University of Massachusetts - Amherst
ScholarWorks@UMass Amherst

Dissertations | Dissertations and Theses

9-1-2009

Phenomenal Acquaintance

Kelly Trogdon
University of Massachusetts - Amherst, ktrogdon@philos.umass.edu

Follow this and additional works at: http://scholarworks.umass.edu/open_access_dissertations

Recommended Citation
Trogdon, Kelly, "Phenomenal Acquaintance" (2009). *Dissertations.* Paper 128.

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

PHENOMENAL ACQUAINTANCE

A Dissertation Presented

By

KELLY TROGDON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2009
Philosophy

© Copyright by Kelly Trogdon 2009
All Rights Reserved

PHENOMENAL ACQUAINTANCE

A Dissertation Presented

By

Kelly Trogdon

Approved as to style and content by:

_____
Joseph Levine, Chair

_____
Louise Antony, Member

_____
Lynne Rudder Baker, Member

_____
Kyle Cave, Member

_____
Phillip Bricker, Department Head
Philosophy

DEDICATION

To Ellen Woodall and my family for their love and support.

ACKNOWLEDGMENTS

I would like to thank first and foremost my advisor, Joseph Levine, for his guidance, support, and friendship. When I first approached him about advising my dissertation, he warned me that he might not be a good advisor. He was wrong. Second, I wish to extend my gratitude to the other members of my committee, Lynne Rudder Baker, Kyle Cave, and especially Louise Antony for their helpful comments and suggestions. Third, thanks are due to Sam Cowling, Timmy Fuller, Barak Krakauer, Casey O'Callaghan, and Alex Sarch for help with specific sections of the dissertation. Fourth, I am grateful to Consuelo Preti, Gary Rosenkrantz, and especially Gene Witmer for encouragement early on in my philosophical career and for sparking my interest in the philosophy of mind. Fifth, thanks to the Graduate School at UMass Amherst for awarding me a Graduate Fellowship for the fall of 2008 and spring of 2009 in order to write this dissertation. Finally, I want to thank my preferred place of work – the Haymarket in Northampton, MA – for letting me sit there for hours upon end, drinking coffee and working away.

ABSTRACT

PHENOMENAL ACQUAINTANCE

SEPTEMBER 2009

KELLY TROGDON, B.A., UNIVERSITY OF NORTH CAROLINA, GREENSBORO

M.A., UNIVERSITY OF FLORIDA

PH.D., UNIVERSITY OF MASSACHUSETTS, AMHERST

DIRECTED BY: PROFESSOR JOSEPH LEVINE

Chapter 1 of *Phenomenal Acquaintance* is devoted to taking care of some preliminary issues. I begin by distinguishing those states of awareness in virtue of which we're acquainted with the phenomenal characters of our experiences from those states of awareness some claim are at the very nature of experience. Then I reconcile the idea that experience is transparent with the claim that we can be acquainted with phenomenal character.

In Chapter 2 I set up a dilemma that is the primary focus of the dissertation. In the first part of this chapter I argue that phenomenal acquaintance has three key features, what I call its 'directness', 'thickness', and 'infallibility'. In the second part I argue, however, that it's really quite puzzling how thoughts about phenomenal character (or any thoughts, for that matter) could have them.

In the next two chapters I consider how we might resolve the dilemma described above. I begin in Chapter 3 by considering an account of phenomenal acquaintance inspired by Bertrand Russell's discussion of acquaintance. The general idea here is to excise mental representation from phenomenal acquaintance, and I ultimately reject the proposal.

Chapter 4 is the core chapter of *Phenomenal Acquaintance*. In it I propose an account of phenomenal acquaintance that doesn't excise mental representation. My account is comprised of three theses. First, token experiences are complex and have instances of phenomenal properties as components. Second, instances of phenomenal properties are mental representations, and they represent themselves. Third, the attention relevant to phenomenal acquaintance is underwritten by self-representation. I argue that my account explains how phenomenal acquaintance is direct, thick, and infallible, thereby resolving our dilemma.

I argue in Chapter 5 that my account of phenomenal acquaintance explains why there is an explanatory gap between the phenomenal and non-phenomenal truths. Accordingly, I conclude that the explanatory gap doesn't pose a problem for physicalism. Here I implement what has come to be called the 'phenomenal concept strategy' for responding to the challenge posed by the explanatory gap.

CONTENTS

Page

ACKNOWLEDGEMENTS v

ABSTRACT vi

LIST OF FIGURES ix

CHAPTER

1. GETTING ACQUAINTED 1

1.1. Introduction 1
1.2. Awareness and Experience 7
1.3. Experiential Transparency 15
1.4. Conclusion 27

2. WHY PHENOMENAL ACQUAINTANCE IS PUZZLING 29

2.1. Introduction 29
2.2. Directness, Thickness, and Infallibility 30
2.3. Three Puzzles 34
2.4. Conclusion 40

3. RUSSELL ON ACQUAINTANCE 43

3.1. Introduction 43
3.2. Two Features 44
3.3. The Objects of Acquaintance 49
3.4. Evaluating Russellian Acquaintance 56
3.5. Conclusion 65

4. AN ACCOUNT OF PHENOMENAL ACQUAINTANCE 67

4.1. Introduction 67
4.2. The Account 67
4.3. Resolving Our Puzzles 73
4.4. Objections 76
4.5. Conclusion 94

5. PHENOMENAL ACQUAINTANCE AND PHYSICALISM 95

5.1. Introduction 95
5.2. The Explanatory Gap and Physicalism 95
5.3. The Phenomenal Concept Strategy 101
5.4. An Objection 117

5.5. Conclusion 125

BIBLIOGRAPHY 126

LIST OF FIGURES

Figure Page

1. A First Pass 37

2. A Second Pass 37

3. Phenomenal Acquaintance 74

4. Balogian Phenomenal Acquaintance 115

CHAPTER 1

GETTING ACQUAINTED

1.1. Introduction

Some mental states are experiential or phenomenally conscious while others are not. A mental state is an experiential state just in case there is "something it's like" to have it. There are various parts or aspects of what an experience is like for its subject. Consider, for example, your current visual experience as of black letters on a white page. There is a blackish aspect to this experience, as well as a whitish aspect. Call each such aspect *a* phenomenal character. *The* phenomenal character associated with an experiential state, then, is the what-it's-likeness of that state in its totality.

What more can we say about phenomenal character? Following Hellie (2007), suppose someone asks you what a certain experience was like. Your answer will have the form "It was *F*", where *F* is a phenomenal character of your experience. '*F*' is a predicate and, we'll presume, expresses some property, a phenomenal property, e.g. *being square-like*, *being reddish*, etc. Hence, *prima facie*, a phenomenal character of an experience is a property, a phenomenal property, and this property is instantiated by that very experience. I wish to stay neutral, however, on whether a phenomenal character is a monadic property of an experience or an *n*-adic relation partially instantiated by an experience. In either case, we can still speak of an experience as "having" a phenomenal character, or a phenomenal character "of" an experience.[1]

[1] I am ruling out the claim that a phenomenal character is a monadic property or an *n*-adic relation that isn't (even partially) instantiated by experiences. In so doing, I rule out, for example, the view that a phenomenal character is a monadic property of an ordinary physical object represented by an experience.

