specifies the orientation of the observer's body in order to pick out a particular spatial direction.

It might now be objected that, while it is plausible to claim that all leftish experiences represent a single particular *relation* or parameterized property (even though they don't represent a single *direction* in space), it isn't plausible to make a similar claim about squarish experiences representing a single relation or parameterized property. According to this objection, leftish experiences seem, *on their face*, to be representing some sort of relation (leftness is *transparently* a relation); but squarish experiences, on their face, seem to be representing a non-relational (i.e., *intrinsic*) property. We discovered the relational aspect of shape properties only through incredibly sophisticated scientific theorizing; so it is not plausible to suggest that all along we somehow represented the relational nature of shape properties in our perceptual experiences of shapes (shapes that seemed to us to be perfectly intrinsic⁵²). A presentationalist view that says experiences of a certain type represent a particular relation is only plausible if that relation shows up *as a relation* in experience; this condition is met for leftish experiences, but not for squarish ones.

I agree that the relational nature of squareness (or, the need for an implicit parameter) is not apparent from within experience itself, the way it (plausibly) is for leftness. But I deny that we can only attribute relational contents to experiences if the relationality shows up *transparently* in those experiences. We can be aware of certain aspects of the contents of our experiences (and concepts), without realizing that more is needed in order for those contents (and concepts) to pick out particular phenomena in the world. When we discover that there is a need for an extra parameter (or, that there is a relational aspect to the phenomenon represented), we do not thereby discover that those concepts fail to have the determinate content we took them to have—we do not, that is, discover that the contents in question are non-presentationalist.

An example of another property with a 'hidden parameter' will be helpful here. Consider Mona, who has lived her whole life in a single city, and has never realized that there are different time zones across the surface of the Earth. Mona has acquired a concept of noon (the time of day when, characteristically, the sun is directly overhead), not realizing that that description could pick out different 'absolute' times, depending on one's location on Earth.⁵³ When Mona discovers that the sun is

⁵² David Lewis emphasizes this point about the seemingly non-relational nature of shape properties (Lewis 1986); I discuss his claims at length in §5.4.

⁵³ The scare quotes on 'absolute' are a way of registering that I am here ignoring the fact that, given STR, *times* are themselves non-absolute (in the sense discussed in fn. 46). The case of Mona will work perfectly well for my purposes if we imagine her living in a Newtonian universe.

not directly overhead at the same time in all locations, she will realize that an additional parameter (or a relativizing term) needs to be included with the proposition ‘It’s noon’ in order to make a determinate temporal claim: ‘It’s noon (in New York)’ is true if and only if it’s the time of day when the sun is directly overhead in New York, while ‘It’s noon’ (with no further parameter filled in) fails to make a determinate claim—one assessable for truth—at all. But, crucially, in learning that her concept of noon needs to be supplemented by an additional parameter, Mona does *not* learn that her concept previously failed to pick out a particular property—that it failed to have the determinate content she took it to have. Similarly, we do not need to propose a placeholder analysis of Mona’s previous ‘noonish’ *experiences*—we do not need to suppose they merely represented *whatever* property in fact caused such experiences in her environment—in order to understand how she could have been representing a property that has a ‘location parameter’. All along, Mona’s noonish experiences represented the determinate property of its being the time of day at which the sun is directly overhead. What Mona realizes when she discovers that there are different time zones is that her noonish experiences, in *presenting* her with that determinate content, also included a *non-presentational* element: unbeknownst to her, those experiences included an *implicit* parameter, which was filled in by her own location.

Similarly, when we discover that, in our universe, squareness is a property objects can have only relative to a particular reference frame, we do not thereby learn that we had no determinate grip on the property of squareness, or that we failed to pick out any particular phenomenon in the world with our attributions of squareness to objects. We simply learn that there is a further parameter to be filled in, which is needed in order for the very property on which we already had a grip—the property of having a quadrilateral surface in a Euclidean plane with equal sides joined at equal angles—to be applicable to the objects we perceive. That is how, even in a relativistic universe, we can hold on to COMMONALITY and VERIDICALITY: our squarish experiences, with the reference-frame parameter implicitly filled in, can attribute the same geometrical property of squareness that we reason about in Euclidean proof to the objects we perceive, while at the same time attributing to them a property they genuinely have.

5.4. IS SHAPE INTRINSIC?

I now want to consider one final objection, which focuses on my claim that the account just sketched—on which all squarish experiences represent Euclidean squareness, relative to the observer—vindicating COMMONALITY. According to this objection, we can grant that the view I’ve proposed counts as a presentationalist one, in that a single *relation* is represented by all squarish experiences. But, the objection goes, this isn’t enough to save COMMONALITY. What COMMONALITY demands is that the property represented by squarish experiences be the same property as that about which we reason in doing *a priori* Euclidean proof. And this latter property is surely *not* relational: there is no ‘hidden parameter’ involved in the property of squareness

that appears in Hippocrates's proof. Squareness, as it features in *a priori* geometrical reasoning, is *intrinsic*; the property of squareness that I've argued is represented by our squarish experiences is *relational*. So the two simply can't be the *same* property; thus, on my own view, COMMONALITY turns out to be false.⁵⁴

In order to respond to this objection, it will be helpful first to consider another puzzle about the nature of shape properties: the problem of 'temporary intrinsics'.⁵⁵ The problem can be spelled out in terms of David Lewis's classic example. At noon, Lewis is standing up straight; but earlier, he was bent over tying his shoes. Being bent and being straight are two incompatible shape properties; nothing—including Lewis—could be both bent and straight. So we need a way to understand how both of these shape properties can be instantiated by the same object without running into contradiction.