I also wish to stay neutral on whether there are phenomenal characters corresponding to non-sensory terms, terms in addition to ones for color, shape, spatial position, and so on. Here the issue is whether there are phenomenal characters corresponding to, for example, appearing to be spoon, a cow, or Humphrey Bogart. Following Siewart (2007), consider any case in which we're inclined to distinguish between something's looking as if it's *F* and looking as if it's *G*, where *F* and *G* aren't sensory terms. The question is whether we can always in principle distinguish via introspection the difference in content of the two visual experiences involved (if any) without speaking of *F* and *G* but only the details of the relevant sensory features. If we can, then it would seem that there are no non-sensory phenomenal characters; while, if we can't, it would seem that there are.

Contemporary philosophers of mind are interested in giving an account of phenomenal character compatible with physicalism. What might such an account look like? Here's a recipe for a particular physicalist account. The general idea behind the proposal is that phenomenal character can be understood in terms of intentionality, and intentionality itself can be accounted for within a physicalist framework. Step 1: argue that representationalism – the thesis that the phenomenal character of an experiential state is exhausted by its intentional properties (cf. Tye (1995, 2000)) – is correct. Step 2: establish the representational theory of mind according to which intentional mental states are instantiated in virtue of token mental representations, and the intentional properties of the former are instantiated in virtue of the intentional properties of the latter (cf. Field (1978) and Fodor (1987)). Step 3: argue that naturalism about mental representation is true, roughly the thesis

that the intentional properties of token mental representations are instantiated ultimately in virtue of non-intentional physical properties (cf. Fodor (1994)).[2]

If you put all of these ingredients together, you have a physicalist account of phenomenal character: in the actual world, the phenomenal character of an experiential state is exhausted by its intentional properties; these properties are instantiated in virtue of the intentional properties of various token mental representations; and these properties in turn are instantiated ultimately in virtue of non-intentional physical properties.[3] Suppose, for example, that you're having a perceptual experience as of a red tomato on the table. On this view, the phenomenal character of your experience is understood simply as your visual system representing (something to the effect) that there is a red figure of a certain shape and texture in your visual field. We can conceive of the phenomenal character of your experience either as a monadic property of it instantiated in virtue of its intentional properties, or as a dyadic relation whose relata are your experience and whatever it represents. The intentional properties of your experience, so the idea goes, in turn are instantiated ultimately in virtue of non-intentional physical properties.

---

[2] Tye (1995, 2000) and Rey (1998) each endorse the representational theory of mind but the former is an *externalist* representationalist while the latter identifies himself as an *internalist* representationalist. So Tye claims that the phenomenal properties of an experiential state are instantiated in virtue of the widely individuated intentional properties of mental representations, while Rey claims that they're instantiated in virtue of the narrowly individuated intentional properties of mental representations. Representationalism (either externalist or internalist) is just an example of one way of working out the idea that the phenomenal can be understood in terms of the intentional. Other theses that proceed upon the same idea include the higher-order monitoring (cf. Lycan (2004) and Rosenthal (2004)) and same-order monitoring (cf. Smith (1989) and Kriegel (2006)) accounts of experience. We will return to these views later in the chapter.

[3] Representationalism as I have characterized it isn't a physicalist thesis *sans* the other theses I describe above. You could, for example, claim that the phenomenal character of a mental state is exhausted by its intentional properties but deny naturalism about mental representation. This view is compatible with dualism.

You might, of course, object to one or more of the theses that goes into this physicalist account of phenomenal character. You might argue, for example, that it's unclear that mental representation has been successfully naturalized. Causal/information accounts (including versions of the asymmetric dependence thesis and teleosemantics) and conceptual role accounts of mental representation, for example, seem to face significant problems. But let's grant for the sake of argument that mental representation can be naturalized. Let's also grant for the moment the representational theory of mind and content externalism, the view that content is widely individuated. Even granting these assumptions, there are interesting, forceful arguments to the effect that, if a token mental state *m* has both intentional and phenomenal properties, the instantiation of the former by *m* isn't sufficient for the instantiation of the latter by *m* across the space of metaphysically possible worlds (cf. Block (1996, 2003); Levine (2003); and Loar (2003)). I won't review these familiar arguments here. Suffice it to say that that if they succeed (externalist) representationalism is false.

Perhaps as a consequence of these objections, along with the perceived failure of other physicalist accounts of phenomenal character, some physicalists have changed course in the following sense. Instead of trying to understand the nature of phenomenal character itself, whether in terms of intentionality or some other notion, the focus has shifted to trying to better understand how we *think* about the phenomenal characters of our current experiences. Many philosophers – physicalists and dualists alike – have remarked that there is something special about phenomenal thought. Physicalists of late are wont to parlay this consideration into a defense of physicalism. The move here consists of something like the following three claims. First, there are special features of phenomenal thought that explain at least in part why we find dualism intuitively plausible. In particular, the features explain those intuitions at work in the conceivability argument (cf. Chalmers (1996) and (forthcoming)),

the knowledge argument (cf. Jackson (1982, 1986)), the explanatory gap (cf. Levine (1983, 2001, 2007)), and so on. Second, we can give an account of these features in a way that is compatible with physicalism. The third claim is a consequence of the first two, but is important enough to state on its own: once these features have been naturalized, we will be in a position to see that our dualist intuitions can be accommodated within a physicalist framework. The latter claim, along with the claim that we have independent reasons to believe that physicalism is true (think causal exclusion arguments), is supposed to clinch the case for physicalism. This, or something close to it, anyway, has come to be known in the literature as the 'phenomenal concept strategy' (cf. Stoljar (2005)).

The deployment of the phenomenal concept strategy typically consists of some interesting speculation about what sort of metaphysical and epistemological features thoughts about phenomenal character would have to have for our dualist intuitions to be innocuous to physicalism. But this, I submit, may be wrongheaded as a piece of methodology for theorizing about phenomenal thought. Now, it's certainly legitimate to require that whatever we say on the topic of phenomenal thought should be compatible with physicalism. But I don't think that our central guiding principle in theorizing here should be specifically that of ending up with an account of phenomenal thought that is able to explain away the dualist intuitions described above. For the issue of phenomenal thought is itself quite interesting and worth thinking about, independently of what implications an account of it might have for other issues that we, *qua* physicalists, are interested in. So the following strikes me as a good way to proceed. First, by considering our (more or less) pre-theoretical conception of phenomenal thought, construct desiderata for an account of it. Second, provide an account of phenomenal thought that meets these conditions. Third, see if such

an account helps us respond to familiar objections to physicalism. In this dissertation I attempt to do all three things.

Implicit in the phenomenal concept strategy is the idea that a phenomenal thought is more tractable than its object, a particular phenomenal character of an experience. The idea that we can make short work of producing an account of phenomenal thought comports with the general impression among physicalists I alluded to earlier that thought in general can be accommodated within a physicalist framework in a straightforward manner. One thing I want to show, however, is that constructing an account of phenomenal thought is more difficult than many seem to think. There is, I claim, a special set of problems we encounter in trying to give an account of it.

The goal of this introductory chapter is twofold. First, I make some general remarks about the notions of attention or focal awareness and peripheral awareness *vis-à-vis* phenomenal thought. Here I point out that the sort of phenomenal thought I'm interested in – what I call 'phenomenal acquaintance' – is underwritten by focal awareness of or attention to experience. On my view, phenomenal acquaintance is a matter of attending to our experiences and grasping particular phenomenal characters of them. Second, I turn to the idea that experience is transparent or diaphanous. You might reject the very idea that we can attend to our experiences, citing the intuition that experience is transparent. I respond to this concern, arguing that, though there are various senses in which experience is transparent, we can attend to our experiences and thereby grasp particular phenomenal characters.

A note about my approach to phenomenal acquaintance before we proceed. When you hear the term 'acquaintance', I bet you think about epistemology, in particular internalism or foundationalism or both (cf. Sellars (1997), Fumerton (1995), and BonJour and Sosa (2003)). I'm sure you've noticed, however, that my discussion of phenomenal

thought so far has been primarily about its metaphysical rather than its epistemological status. Though in the course of things I touch on epistemological issues relevant to phenomenal acquaintance, the rest of the chapter and the dissertation as a whole proceeds in the same manner. Following Horgan and Kriegel (2007), my general impression is that it's the metaphysics of our cognitive relation to phenomenal character that will explain the relation's epistemological features rather than the other way around. So, while it isn't a good idea to give an "account" of phenomenal thought by imputing to it whatever features might explain away our dualist intuitions, so too is it not a good idea, as far as I can tell, to approach phenomenal thought first from the angle of epistemology rather than that of metaphysics.

1.2. Awareness and Experience

Many think that to have an experience is simply to instantiate a certain sort of state of awareness. Though many are sympathetic with this general proposal, there is much disagreement about the content of experience-making states of awareness. Representationalists claim these states typically represent physical objects (for example, a part of your body or your shoe) as instantiating various properties (cf. Tye (1995, 2000)), while higher-order and same-order monitoring theorists claim that they instead represent mental states, the very states whose experiential status is the object of explanation (cf. Rosenthal (2004) and Kriegel (2006), respectively).

Suppose that you're having a token experience *e*, say of a red tomato. Moreover, suppose for the moment that a particular version of the same-order monitoring account of experience is true, one according to which: (i) *e* is a complex mental state token, one of whose components is a further mental state token *m*; (ii) *m* is a state of awareness to the

effect that you're having *e* (i.e. for *m* to occur is for you to be aware of having a particular experience of a red tomato); and (iii) *e* is an experience in virtue of having *m* as a component. In this case, you might think that *m*, the constitutive experience-making state of awareness *vis-à-vis e*, is a token state of phenomenal acquaintance whose content is *e* or perhaps a phenomenal character of *e* (cf. Smith (1989, 70)). A consequence of this proposal is that *all* of our current experiences involve phenomenal acquaintance. Here phenomenal acquaintance comes with, as it were, each of your current experiences because it's that in virtue of which each of your experiences is an experience.

I bring this up to point out that this is *not* how I'm thinking of phenomenal acquaintance. Phenomenal acquaintance as I'm conceiving of it involves *attending* to an experience and thereby grasping a phenomenal character of it. In my view it's simply the cognitive relation you bear to a particular phenomenal character in virtue of attending to an experience that has it. Attention or focal awareness is cognitively costly, and, as such, at any given moment it's not the case that you're acquainted with each phenomenal character of each of your current experiences. Indeed, my contention is that often you aren't acquainted with *any* phenomenal character. So the point is that if there is a sort of awareness by virtue of which your experiences are experiences, this isn't the sort of awareness at issue with phenomenal acquaintance.

Recall Block's (1995, 2003) distinction between attention and awareness. Suppose, for example, that you're involved in an intense conversation while a jackhammer outside causes you to raise your voice, but you don't attend to the noise until your interlocutor later comments on it. You having adjusted the volume of your voice, so the idea goes, shows that you were *aware* of the loud noise in the absence of *attending* to it. So if the same-order monitoring account is true, perhaps experience-making states of awareness are states of mere

awareness rather than states of attention, while attention but not mere awareness is involved in phenomenal acquaintance.

Sosa (2003) makes a similar claim in distinguishing between what he calls 'n-awareness' – awareness *qua* attention – and 'e-awareness' – awareness *qua* peripheral awareness that is perhaps constitutive of experience. In his terminology, phenomenal acquaintance is a matter of n-awareness rather than e-awareness. Let 'E' be an experiential type such that, for any token experience e, e is an E-experience just in case e has phenomenal characters C_1 and C_2… and C_n. Now, suppose that you have an E-experience at a time t_1, and let 'e_1' be the E-token that occurs then. On one way of reading Sosa (I'm not sure if this is actually his view), you're peripherally aware of the occurrence of e_1 at t_1 (given the assumption that a mental state token is an experience in virtue of you being peripherally aware of its occurrence), and if you focus your attention, you form an *additional* state of awareness at a latter time t_2, one of a token e_2 of E distinct from e_1. (I'm assuming that e_1 and e_2 are distinct because they occur at different times.) So on this view, when you attend to your having an E-experience at t_2 (i.e. you attend to e_2 at t_2) you are, as it were, *doubly* aware of having E; at t_2 two token states of awareness p_2 and f_2 occur with the same content (they both represent e_2), where the former is peripheral and the latter is focal in nature. p_2, we're assuming for the moment, is that in virtue of which e_2 is an experience, and f_2 is that in virtue of which you're acquainted with some phenomenal character of e_2 at t_2.

According to the proposal above, the experience token you're e-aware but not n-aware of occurring at t_1 and the experience token you're both e-aware and n-aware of occurring at t_2 are both E-experiences – they have the same phenomenal characters C_1 and C_2… and C_n. You might think, however, that certain aspects of the phenomenal character of your experiences supervene on which experiences you happen to attend to (Siewart (2007)),

so these token experiences can't both be *E*-experiences.[4] To this end, you might claim, for example, that when you're merely peripherally aware of having a particular experience as of a red tomato, its phenomenal color is presented as being essentially general in nature (e.g. it's presented as being reddish but not as having any more determinate phenomenal color), while when you're focally aware of having such an experience, its phenomenal color is presented as being more specific in nature (e.g. as being bright-reddish). From this you might conclude that the phenomenal color of the former is not only presented as being essentially general in nature but it is so in reality; and the phenomenal color of the latter is not only presented as being more determinate but it is in fact more determinate. If this is right, then perhaps we should say that e_1 and e_2 from above are instead type-distinct in the sense that they aren't both *E*-experiences (they don't have the same overall phenomenal character), contrary to our initial assumption. (In this case we could grant that they're type identical in the sense that there are particular phenomenal characters, e.g. *reddishness*, they share, but they're type distinct in the more general sense mentioned above.) So we might want to say instead the following: (i) at t_1 you're peripherally aware of the occurrence of e_1; (ii) at t_2 you're doubly aware of the occurrence of e_2 in the sense that both p_2 and f_2 occur; but (iii) e_1 and e_2 don't have the same overall phenomenal character.

So now we have two proposals, and their dimension of difference concerns whether you can be aware of an experience with the same phenomenal character when focal awareness enters the picture. The first is that when you have, say, an experience as of a red tomato with a certain overall phenomenal character, you're peripherally aware of having it,

[4] In the next section, I introduce the idea that the phenomenal characters of our experiences are *appearances*, i.e. ways non-experiential aspects of the world appear. On this conception of phenomenal character, the proposal above is that the way non-experiential aspects of the world appear changes depending on which experiences you attend to.

and you can add to peripheral awareness a layer of focal awareness of having an experience with that phenomenal character. The second is that, when you have an experience as of a red tomato with a certain overall phenomenal character, you're peripherally aware of having it, and when you focus your attention you add an additional layer of awareness, but you always latch onto an experience with a different overall phenomenal character. Here it is thought that the addition of focal awareness is accompanied by a change in experience; you start with a token experience whose phenomenal color is essentially generally in nature, and then when focal awareness enters the picture it gets replaced with a token experience with a more specific phenomenal color.