We might propose to do this by noting that Lewis is not *simply* straight and bent; he is, rather, straight-at- t_1 and bent-at- t_2 (where t_1 and t_2 are different times). While being straight and being bent are incompatible, being straight-at- t_1 and bent-at- t_2 are not. They are compatible *relational* properties.

According to Lewis, this solution to the problem of temporary intrinsics cannot work because it implies that,

contrary to what we might think, shapes are not genuine intrinsic properties. They are disguised relations, which an enduring thing may bear to times. One and the same enduring thing may bear the bent-shape relation to some times, and the straight-shape relation to others.... This is simply incredible.... If we know what shape is, we know that it is a property, not a relation. (Lewis 1986, p. 204)

⁵⁴ This objection can allow that there is *some* connection between the monadic *is-square* predicate that features in Euclidean geometry and the dyadic *is-square-in* predicate that relates objects to reference frames in STR; but it denies that the two predicates pick out the *same property*, since one is monadic, and the other dyadic. On a view that allows for predicates of *variable arity* (or 'multigrade predicates'), the fact that Euclidean squareness, as it features in our *a priori* reasoning, is expressed by a monadic predicate, would not rule out that that very same property could be instantiated in a relativistic universe *relationally*. That is, we could take a single *is-square* predicate, Q , to express the very same property of Euclidean squareness in *a priori* geometry (where it appears in monadic form: $Q(o)$) and in attributions of squareness to physical objects, relative to a reference frame, in our universe (where it appears in *dyadic* form: $Q(o,r)$). For an argument that we should allow for such multigrade predicates, see Oliver and Smiley (2004). I will not here take a stand on whether we should allow for multigrade predicates; instead, I hope to show that such predicates are not needed to understand shape perception in STR, since Euclidean squareness is represented in our experiences *non-relationally*.

⁵⁵ See Lewis (1986), Johnston (1987), Lowe (1988), Haslanger (2003), and Sider (2007).

Lewis's claim here is that shape properties are *essentially* intrinsic, while the properties attributed to an object according to the proposal under consideration—straightness-at-t1, bentness-at-t2—are *relational*. So, if we accept this proposal, it turns out that the properties we attribute to objects are not *shape* properties at all.

We may suppose that Lewis would agree that shape properties, as he conceives of them, are the ones we reason about in Euclidean geometry (which are not relational); his claim, then, is that those properties, in their very nature, are intrinsic, and thus no relation could be Euclidean squareness. This means that no view that analyzes the contents of squarish experiences in STR in terms of relational properties could be one on which COMMONALITY turns out to be true. And so, even if the view I've sketched preserves some form of *presentationalism*, it cannot save COMMONALITY.

The key claim in Lewis's objection takes the form of a conditional: 'If we know what shape is, we know that it is a property, not a relation'. I have, in effect, been insisting that the antecedent of this conditional is true: we *do*, contrary to what the placeholder view implies, know what squareness is. If the placeholder view were true, then Lewis's claim would pose no threat to the thesis that squareness is a relation, since the antecedent of the conditional would be false. According to the placeholder view, we *don't* know what shape is: we have only a 'placeholder' understanding of the property we ascribe to objects in taking them to be square. We know squareness as 'that property, whatever it turns out to be, that plays the relevant role in generating our squarish experiences'. And that kind of 'knowledge' of shape in no way rules out that squareness could be a relation; indeed, as we saw above, it does not rule out that squareness could turn out to be a completely unfamiliar property of the Minkowski space-time manifold.⁵⁶ But what I emphasized in the Introduction is that our knowledge of shape is not like this: while it is plausible that we have only a kind of placeholder concept of *color* properties, we know—from the perspective of experience itself—which determinate property *squareness* is, which specific property our squarish experiences attribute to the objects we perceive. As COMMONALITY insists, it is the very property of Euclidean squareness about which we reason in doing *a priori* geometrical proof. In trying to vindicate both VERIDICALITY and COMMONALITY, then, what we need is an account that allows us to say that it is the property of squareness on which we have an *a priori* grasp—the property that shows up *unrelativized* in Euclidean proof—that our perceptual experience veridically attributes to objects.