Horgan and Kriegel (2007), as I read them, disagree with the second proposal in claiming that the token experience you're aware of at t_1 and the token experience you're aware of at t_2 are (or at least can be) both *E*-experiences. And they disagree with both proposals in claiming that when you're focally aware of having an experience, you aren't doubly aware of having it in the sense that tokens of type distinct states of awareness (peripheral and focal) occur. For they claim that the states of peripheral awareness by virtue of which experiences are experiences *become* states of focal awareness by virtue of the shifting or redirection of attention. So for them at t_1 you're peripherally aware of the occurrence of e_1, at t_2 you're focally but not peripherally aware of the occurrence of e_2, and e_1 and e_2 have (or at least can have) the same overall phenomenal character. (Here we would have to assume that focal awareness, in addition to peripheral awareness, is sufficient for experience, so e_2 is an experience in virtue of the occurrence of f_2.)

Another dimension of difference, then, for accounts of the relationship between peripheral and focal awareness on the one hand and experience on the other concerns how we think about the relationship between these types of awareness. A third proposal is that

when you have an experience as of a red tomato with a certain overall phenomenal character, you're peripherally aware of having it, and you can also come to be focally aware of having an experience with that phenomenal character, but not in a way that involves doubling up on awareness. A fourth proposal is that when you have an experience as of a red tomato with a certain overall phenomenal character, you're peripherally aware of having it, and when you focus your attention, you latch onto an experience with a different phenomenal character like the one described by the second proposal, but again not in a way that involves double awareness.

What should we make of the four proposals outlined above? I wish to officially stay neutral on which of the proposals (if any) is correct. The important thing to keep in mind is that phenomenal acquaintance is a matter of focal awareness: to be acquainted with a phenomenal character is to be focally aware of having an experience and thereby grasp one of its phenomenal characters in thought. This is compatible with each of the four proposals. But I would like to say something about which proposal I think is the most plausible.

To begin, I think that the second and fourth proposals are implausible because they proceed upon an implausible claim about the instantiation of phenomenal character. Recall the idea that e_1 and e_2 from above aren't both *E*-experiences because the phenomenal color of the former is essentially general in nature (e_1 instantiates the determinable *reddishness* but fails to instantiate any determinate of this property) while the phenomenal color of the latter is more specific in nature (e_2 instantiates *reddishness* and various determinates of it). But given the logic of the determinable/determinate relation, there just can't be such a thing as an experience that instantiates *reddishness* but fails to instantiate any determinate of it. For the having of a determinate is a way of having a determinable, and if you instantiate a determinable, there must be some way you have that determinable. So if, for example, you're

red, there must be some way in which you are red, for example scarlet. The same goes for *reddishness*. So if the plausibility of the claim that e_1 and e_2 must have different overall phenomenal characters rests on the model set out above according to which e_1 is essentially general in nature, it seems that we should go with either the first or third proposal claiming that e_1 and e_2 are (or at least can be) *E*-experiences. Though the model set out above is implausible, we can say, however, this: in being peripherally aware of having an *E*-experience at t_1, you represent it as instantiating *reddishness* but no determinate of this property, while, in virtue of being focally aware of having an *E*-experience at t_2, you represent it as instantiating *reddishness* as well as various determinates of this property. So, though there isn't a difference in the specificity of properties of the objects of peripheral and focal awareness *per se* at t_1 and t_2, there is a difference concerning the specificity of the properties those experiences are represented as having.

If the second and forth proposals are out, what should we say about the first and third proposals? I think that the choice between these two is a bit less clear-cut. Against the first proposal, it sounds a little odd to say that you can be both peripherally and focally aware of something simultaneously. Doesn't one preclude the other? Here's a reason why you might think so. As I have already suggested, awareness is underwritten by mental representation. In the next section I'll argue that to attend to an object is for you to token a mental representation of it that plays a certain role in your cognitive economy (the role characteristic of attention), and the same goes for states of peripheral awareness. Assuming that awareness is underwritten by mental representation and types of awareness are individuated in terms of the role that constituent mental representations play, we may have a problem. For it seems that if a mental representation plays the role characteristic of peripheral awareness, it can't simultaneously play the role characteristic of focal awareness.

To see why, suppose that you're peripherally aware of having an experience as of a red tomato. What is the role of the mental representation in virtue of which you're peripherally aware of having this experience? This is a difficult question, but we can at least say this much: the content of the mental representation in question – that you're having a reddish experience – is available to central processing so as to be, for example, poised for use as a premise in reasoning, but it's not as readily available as it would have been were the mental representation in question to have played the role characteristic of focal awareness. But if this is right, then it's clear that you can't be both peripherally and focally aware of an experience in virtue of tokening a single mental representation of it.

The advocate of the first proposal, however, doesn't claim that a single mental representation plays two different roles. She claims instead that being simultaneously peripherally and focally aware of having a particular experience involves two mental representations, where a token of the first plays the role characteristic of attention while a token of the second plays the role characteristic of peripheral awareness. There is nothing incoherent about the proposal so understood. But the intuition that there is something wrong about the first proposal persists (for me, anyway). The problem, I think, is that the proposal has an air of profligacy. For notice that anything that the token mental representation that plays the role characteristic of peripheral awareness does the token mental representation that plays the role characteristic of attention does as well. On the first proposal there is a sense, then, in which a single contribution to your cognitive economy is made twice over whenever you attend to an experience. Here the execution of certain role in your cognitive economy is overdetermined, and methodological considerations involving simplicity and economy recommend against such overdetermination. This, though by no means a decisive objection against the first proposal, gives us some reason, I think, to go

with the third proposal instead. So I think that, of the four proposals discussed above, the third one may be the most plausible.

1.3. Experiential Transparency

Above I talked as if it's quite obvious that you can attend to an experience and thereby grasp an aspect of its what-it's-likeness, a particular phenomenal character of that experience. But consider the intuition that experience is transparent, the idea, very roughly, that in introspection we "see through" our experiences, grasping only features our experiences represent various non-experiential aspects of the world as having (cf. Harman (1990) and Tye (1995)).[5] Here I use the term 'aspect of the world' so as to stay neutral on the ontological category of those entities represented by our perceptual and non-perceptual intentional states, be these states veridical, illusory, or hallucinatory. (At times I speak as if ordinary physical objects are those items represented, but I officially stay neutral on this issue.) Suppose, for example, that there is a red tomato on the table before you, and you're having a visual experience that represents it as being red. The transparency intuition applied to this case is something like this: when I ask you to attend to your visual experience and grasp its reddish character, what you end up singling out in thought is the property *redness*, and you grasp this feature as a property of the tomato. In other words, in this case what you are presented with in introspection is just the property *redness*, and it's presented as a feature

[5] Some have proposed weaker readings of experiential transparency. You might, for example, claim only that a particular type of experience, e.g. visual experience, is transparent in the sense specified above (cf. Dretske (1995, pg. xv)). Alternatively, you might claim that in introspection you can grasp features as features of your experience, but these features are fairly non-specific and uninteresting, such as being such that you're aware of a particular green (cf. Lycan (1995)).

of the tomato. In this case you do not, so the idea goes, grasp any feature as a feature of your experience; no such feature is presented in introspection as a feature of your experience.