My claim is that the account I've proposed can achieve this. In order to see how, we can first consider what has been called the 'adverbial' response to the problem of

⁵⁶ The sense in which we don't know *what squareness is*, on the placeholder view, is that we lack *acquaintance* with it. We don't know *which* property squareness is (at least prior to doing a lot of empirical research), if we only know that it is *whatever* property in fact typically generates squarish experiences. See Campbell (2011), Johnston (1996), and Evans (1982) on the significance of this type of acquaintance, or 'knowledge which'; the *locus classicus* is of course Russell (1912).

temporary intrinsics.⁵⁷ Lewis's challenge is that, in order to avoid contradiction in attributing shape properties to an enduring object like a person—in order to avoid the conclusion that a single object is both bent and straight—we would need to introduce some extra *things*, over and above the object and the properties: namely, the *times* at which the object has the properties. This would require us, by Lewis's lights, to insert an extra *argument place* in our attributions of shapes to objects: it is not that an object, *o*, is square (which we could formalize, using *Q* as the predicate 'is-square', as ' $Q(o)$ '); it is that *o* bears the square-at *relation* to a further thing, a time *t* (thus, what we really have is ' $Q(o, t)$ '). The is-square predicate *Q* turns out to be two-place; the is-square property turns out to be the is-square-at relation.

The adverbial response to Lewis's challenge is that we needn't understand the temporal modifier 'at *t*'—in a sentence like 'Lewis is bent at noon'—as introducing an additional *object*, a time *t*, to which the subject bears some relation. Instead, we can see 'at *t*' as expressing the *manner* in which the shape property is *instantiated*; we can, in the words of E.J. Lowe, take "at *t*" at its face value as having adverbial (or predicate modifier) status' (Lowe 1988, p. 73). Temporal modifiers, on this account, have an *adverbial* role: a sentence like 'Lewis is bent at noon' is to be parsed as 'Lewis is-noonly bent'.⁵⁸

I do not want to take a stand on whether the adverbial response is the best solution to the problem of temporary intrinsics. What I do want to suggest is that we can apply a version of the adverbial response to a parallel problem raised by STR. Recall how we described Third Base_L, the rest-square object on Lorentz Earth: relative to Lorentz Earth, it is square; relative to Earth, it is 2:1-rectangular. Being square and being 2:1-rectangular are incompatible properties; so, if we take them to be *intrinsic* properties of Third Base_L, we seem to have a contradiction. We might, then, think that we are forced to acknowledge that the shapes of objects are *not* intrinsic: STR reveals that we need to introduce another *thing*, a reference frame, relative to which objects have shapes. Third Base_L can then be seen as having two compatible relational properties: it bears the square-in *relation* to Lorentz Earth's frame, and the 2:1-rectangular-in relation to Earth's frame. In general, objects bear the square-in relation to frames; they do not instantiate the intrinsic property of *Euclidean* squareness.⁵⁹ This is just another way of drawing out the idea that, given STR, we are forced to reject COMMONALITY.

⁵⁷ See Johnston (1987), Lowe (1988), Haslanger (2003).

⁵⁸ This is a somewhat awkward construction, but there are other temporal modifiers for which the explicitly adverbial rendering is more natural: we can say 'Lewis is-presently bent' and 'Lewis is-formerly straight', giving us two felicitous (and transparently non-contradictory) sentences. Compare Johnston (1987, p. 128): 'Temporal qualification has to do with the ways individuals have properties. Unproblematically Sam may have the property of being fat in the *t** way and have the property of being thin the *t* way. Temporal qualifiers are often adverbs. Sam is presently fat. But he is *thly* thin'.

⁵⁹ Skow (2007) argues, based on considerations entirely independent of STR, that objects do not have intrinsic shape properties. I will not be addressing his argument here.

My response to this challenge should by now be apparent: Reference frames are not additional *things* to which objects bear some relation (they do not fill an additional argument place in uses of the ‘is-square’ predicate). Instead, the role reference frames play in STR is to determine different *ways* in which spatial properties can be *instantiated* by objects in our universe. The properties themselves are not relations between reference frames and objects; a shape property like squareness remains the very same intrinsic property about which we reason in Euclidean proof.

On this picture, Third Base_L instantiates squareness in the *Lorentzian manner*, and it instantiates 2:1-rectangularity in the *Earthly manner* (we might say that it is-Earthwise 2:1-rectangular, and it is-Lorentzwise square). These two properties—2:1-rectangularity and squareness—are the same intrinsic properties (expressed by the same monadic predicates) as those that feature in Euclidean proof.

Turning to the question of shape *perception*, we can see that this picture provides a natural account of the contents of squarish experiences. Perception is, in its nature, first-personal. When a subject *S* has a squarish experience, then, the manner of instantiation of squareness that sets veridicality conditions for her experience is the one determined by *her own* (current) state of motion: *S* represents the object she perceives as instantiating Euclidean squareness *S-wise*.⁶⁰ On this picture, a squarish experience presents a subject with Euclidean squareness—Euclidean squareness is the *presentational* content of such experiences. In a relativistic universe like ours, the contents of such experiences *also* include a further, *non-presentational* element—a specification of the *manner* in which that property is instantiated, determined by the subject’s state of motion.

⁶⁰ On this picture, perceptions are veridical if and only if they represent the perceived object as having the shape it instantiates relative to the perceiver’s current reference frame. We thus get the intuitively correct veridicality judgments about the shape experiences of Astronaut Albert, in the cases discussed in §5.2. When we widen our gaze and consider non-perceptual shape contents, we can also see how this account allows for a natural understanding of the common locutions of STR. Once we are familiar with the theory, we can *make explicit* the manner of instantiation of spatial properties that we wish to discuss. Even while on Earth, we can *say* ‘Third Base_L is square in Lorentz Earth’s frame’; here, we make the claim that Third Base_L instantiates squareness *Lorentzwise*. The flexibility of our language—a flexibility that is absent in the domain of perception, where contents are necessarily first-personal—allows us to flag that we are discussing different manners of instantiation of shape properties, while holding constant the meaning of our shape terms themselves. When, in more mundane circumstances, we describe an object as square without specifying a reference frame, we should interpret our statements as making claims about the shapes objects instantiate in the manner that is most relevant contextually (namely, the manner of instantiation associated with the speaker’s current reference frame; this explains why Astronaut Albert *speaks truly* when he says ‘Third Base_A is square’). This account thus captures the intuitive idea that the meaning of ‘square’ does not shift when we say, in everyday life, ‘Chessboards are square’, and then, in a discussion of STR, ‘Third Base_L is square relative to Lorentz Earth’.

With this account onboard, we can now (finally) put our finger on where Chalmers's argument against shape presentationalism (from §5.2) goes wrong. Chalmers claims (in step (4) of the argument) that Twin Albert's squarish experience of Third Base_L represents the shape property that Third Base_L actually has. Since (by (5)) that property is 2:1-rectangularity, Chalmers concludes (step (7)) that Twin Albert's squarish experience represents 2:1-rectangularity; and so we have a case in which a squarish experience represents something other than squareness (step (9)), thereby providing a counterexample to shape presentationalism. But if objects can, as I've been suggesting, instantiate *multiple* shape properties—even incompatible shape properties—in different *manners*, there is a failure of reference in the claim, in step (4), that Twin Albert's squarish experience of Third Base_L represents *the* shape property that the base actually has. Third Base_L instantiates *both* 2:1-rectangularity (Earthwise) *and* squareness (Lorentzwise); there is no *unique* shape that is *the* shape it has. The veridicality of Twin Albert's experience thus does not imply that his experience represents 2:1-rectangularity; it only implies that his experience represents *some* shape property that Third Base_L has.⁶¹ And, given that perceptual experiences represent the properties that are instantiated in the first-personally-relevant manner, Twin Albert's squarish experience will represent the shape property Third Base_L instantiates Lorentzwise—namely, *squareness*. Twin Albert's squarish experience is therefore not a counterexample to shape presentationalism.

Importantly, I am not claiming that we were getting *nothing* wrong in our conception of shape properties before Einstein. Prior to STR, we did think that Euclidean squareness was instantiated by objects *absolutely*—without any adverbial modification⁶²—as it would be in a Newtonian universe. So we learned that we were getting something wrong when Einstein revealed that the universe is non-Newtonian. And this seems quite correct: Einstein's theory would not have been as revolutionary as it was had it not revealed an error in our ordinary thinking. But what STR reveals is simply that the *manner* of instantiation of shape properties is different from what we took it to be; the properties themselves, instantiated by objects in this novel way, are the very Euclidean properties with which we were already acquainted. On this account, we can understand why we describe STR in the terms we do, using ordinary

⁶¹ Strictly speaking, Chalmers need not claim that Third Base_L instantiates only one shape property; what he needs for his argument to go through is the claim that, whatever shape properties Third Base_L *does* instantiate, the property represented by *Albert's* squarish experience—squareness—is not among them. Modifying the argument to include only this less demanding premise, though, does not remove the problem. There is now a crucial ambiguity in the modified premise 'Third Base_L does not instantiate squareness'. It is true that Third Base_L does not instantiate squareness *Earthwise*; but false that it does not instantiate squareness *Lorentzwise*. In moving from the true premise 'Third Base_L does not instantiate squareness [Earthwise]', via the veridicality premise, to the conclusion, 'Twin Albert's squarish experience does not represent squareness', the argument would thus be equivocating (since it is the *Lorentzwise* instantiation of squareness that is relevant to the truth of the conclusion).

⁶² Or, perhaps, in the 'absolute manner', as the adverbial phrasing suggests.

Euclidean shape terms (in statements like ‘a square object moving relative to us will, relative to our reference frame, be an elongated rectangle’). We do not have to convict our perceptual experience of radical error (allowing us to hold on to VERIDICALITY), and we can (as COMMONALITY urges) understand the contents of our shape experiences in terms of the very unrelativized, Euclidean shape properties on which we have an intuitive, *a priori* grasp.

5.5. CONCLUSION

According to Thomas Reid, there is a genuine distinction between primary and secondary qualities, in that:

our senses give us a direct and a distinct notion of the primary qualities, and inform us what they are in themselves: But of the secondary qualities, our senses give us only a relative and obscure notion. They inform us only, that they are qualities that affect us in a certain manner, that is, produce in us a certain sensation; but as to what they are in themselves, our senses leave us in the dark. (Reid 1785, 2002; Essay II, Chapter XVII, Section I, p. 201)

On this Early Modern picture, experience provides us with only a placeholder conception of the *secondary* qualities; but, because perception includes ‘a direct and a distinct’ representation of the *primary* qualities, we are not left *entirely* ‘in the dark’ about the nature of the empirical world.