Here's another way to put the point about experiential transparency. In the previous section I distinguished between two sorts of awareness – peripheral and focal awareness of experience – but Levine (2006a) usefully distinguishes between three sorts of awareness in the vicinity. His tripartite distinction goes like this. Suppose you're having a visual experience. First, there is a relation between you and the primary object of your experience, whatever state of the external world is presented to you in your experience. This is the first sort of awareness. Second, in having this experience, it seems that you're simultaneously aware of your having the experience, though this awareness doesn't involve attention. This is the second sort of awareness, experience-making awareness or e-awareness. Third, you can explicitly contemplate your experience, focusing your attention on it. This third type of awareness, focal awareness or n-awareness, I claim, is the one relevant to phenomenal acquaintance. With respect to this tripartite distinction, the claim that experience is transparent is the claim that experiential introspection is exhausted by the first sort awareness, awareness with respect to the intentional content of experience. So, *contra* Levine, it's unclear that there is anything like peripheral or focal awareness of experience as such, or so the transparency intuition goes.[6]

The worry for my conception of phenomenal acquaintance should be pretty obvious. If experiences are transparent in the sense specified above – if in introspection we fail to grasp features as features of our experiences – it would seem that in introspection we fail to

[6] Philosophers typically discuss the phenomenon of experiential transparency in the context of defending representationalism as a physicalist account of experience; see, e.g. Tye (1995, 2000). For the record, I don't think that experiential transparency in fact supports the thesis (cf. Loar (2003), Kind (2003), Siewart (2004), and Schroer (2007)).

grasp our experiences as such. This means, among other things, that we can't (or at least we never succeed in our effort to) attend to our experiences. But if we never attend to our experiences, we're obviously never acquainted with the phenomenal characters of them in my sense, given that in my view phenomenal acquaintance is underwritten by experiential attention. I agree that there is something to the transparency intuition, so I have some work to do.

Let's begin by getting clearer on the nature of attention. What is attention, anyway? Well, taking a step back, what is it to have a thought in the first place? I endorse the representational theory of mind, which says that to have a thought is to token a mental representation that plays a certain role in your cognitive economy. There are obviously different kinds of thoughts, and thoughts are individuated in terms of the role of their constituent mental representations. On this view, to have a state of attention is to token a mental representation and for this mental representation to play the role characteristic of attention. So suppose that you're sitting in your office and looking at the various objects on your desk. Here you token various mental representations, two of which we will suppose are of a coffee cup and a pencil. For you to attend to the coffee cup instead of the pencil is for your coffee cup representation to come to play a role that your pencil representation doesn't – the role characteristic of attention. Specifying the nature of the role characteristic of attention (or any type of thought) is a difficult matter as I noted in the previous section, and it falls outside the scope of this chapter to pursue this matter in any detail. But part of the role of attention obviously includes relations to beliefs. When you attend, for example, to the

coffee cup instead of the pencil, you come to have new perceptual beliefs about the coffee cup but not necessarily about the pencil.[7]

With the above conception of attention in mind, we can at least model the distinction between attending to an experience and grasping a feature as a feature of it on the one hand and attending to a non-experiential aspect of the world and grasping a feature as a feature of it on the other. Suppose again that you're having a veridical perceptual experience as of a red tomato on the table. To attend to your experience and grasp its reddish character rather than to attend to the tomato and grasp its redness is for you to token at least two mental representations – one that represents your experience as being reddish and one that represents the tomato as being red – and for the former but not the latter to come to play the role characteristic of attention in your cognitive economy. But it's one thing to provide a model for this distinction and quite another to show that we actually implement the model, that we have the sorts of minds that token mental representations of our experiences as having various features that play the role characteristic of attention. We can understand the objection from experiential transparency as the claim that our minds just don't seem to work this way.

Can we do better by way of defending the claim that we can attend to our experiences, grasping features as features of them? The remainder of this section is devoted to answering this question. First, I consider one way of defending the claim that appeals to imagination and hallucination, but I argue that it may not work. Second, I propose what I take to be a better argument appealing to illusory experiential perception. Third, I argue that

[7] Another important issue with respect to understanding attention is this. Attention is a form of *de re* thought as opposed to belief, which is *de dicto* in the sense that canonical belief ascriptions have a predicative element ("He believes that x is F"), whereas canonical ascriptions of attention do not ("She attends to an F"). Explaining the *de re* character of attention is an interesting a formidable task, but one I won't pursue here.

we can motivate the claim that we can attend to our experiences independently from considerations involving non-standard cases of experiential perception. Fourth, I offer a diagnosis of why some wrongly think that considerations involving experiential transparency show that we can't attend to our experiences. Fifth and finally, I respond to an argument for the claim that we can't attend to our experiences, one that appeals to the idea that there is no phenomenological difference between attending to an experience and attending to what it represents.

Let's begin with hallucinatory experience. While we can grant for the sake of argument that there is a dominant, untutored manner of conducting experiential introspection such that we're only presented with features as features of non-experiential aspects of the world, perhaps there is a way of supplementing experiential introspection such that we're presented with features as features of our experiences. The idea is that what we need to do is supplement straightforward experiential introspection by performing certain imaginative tasks. The proposal in particular is that you can make changes in imagination to your perceptual environment, taking you from a veridical perceptual experience (on the assumption that your original experience in fact is veridical) to a corresponding hallucinatory perceptual experience (cf. Loar (2003)). No matter how the content of your original experiential state was determined, be it solely by external environmental features, features internal to you, or some combination of both, something associated with your experience holds constant through the imaginative change. This experiential residue, according to the proposal, is a phenomenal character of your experience, a phenomenal property instantiated by it. The proposal, then, is that, when you supplement straightforward experiential introspection with the sort of imaginative task described above, you are presented with

features as features of your experiences rather than as features of non-experiential aspects of the world; you grasp these features as features of the former rather than the latter.

I think it's clear that we can perform the sort of imaginative task described above and thereby single out in thought the commonality between actual veridical and imagined hallucinatory perceptual experiences, and it's natural to think that this commonality is a phenomenal character the experiences share.[8] The question, however, is whether such a commonality is presented as a feature of a non-experiential aspect of the world or as a feature of experience. The representationalist will presumably claim that it's presented as a feature of some non-experiential aspect of the world, given her claim that the commonality is nothing over and above the intentional content shared by both experiences. When I perform the imaginative task myself, grasping in introspection the commonality between my actual veridical perceptual experience and my imagined hallucinatory one, I don't clearly register it as a feature of experience or as a feature of a non-experiential aspect of the world. So, from my perspective, anyway, the argument under consideration is inconclusive.