An intuitively appealing way of spelling out Reid’s idea is the following: experience provides us with direct and distinct representations of primary qualities because it presents objects as instantiating the Euclidean spatial properties of which we have an *a priori* understanding.⁶³ The Einsteinian revolution, however—by calling into question the applicability of our *a priori* spatial concepts to the physical universe—threatened to undermine the traditional distinction between primary and secondary qualities: if the objects we perceive do not in fact instantiate Euclidean spatial properties, we might feel compelled—in order to preserve VERIDICALITY—to retreat to a placeholder account of *every* aspect of perceptual content, and to accept that our experience does leave us utterly in the dark about the nature of the world.

Fortunately, though, Einstein’s theories do not force us to accept a placeholder account of spatial perception. The objects we perceive do, even within STR, instantiate the very shape properties of which Reid had ‘a direct and a distinct notion’. What STR reveals is that these Euclidean shape properties are instantiated in our universe

⁶³ Reid himself would have been pleased to acknowledge some *a priori* element at work in our perception of the primary qualities; but he would not have accepted that it is *Euclidean* geometry that is represented in *visual* experience. Indeed, Reid is credited by many with the invention of *non-Euclidean* geometry, precisely as a theory of the ‘geometry of visibles’. Reid did hold, however, that *tactual* geometry is Euclidean (see Van Cleve (2002)).

not *absolutely*, but in a *manner* that Reid could not have foreseen—namely, *relative to particular frames of reference*. Understanding STR in this way allows us to hold onto the crucial Early Modern idea that, in virtue of our spatial experience, we are indeed acquainted with a central aspect of the world we perceive.

CONCLUSION

At the heart of Kant's critical philosophy lies the following question: How can we have knowledge of empirical space in virtue of our *a priori* geometrical reasoning? It was answering this question, in large part, that led Kant to the radical doctrine of transcendental idealism; but, from a contemporary perspective, Kant's "how possible" question itself has come to seem misguided. Having discovered that there are consistent non-Euclidean geometries, and having accepted that our own universe is not a perfectly Euclidean space, we now simply reject Kant's epistemological presupposition – namely, that we have genuine *a priori* knowledge of the spatial character of empirical reality.

The contemporary rejection of Kant's account of empirical knowledge is well founded: after Einstein, we can no longer take seriously the idea that armchair reasoning guarantees us insight into the nature of physical space. But the question at the core of Kant's project remains unresolved. For Kant's great insight was that there is a crucial *cognitive* connection between the theorems we prove in Euclidean geometry and the spatial features we perceive physical objects to have. Having proven the Pythagorean theorem, for instance, a carpenter will expect a particular relation to hold among the lengths of the sides of a right triangle she is constructing from wooden beams. This shows that we *represent* the empirical objects we perceive as conforming to the results we obtain from Euclidean proof, even if we no longer think that such representations are guaranteed to be correct.

Transposed from the epistemological key to a representational one, Kant's "how possible" question about the connection between *a priori* reasoning and perceptual experience is well worth asking: "How must we conceive of perceptual experience and *a priori* thought, in order to make sense of the fact that we *represent* the empirical world as instantiating the very spatial properties about which we reason in performing *a priori* Euclidean proof?"

This "transposed" Kantian question has provided the framework for my discussion. In the first chapter, I showed how attention to the connections between our *a priori* and experiential representations of space can solve the PROBLEM OF INTERNAL ACQUAINTANCE. Perceptual experience seems to provide us with a point of view on a world that is independent of our awareness – a world of material objects arranged in space, with various colors, shapes, and other sensible qualities. A pervasive narrative in contemporary philosophy holds that this observation about the *world-directedness* of experience is incompatible with another, equally-natural thought: according to the default modern picture of the mind, our conscious lives, including our perceptual experiences, are constituted solely by our own internal states. On this *internalist* conception of experience, the thought goes, the world itself lies forever beyond our ken, leaving us without any grip on mind-independent reality. Just as the intrinsic properties of linguistic representations—the loops and lines of the typographic string BOOT, for example—do

not *themselves* represent any particular entity, experiences, in virtue of their internally-fixed phenomenology, could not *themselves* establish any connection to external-world features. The natures of the shapes and colors of the external world could, therefore, never be known through experience as conceived on the internalist model; those properties could show up only as the unknown external causes of a subject's self-contained inner life.

I argued that this challenge to internalism fails because it construes the resources available to the internalist too narrowly. We are not forced to explain our grip on the features of the world in terms of their effects on our sensibility. Rather, I argue, we grasp the nature of some of those features—the *primary* qualities—*directly*, in virtue of our possession of *a priori* spatial concepts. The property represented by an experience of an object as square, for example, is the very property that features in abstract geometrical proof – when, say, we utilize our understanding of squareness to prove the Pythagorean theorem. Thus, by attending to the connection between Euclidean proof and experience, we can see how our internally-constituted experiences of space can represent a mind-independent world.

Having thus set out the basic thesis of the dissertation, I was left with two central tasks, one concerning my claims about Euclidean proof, the other concerning my picture of spatial experience. First, in Chapter 2, I argued that the concepts employed in Euclidean proof are genuinely *a priori*. Then, in Chapter 3, I argued that these very concepts—the *a priori* geometrical concepts of Euclidean proof—also feature in the contents of our perceptual experiences. Such a link between the contents of *a priori* Euclidean geometrical reasoning and those of spatial experience was needed, I showed, in order to explain cases, like that of the carpenter, in which we immediately apply the results of Euclidean proof to the objects we perceive. We see objects as instantiating the properties—such as right triangularity—that we reason about in abstract geometry. This connection, I suggested, arises because both our experience of space and our more intellectual geometrical reasoning employ the same set of spatial *proto-concepts* – they are both expressions of our innate grasp of space.

In the final two chapters, I defended this picture of spatial experience against two kinds of challenges. First, in Chapter 4, I showed that no satisfactory account of spatial experience is possible if we limit ourselves to the resources provide by our empirical interactions with the world. For *color* experience—where, lacking any relevant *a priori* concepts, we really are limited to empirical resources—fails to have any determinate content. Second, in Chapter 5, I argued that linking the contents of our spatial experience to our *a priori* Euclidean reasoning does not, even in the face of Einstein's discoveries about the surprising nature of space-time, force us to the implausible conclusion that we suffer an illusion in perceiving the space around us.

Through careful analysis of spatial cognition across multiple domains—sensory experience of the shapes of physical objects; *a priori* mathematical reasoning; and contemporary scientific theories of space-time—I hope to have shed some light on the nature of one central aspect of our cognitive lives. I also hope to have provided a general model of how attention to the complex interconnections of the human mind can help us make progress on some of our most intractable philosophical puzzles.

BIBLIOGRAPHY

- Bennett, Jonathan (1971). *Locke, Berkeley, Hume: Central Themes*. Oxford University Press.
- Block, Ned (1990). "Inverted Earth." *Philosophical Perspectives*, 4:53-79.
- Brewer, Bill (2006). "Perception and Content." *European Journal of Philosophy* 14 (2):165-181.
- Burge, Tyler (1979). "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1):73-122.
- Burge, Tyler (1986). "Individualism and Psychology." *Philosophical Review* 95 (January):3-45.
- Burge, Tyler (2006). "Perceptual Objectivity." *Philosophical Review* 118 (3):285-324.
- (1988). "Individualism and Self-Knowledge." *Journal of Philosophy* 85:649-663.
- Campbell, John (2002). "Berkeley's Puzzle." In Tamar S. Gendler & John Hawthorne (eds.), *Conceivability and Possibility*. MIT Press.
- (2011). "Consciousness and Reference." in B. McLaughlin, A. Beckermann & S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford University Press.
- Campbell, John and Cassam, Quassim (2014). *Berkeley's Puzzle: What Does Experience Teach Us?* Oxford University Press.
- Chalmers, David J. (2005). "The Matrix as Metaphysics." In Christopher Grau (ed.), *Philosophers Explore the Matrix*. Oxford University Press. pp. 132.
- (2006). "Perception and the Fall from Eden." In Tamar S. Gendler & John Hawthorne (eds.), *Perceptual Experience*. Oxford University Press. pp. 49--125.
- (2010). "The Matrix as Metaphysics." In D.J. Chalmers, *The Character of Consciousness*. Oxford University Press.
- (2012). *Constructing the World*. Oxford University Press.
- (forthcoming). "Three Puzzles About Spatial Experience." In A. Pautz & D. Stoljar (eds.), *Themes from Ned Block*. Oxford University Press. (<http://consc.net/papers/spatial.pdf>)
- Davies, Martin (1993). "Aims and Claims of Externalist Arguments." *Philosophical Issues* 4:227-249.
- Dedekind, Richard (1901). *Essays on the Theory of Numbers*, W.W. Beman (trans.). Open Court Publishing Co.
- Descartes, René (1641/1996). *Meditations on First Philosophy*, J. Cottingham (ed.). Cambridge University Press.
- DiSalle, Robert (2006). *Understanding Space-Time: The Philosophical Development of Physics From Newton to Einstein*. Cambridge University Press.
- Dretske, Fred (1995). *Naturalizing the Mind*. MIT Press.
- Egan, Andy (2010). "Projectivism Without Error." In Bence Nanay (ed.), *Perceiving the World*. Oxford University Press. pp. 68-80.