Now let's turn to illusory experience; perhaps here we will find a better way of reconciling the intuition that experience is transparent with the idea that we can attend to our experiences, grasping phenomenal characters as features of them. Before I get to the proposal, however, I need to say something more about phenomenal character. Now, I want to stay as neutral as I can on the nature of phenomenal character, but here I'll set out a

[8] I find the claim that there is a single phenomenal property that both the veridical perceptual experience and the imagined hallucinatory one instantiate quite plausible. Disjunctivists about phenomenal character, however, e.g. Langsam (1997), deny this; they claim that, though the phenomenal character of each experience may be indiscernible with respect to what it's like to have experiences with that phenomenal character, the phenomenal character of the veridical perceptual experience and that of the hallucinatory one are type distinct.

substantive thesis about it that strikes me as plausible.[9] Paradigmatically, a particular phenomenal character is an *appearance* – a particular way a non-experiential aspect of the world appears – and an experience is a cluster of appearances, or so I claim.[10] Suppose again that you're having a visual perceptual experience with a reddish character. Given my claim above, this is just to say that you're having an experience an aspect of which is a non-experiential aspect of the world looking red.[11] Some claim that if your perceptual experience is veridical or illusory, it's a physical object that appears red, while if your experience is hallucinatory, something from a different metaphysical category is that which does the appearing (cf. Langsam (1997) and Levine (2006b)).[12] Others claim that in either case the same sort of thing appears red, and among our options here are regions or locations (cf. Clark (2004)) and intentional inexistents or virtual objects (cf. Levine (2008)). Just as I stay

---

[9] I appeal to this thesis about phenomenal character in responding to the concern about transparency here and later in this section. If you're inclined to reject this thesis, note that later in this section I provide another way of resolving the transparency concern that doesn't appeal to it. I do, however, appeal to the thesis in the fourth chapter in responding to an objection to my account of phenomenal acquaintance.

[10] I have more to say about the structure of experience in the fourth chapter. There I argue that a token experience is itself a property instance and is composed of further property instances. So a token visual experience as of a red tomato – itself a property instance – is composed by further property instances, including an instance of the phenomenal property *reddishness*.

[11] My current tactile experience consists of various tactile appearances (non-experiential aspects of the world feeling shaped or textured); my current gustatory experience consists of various gustatory appearances (non-experiential aspects of the world tasting various ways); and so on. Each of these appearances, so the idea goes, is a phenomenal character.

[12] Langsam (1997) claims that phenomenal characters are appearances and appearances are relations. He claims that different kinds of objects do the appearing in cases of veridical and hallucinatory experiential perception, so he claims that the phenomenal characters of veridical and hallucinatory experiences *qua* relations take different kinds of relata. This is why he claims that the phenomenal characters of veridical and hallucinatory experiential perceptions are type distinct; see note 8.

neutral on the ontological category of those entities represented by our intentional states, here I likewise remain neutral on just what sort of thing does the appearing.[13]

If phenomenal characters are appearances as I claim above, then, if in introspection we're presented with sensory appearances as features of our experiences rather than as features of non-experiential aspects of the world, we can indeed attend to our experiences. Perhaps it would be sufficient to show, for example, that in introspection you can grasp a non-experiential aspect of the world as *looking* red, elliptical, etc. in contrast to grasping a non-experiential aspect of the world as *being* red, elliptical, etc. (cf. Siewart (2004)). So suppose, for example, that there is a red tomato before you on the table that looks blue, though you know both that it's actually red and that it appears blue only due to unusual lighting conditions. It seems clear that in this case you can single out in thought the blue appearance, the tomato looking blue. We're identifying sensory appearances with phenomenal characters, so in this case you grasp in introspection the bluish character of your experience. The objection from experiential transparency is that in introspection we fail to grasp features as features of our experiences; instead, we grasp them as features of non-experiential aspects of the world. In our example, the objection would be that in introspection you only grasp the property *blueness*, and you represent this feature as a feature of the tomato. But this patently isn't the case – in our example you don't represent the tomato as being blue; instead, you represent it as being red.

What we've done so far is establish that there is a clear difference between grasping a bluish phenomenal character and grasping *blueness*. So now the question is this: is the blue

[13] We also have imagery experiences; I can, for example, visualize Santa Clause, silently "talk to myself" or "listen to a song in my head", and so on. The phenomenal character as appearance model can accommodate such experiences. In the case of visualizing Santa Clause, for example, there is a sense in which the image I entertain in thought appears red, white, and so on. I'll put imagery experiences to the side, however, in what follows.

appearance (the bluish phenomenal character) from the example above presented as a feature of your experience in introspection? At the very least it's clear that we *can* represent appearances – things appearing certain ways – as features of our experiences. And when I, for example, look at the penny on the table before me, knowing that while it appears elliptical it's actually circular, it seems to me that the elliptical appearance is presented as something that belongs, as it were, to my experience rather than to the penny.

Here it seems that we have cases such that in introspection we grasp certain features as features of our experiences. These cases suggest not only that we can attend to our experiences, but they also give us an idea of how our being able to so is compatible with the intuition that experience is transparent. If phenomenal characters are indeed sensory appearances and experiences are clusters of appearances, we can all agree that experience is transparent in the sense that it has a "directional" character; it concerns non-experiential aspects of the world appearing certain ways (cf. Levine (2008)). And it seems that, once you've appreciated that an object, say, appearing red and that object being red are different, it isn't difficult to frame your attention in such a way that a red appearance is presented straightaway as a feature of your experience, even if you believe that your experience is veridical in nature.

I think that we can also motivate the claim that we can attend to our experiences without appealing to illusions or hallucinations. Suppose, for example, that, though you've never had wine before, you now find yourself at a wine tasting, sampling various wines. Suppose you ask yourself, "What are some of the differences with respect to the taste of wines *A* and *B*?" and all you can come up with is something like "*A* tastes like this to me, while *B* tastes like that to me". Here it seems that the tastes of *A* and *B* are presented to you in thought only under descriptions that appeal to appearances; they lack non-experiential

modes of presentation. You conceive of the tastes of *A* and *B* only relative to their appearing to you this way or that. In this case you attend to *A* and *B* and thereby grasp their respective tastes in gustatory perception in virtue of (and only to the extent that) you grasp certain features as sensory appearances, as features of your experience. So it seems that in cases in which you lack non-experiential modes of presentation of the features of the objects of perception (be your perceptual experience veridical, illusory, or hallucinatory), a requirement for attending via perception to such objects is that you're presented with certain features as appearances of those objects, and these appearances are presented as features of experience (cf. Siewart (2004)).

Above I proposed three arguments for the claim that we can attend to our experiences, two of which I find plausible. If I'm right in thinking that we can attend to our experiences, why is it that many philosophers seem to think that considerations involving experiential transparency suggest that we can't do so? Well, consider the fact that we can shift our attention from one object to another. Supposing that there is a pencil and a coffee cup on your desk, you can, for example, draw your attention away from the coffee cup towards the pencil. I suspect that many think that if you have a veridical perceptual experience as of a pencil, attending to this experience would require that you draw your attention *away* from the pencil *towards* your experience of it. I suspect, moreover, that many think that experience is transparent precisely in the sense that, try as you might, such a shift in attention can't be achieved; you can't move the spotlight of attention from the pencil to your experience as of a pencil (cf. Martin (2002)). But as Siewart (2004) and Schroer (2007) point out, we're free to claim that attention to those non-experiential aspects of the world represented by our perceptual experiences and attention to those experiences *come as a package*. In this case, the idea that attending to an experience requires drawing your attention

away from a non-experiential aspect of the world toward the experience itself is just confused. So it may be that attention to a perceptual experience can't "block" attention to what it represents, but this, of course, is consistent with the claim that you can attend to the experience. It's just that when you attend to an experience, you also attend to whatever it represents.[14]

Here is a final observation on experiential transparency. In discussing experiential transparency, many claim that they have no introspective purchase on the idea that we can attend to our experiences; they claim that in introspection they can't tell what the difference is supposed to be between attending to experiences and attending to what they represent. Here the claim might be that there is no relevant *phenomenological* difference registered in experiential introspection. These considerations suggest something like the following argument against the idea that we can attend to our experiences: there is something to attending to an experience over and above attending to what it represents only if there is a phenomenological difference between these putatively distinct cognitive operations; but, since there is no such difference, there is no attending to experience as such. So suppose that you're having a veridical perceptual experience as of a red tomato. The claim is that there is a distinction between attending to the tomato and attending to your experience of it only if there is a corresponding difference in phenomenology; but, since there is no such phenomenological difference, there is no attending to your experience as of a red tomato over and above attending to the red tomato.