- Einstein, Albert (1911). "The Theory of Relativity." In Martin J. Klein, A. J. Kox, Jürgen Renn & Robert Schulmann (eds.), *The Collected Papers of Albert Einstein, Volume 3: The Swiss Years: Writings 1909-1911*. Princeton University Press, pp. 340-350.
- (1922). *Sidelights on Relativity*, G.B. Jeffery and W. Perrett (trans). (http://www.ibiblio.org/ebooks/Einstein/Sidelights/Einstein_Sidelights.pdf)
- Evans, Gareth (1980). "Things Without the Mind." In Zak van Straaten (ed.), *Philosophical Subjects: Essays Presented to P.F. Strawson*. Clarendon Press, 76-116.
- (1982). *The Varieties of Reference*. Oxford University Press.
- Foster, John A. (2000). *The Nature of Perception*. Oxford University Press.
- Friedman, Michael (1983). *Foundations of Space-Time Theories Relativistic Physics and Philosophy of Science*. Princeton University Press.
- (1992). *Kant and the Exact Sciences*. Harvard University Press.
- Giaquinto, Marcus (2011). "Crossing Curves: A Limit to the Use of Diagrams in Proofs." *Philosophia Mathematica* 3(19):281–307.
- Gibbard, Allan F. (1996). "Visible Properties of Human Interest Only." *Philosophical Issues* 7:199-208.
- Graff, Delia (2001). "Phenomenal Continua and the Sorites." *Mind* 110:905-935.
- Hardin, Clyde L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Hackett.
- Haslanger, Sally (2003). "Persistence Through Time." In Michael J. Loux & Dean W. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*. Oxford University Press, pp. 315--354.
- Hilbert, David (1989). "Letter to Frege of December 29, 1899." Quoted from Blanchette, Patricia, "The Frege-Hilbert Controversy." In *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.). (<http://plato.stanford.edu/archives/spr2014/entries/frege-hilbert/>)
- Hopkins, James (1973). "Visual Geometries." *Philosophical Review* 82(1):3-34.
- Horgan, Terence and Tienson, John (2002). "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press, 520-533.
- Horwich, Paul (1978). "On the Existence of Time, Space and Space-time." *Noûs* 12 (4):397-419.
- Izard, Véronique, Pierre Pica, Danièle Hinchley, Stanislas Dehaene & Elizabeth Spelke (2011). "Geometry as a Universal Mental Construction." In Stanislas Dehaene & Elizabeth Brannon (eds.), *Space, Time and Number in the Brain*. Oxford University Press.
- Hurley, Susan L. (1998). *Consciousness in Action*. Harvard University Press.
- Jacovides, Michael (1999). "Locke's Resemblance Theses." *Philosophical Review* 108 (4):461-496.
- Johnston, Mark (1987). "Is There a Problem About Persistence?" *Proceedings of the Aristotelian Society* 61:107-135.
- (1996a). "Is the External World Invisible?" *Philosophical Issues*, 7:185-198.

- (1996b). “A Mind-Body Problem at the Surface of Objects.” *Philosophical Issues*, 7:219-229.
- Langton, Rae (1998). *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford University Press.
- Lewis, David K. (1986). *On the Plurality of Worlds*. Blackwell.
- (2009). “Ramseyan Humility.” In David Braddon-Mitchell and Robert Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*. MIT Press, 203-222.
- Loar, Brian (2003). Phenomenal Intentionality as the Basis of Mental Content. In Martin Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. MIT Press, 229-258.
- Lowe, E.J. (1988). “The Problems of Intrinsic Change: Rejoinder to Lewis,” *Analysis* 48.
- Lycan, William G. (2001). “The Case for Phenomenal Externalism.” *Philosophical Perspectives* 15 (s15):17-35.
- MacFarlane, John (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press.
- Mackie, J. L. (1976). *Problems from Locke*. Clarendon Press.
- Manders, Kenneth (1995). “Diagram-Based Geometric Practice.” In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*. Oxford University Press, 65-79.
- (2008). “The Euclidean Diagram.” In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*. Oxford University Press, 80-133.
- Martin, Michael G. F. (2004). “The Limits of Self-Awareness.” *Philosophical Studies* 120 (1-3):37-89.
- Maudlin, Tim (2012). *Philosophy of Physics: Space and Time*. Princeton University Press.
- (2014). *New Foundations for Physical Geometry: The Theory of Linear Structures*. Oxford University Press.
- McDowell, John (1985). “Values and Secondary Qualities.” In Ted Honderich (ed.), *Morality and Objectivity*. Routledge. pp. 110-129.
- (1986). “Singular Thought and the Extent of ‘Inner Space’.” In John McDowell and Philip Pettit (eds.), *Subject, Thought, and Context*. Clarendon Press.
- (2011). “Colors as Secondary Qualities.” In Jason Bridges, Niko Kolodny, and Wai-hung Wong (eds.), *The Possibility of Philosophical Understanding: Reflections on the Thought of Barry Stroud*. Oxford University Press.
- McGinn, Colin (1989). *Mental Content*. Blackwell.
- McKinsey, Michael (1991). “Anti-Individualism and Privileged Access.” *Analysis* 51(1): 9-16.
- McLaughlin, Brian P. and Tye, Michael (1998). “Is Content-Externalism Compatible with Privileged Access?” *Philosophical Review*, 107 (3): 349-380.
- Nagel, Thomas (1986). *The View from Nowhere*. Oxford University Press.
- Oliver, Alex & Smiley, Timothy (2004). “Multigrade Predicates.” *Mind* 113 (452):609-681.

- Ney, Alyssa (2012). "The Status of our Ordinary Three Dimensions in a Quantum Universe." *Noûs* 46 (3):525-560.
- Papineau, David (2014). "The Presidential Address: Sensory Experience and Representational Properties." *Proceedings of the Aristotelian Society* 114 (1pt1):1-33.
- Pautz, Adam (2006). "Can the Physicalist Explain Colour Structure in Terms of Colour Experience?" *Australasian Journal of Philosophy* 84 (4): 535-564.
- Peacocke, Christopher (1989). *Transcendental Arguments in the Theory of Content an Inaugural Lecture Delivered Before the University of Oxford on 16 May 1989*.
- Peacocke, Christopher (2015). "Magnitudes: Metaphysics, Explanation, and Perception." In Annalisa Coliva, Volker Munz & Danièle Moyal-Sharrock (eds.), *Mind, Language and Action: Proceedings of the 36th International Wittgenstein Symposium*. De Gruyter. pp. 357-388.
- Putnam, Hilary (1973). "Meaning and Reference." *Journal of Philosophy* 70(19): 699-711.
- (1981). "Brains in a Vat," Ch. 1 of *Reason, Truth and History*. Cambridge University Press.
- Reichenbach, Hans (1958). *The Philosophy of Space and Time*. Dover Publications.
- Reid, Thomas (1785/2002). *Essays on the Intellectual Powers of Man*, ed. Derek R. Brookes. Edinburgh University Press.
- Russell, Bertrand (1912). *The Problems of Philosophy*. Oxford University Press.
- Sattig, Thomas (2015). *The Double Lives of Objects: An Essay in the Metaphysics of the Ordinary World*. Oxford University Press.
- Searle, John R. (1983). *Intentionality: an Essay in the Philosophy of Mind*. Cambridge University Press.
- Sider, Theodore (2007). "Temporal Parts." In Theodore Sider, John Hawthorne & Dean W. Zimmerman (eds.), *Contemporary Debates in Metaphysics*. Blackwell. pp. 241--262.
- Siegel, Susanna (2010). *The Contents of Visual Experience*. Oxford University Press USA.
- Skow, Bradford (2007). "Are Shapes Intrinsic?" *Philosophical Studies* 133 (1):111 - 130.
- Sosa, David (1996). "Getting Acquainted with Perception." *Philosophical Issues*, 7: 209-214.
- Spelke, Elizabeth, Sang Ah Lee & Véronique Izard (2010). "Beyond Core Knowledge: Natural Geometry." *Cognitive Science* 34 (5):863-884.
- Spelke, Elizabeth S. (2011). "Natural Number and Natural Geometry." In Stanislas Dehaene & Elizabeth Brannon (eds.), *Space, Time and Number in the Brain*. Oxford University Press. pp. 287--317.
- Stazicker, James (2011). "Attention, Visual Consciousness and Indeterminacy." *Mind and Language* 26 (2):156-184.
- (forthcoming). "The Visual Presence of Determinable Properties." In F. Dorsch, F. Macpherson & M. Nida-Rumelin (eds.), *Phenomenal Presence*. Oxford University Press.

- Strawson, P. F. (1966). *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. Methuen.
- Stroud, Barry (2002). *The Quest for Reality: Subjectivism and the Metaphysics of Colour*. Oxford University Press.
- (2004). "Replies to Comments on *The Quest for Reality: Subjectivism and the Metaphysics of Colour*." *Philosophy and Phenomenological Research* 68 (2):423-442.
- Tarski, Alfred (1959), "What is Elementary Geometry?" In Leon Henkin, Patrick Suppes & Alfred Tarski (eds.), *The axiomatic method. With special reference to geometry and physics. Proceedings of an International Symposium held at the Univ. of Calif., Berkeley, Dec. 26, 1957-Jan. 4, 1958*. Studies in Logic and the Foundations of Mathematics, 16–29.
- Thompson, Brad J. (2010). "The Spatial Content of Experience." *Philosophy and Phenomenological Research* 81 (1):146-184.
- Travis, Charles S. (2004). "The Silence of the Senses." *Mind* 113 (449):57-94.
- Tye, Michael (2000). *Consciousness, Color, and Content*. MIT Press.
- Van Cleve, James (2002). "Thomas Reid's Geometry of Visibles." *Philosophical Review* 111 (3):373-416.
- Child, T.W. (1994). *Causality, Interpretation and the Mind*. Oxford University Press.
- Winnie, John A. (1977). "The Causal Theory of Space-Time." In John Earman, Clark Glymour & John Stachel (eds.), *Foundations of Space-Time Theories*. University of Minnesota Press.
- Woozley, A. D. (1964). "Introduction." In John Locke, *An Essay Concerning Human Understanding* (A. D. Woozley, ed.). Collins Publishing.
- Wright, Crispin (1986). *Realism, Meaning, and Truth*. Blackwell Publishing.