[14] Block (1996, 2003) rejects the transparency intuition from the get go and denies that these sorts of awareness come as a package, claiming that we can attend to certain non-visual experiences (e.g. sensations and moods) as well as various visual experiences (e.g. phosphene experiences) without also attending to those non-experiential aspects of the world that they represent (if they represent anything at all).

Is this a good argument? Well, it's only as good as its central assumption, that attention to experience, if possible, has a distinctive phenomenology, one distinct from that of attention to non-experiential aspects of the world. But is this assumption correct? I'm not so sure that is. We should all agree that there is something it's like to attend to a red tomato, grasping its red color. This isn't under dispute. Now, if there is something it's like to attend to your experience, what is it like? In other words, what would the (or a) phenomenal character of this experiential state be? Well, according to the appearance model of phenomenal character outlined earlier, the reddish character of your state of attending to the tomato is just a non-experiential aspect of the world looking red. Assuming that the appearance model of phenomenal character is perfectly general, we would have to say something similar about the phenomenal character of you attending to your experience and grasping its reddish character. It would seem that in this case what we should say is that the phenomenal character of you attending to your experience and grasping its reddish character is this: an experience appearing to be reddish. (Here we, of course, have to drop the requirement that what does the appearing is a non-experiential aspect of the world.) This is just for the experience to appear such that an aspect of it is a non-experiential aspect of the world looking red. Notice, however, that there is no phenomenological distinction between appearing to look red and looking red. So the picture we end up with here is this: there is something it's like to attend to your experience and grasp its reddish character, but this what-it's-likeness is just the what-it's-likeness of attending to a tomato and grasping its red color.

It seems, then, that we have some reason to reject the idea that attention to experience requires a distinctive phenomenology. There is certainly something it's like to attend to an experience, but it's unclear that this what-it's-likeness outstrips the phenomenology associated with attending to those non-experiential aspects of the world

perceptual experiences represent. And it's unclear why we should think that it should do so. But let's grant for the sake of argument that attention to experience requires a distinctive phenomenology. In this case, we can say that the overall phenomenology of attending to an experience contains a distinctive aspect as well as a non-distinctive derivative aspect, one shared or inherited from the phenomenology of experiential perception. If this is right, then perhaps those who think that there is no attending to experiences over and above attending to what experiences represent are focusing solely on the derivative, shared aspect of the phenomenological status of attending to experience. So on this proposal, we're granting that attention to experience requires a distinctive phenomenology, and we've offered an explanation of why some claim that no such phenomenology is registered in experiential introspection.[15, 16]

1.4. Conclusion

In this introductory chapter I introduced the notion of phenomenal acquaintance. Phenomenal acquaintance is a matter of attending to your experiences and grasping particular phenomenal characters of them. So the objects of phenomenal acquaintance are phenomenal characters, and you come to be acquainted with a particular phenomenal

---

[15] Just what might the distinctive phenomenology of attention to experience be? While I officially stay neutral on this issue, in the next chapter I introduce what I take to be three key features of phenomenal acquaintance, and you might claim that some or all of these features are distinctive *phenomenological* features of phenomenal acquaintance. Of the three features I discuss, I think that what I call 'directness' is the best candidate for being a phenomenological aspect of phenomenal acquaintance.

[16] Spener (unpublished ms) argues that the fact that philosophers disagree about whether particular experiences (and experience in general) are transparent suggests that experiential introspection isn't a reliable belief-forming mechanism. Schwitzgebel (2008) offers his own argument from disagreement in support of the same conclusion, one appealing to the debate concerning whether there is a phenomenology distinctive of cognition. See Siewart (2007) for (to my mind persuasive) responses to these challenges.

character in virtue of attending to an experience that has it. I argued that if we wish to give an account of phenomenal acquaintance we need to do more than merely impute to it whatever features might explain away our dualist intuitions at work in familiar objections to physicalism. Then I contrasted the states of awareness at issue with phenomenal acquaintance with those states some claim are at the very nature of experience. Finally I argued that considerations involving experiential transparency in the end don't tell against the idea that we can attend to our experiences.

In this dissertation I propose an account of phenomenal acquaintance. But what precisely is the explanatory project here? In other words, what precisely are we trying to explain or account for in constructing an account of phenomenal acquaintance? And why be interested in this project in the first place, provided that we can give some content to it? Addressing these issues takes up the bulk of the next chapter. First I set out what I take to be three central features of phenomenal acquaintance. I treat the idea that phenomenal acquaintance has these features as something like a pre-theoretical datum. These features, I argue, provide content to our explanatory project in the sense that they provide desiderata for an account of phenomenal acquaintance. An adequate account of phenomenal acquaintance, I argue, must explain why it has these features, or at least why it seems to have them. Then I argue, however, that with phenomenal acquaintance we face a certain dilemma, for though it certainly seems that it has these features, it's really quite puzzling how phenomenal acquaintance (or any cognitive relation, for that matter) could have them. So one reason to be interested in the notion of phenomenal acquaintance is that, on the face of it, it's quite difficult to understand. Resolving this dilemma is the main project of the dissertation, and I do so in the fourth chapter. In the fifth chapter I trace out the consequences of my account for physicalism.

CHAPTER 2

WHY PHENOMENAL ACQUAINTANCE IS PUZZLING

2.1. Introduction

In this chapter I first argue that there are three central features of phenomenal acquaintance – I claim that it's direct, thick, and infallible in senses to be specified – and then I propose three corresponding adequacy conditions for an account of it. The idea that phenomenal acquaintance has these features strikes me as something like a pre-theoretical datum, and we'll use these features as constraints on our theorizing about it. Second, I consider these features from a different perspective, arguing that it's actually quite puzzling how phenomenal acquaintance (or any kind of thought, for that matter) could be direct, thick, and infallible. We find ourselves, then, in a kind of dilemma. On the one hand, it seems that, in virtue of attending to our experiences, we have thoughts about particular phenomenal characters that are direct, thick, and infallible. Yet, on the other hand, it's hard to see how any thought could really have these features. Third, I contrast two general conceptions of phenomenal acquaintance, what I call the 'subjective' and 'representational' conceptions. In later chapters I consider accounts of phenomenal acquaintance from both conceptions: in the third chapter I consider a version of the subjective conception inspired by (the early) Russell's discussion of acquaintance, and in the fourth chapter I set out and defend a version of the representational conception. I argue that the latter but not the former is up to the task of explaining how phenomenal acquaintance could be direct, thick, and infallible.

2.2. Directness, Thickness, and Infallibility

As you read these words, you're having various experiences. You're having, for example, a visual experience as of a white page with black letters on it, perhaps a tactile experience as of smooth paper in your hands, perhaps a gustatory experience as of the bitter taste of coffee in your mouth, and so on. Focus your attention on your visual experience and zero in on its blackish character. There is, I think you would agree, something special about this thought – your acquaintance with the blackish character of your experience – that your thoughts about non-experiential things lack. Three features of this thought come to mind.

First, it seems, following Russell (1911, 1912, 1913a), that your cognitive access to the object of your thought – the sense in which the blackish character of your experience is available to cognition – is *direct* in that you don't register any cognitive distance between yourself and its object. Contrast this with your grasp of non-experiential things. Suppose, for example, that, in looking at the objects on your desk, you single out your cell phone and grasp it in thought. Assuming the representational theory of mind, your cell phone is before your mind in thought in the sense that it's the content of a mental representation. So we can say that your cognitive relation to it is mediated in that you make cognitive contact with it only through an intermediary, a particular mental representation. So here we have three items – you, a mental representation, and a cell phone – such that the second item mediates the connection between the first and third. Intuitively, your cognitive relation to the blackish character of your experience isn't mediated in this sense. It seems that you don't make cognitive contact with this experiential feature only by proxy, via some further item (a mental representation) distinct from both you and it. Your cognitive contact seems more intimate than this. The blackish character of your experience, in other words, seems to enjoy

a special kind of cognitive presence when you attend to that experience; it's as if the blackish character is somehow included in your thought about it.

Second, the grasp of the blackish character that attending to your visual experience affords you, following Levine (2001), seems *substantive* and *determinate*. It seems substantive in the sense that you don't grasp the blackish character as a "know-not-what". Contrast the substantive nature of this thought to what commonly happens in demonstrative thought. With demonstrative thought, you can have a thought about something, yet really have very little grip on what it is you're thinking about, as when you think, for example, "What the hell *is* that?" Your grasp of the blackish character seems determinate in the sense that, when you attend to your visual experience, you grasp its blackish character as a specific quality, identifiable in its own right, not merely by its relation to other qualities. Contrast this feature of your grasp of blackish character to the way you grasp, say, the property of being a heart. You grasp this property merely as something that stands in various relations to other features like veins, arteries, and so on. It seems, then, that there is a clear sense in which your grasp of the latter is relational, while your grasp of the former is not. Let's call the substantive and determinate nature of your grasp of the blackish character of your experience its 'thickness' for short.

Horgan and Kriegel make a claim similar to the claim that phenomenal acquaintance is thick in their discussion of "bracketing" modes of presentation of phenomenal character:

> This is a mode of presentation that brackets out all relational information about the experience and its phenomenal character, including how experiences of this sort are classified by other subjects, how they are classified by oneself on other occasions, what their typical causes are, etc. It focuses (so to speak) on how the experience appears to the subject at that moment (2007, 128).

Here the claim is that you can entertain in thought a phenomenal character under a mode of presentation that, instead of presenting it as something that stands in various relations to

other items, just presents the what-it's-likeness of that phenomenal character (how it appears) which is, of course, just the phenomenal character itself.[1]

Third, as many have remarked, it seems that a certain sort of *infallibility* attends your grasp of the blackish character of your experience. Now, I don't think that anyone wishes to claim that any experiential judgment whatsoever can't go wrong – this claim is certainly too strong. Sometimes we make false judgments about the nature of our experiences when we're distracted. Moreover, we can obviously misapply phenomenological vocabulary in describing our experiences – certainly some of us have a better facility with respect to describing our experiences than others and are more careful in doing so. Though it's not the case that experiential thought in general is infallible, the intuition that there are nonetheless ways of thinking about our experiences in which we can't go wrong persists. Suppose, for the moment, that it's an open question as to what phenomenal characters your current visual experience has. Let's also suppose that you carefully attend to your current visual experience, and it's presented as having a blackish character. Finally, suppose that you have a good handle on phenomenological vocabulary and, in virtue of attending to your experience, you judge that it has a blackish character. This is the sort of experiential judgment that strikes

[1] When you conceive of a phenomenal character in this way, there is some question about how to describe the object of your thought in a public language. As Horgan and Kriegel (2007) note, words in public languages carry presuppositions of a kind that are suspended by bracketing modes of presentation, e.g. presuppositions to the effect that your present use of a word is semantically in accord both with how you have used it previously and with the standard uses of that word in your linguistic community. My go-to examples of phenomenal characters so far have been what I call 'blackish' and 'reddish' phenomenal characters. The use of these terms, however, appeals to these very sorts of presuppositions. Horgan and Kriegel note that perhaps the best you can do in describing a phenomenal character conceived of under a bracketing mode of presentation in a public language without flouting ordinary usage is something like "This experience has this feature". While this seems right, for ease of presentation I'll continue to say things like "Your cognitive relation to *blackishness* is thick", though keep in mind that, if we understand the thickness of phenomenal acquaintance in terms of bracketing modes of presentation, 'blackish' in this context doesn't function as you would expect it to.

many as the sort that can't lead us into error, and such judgments are underwritten by states of phenomenal acquaintance. This is the sense in which phenomenal acquaintance seems infallible.[2]

Above I claimed that it seems that phenomenal acquaintance is direct, thick, and infallible. Now, there are various ways you can think about a phenomenal character where your thoughts about it don't seem to have any of these features. You can, for example, think "I'm currently not having an experience with a blackish character" or "I'm sure I've had experiences with a blackish character before", and these thoughts seem to lack some or all of the features I've discussed (cf. Papinaeu (2007)). But, as I said above, when you *attend* to your experiences in the having of them, the resultant thoughts you have about particular phenomenal characters do seem direct, thick, and infallible in the senses specified above.

You might argue, however, that, contrary to first appearances, such thoughts are mediated in some important sense (cf. Tye (2000, 50)), our grasp of phenomenal character (or at least certain phenomenal characters) is never actually substantive and determinate (cf. Schroer (unpublished ms)), or experiential judgments underwritten by phenomenal acquaintance can go wrong (cf. Schwitzgebel (2008)). Part of this approach, of course, would be to explain why these thoughts seem to have these features when in fact they don't. A related strategy would be to agree that such thoughts have the features in question, but to

[2] Dennett (1991, 2001) and Schwitzgebel (2008) make much of the fact that subjects are prone to make false judgments about the size of the portion of their visual fields that enjoy uniform clarity. Does this finding tell against the infallibility of phenomenal acquaintance? I don't think so. I'm inclined to say that when a subject, for example, has a perceptual experience as of a desk and wrongly judges that the desk is presented in experience with uniform clarity, it's the subject's expectation of how desks *should* look that leads her astray. Here the idea is that, from our folk ontological belief that desks are unified, stable objects, we tend to expect that this should be reflected somehow in how desks are presented to us in visual experience. Making this mistaken judgment for this reason is compatible with the claim that experiential judgments underwritten by phenomenal acquaintance alone can't go wrong.

give some or all of them a deflationary treatment. Here the idea would be to argue, for example, that there is indeed a sense in which phenomenal acquaintance is infallible, but its infallibility is fairly trivial or uninteresting.

Though the debunking and deflationary strategies sketched above are certainly options, I think it's clear that they are to be pursued only upon losing confidence in the idea that such thoughts *seem* to have these features because they *do* in fact have them, and that these features are robust, as they seem to be. This is at least how the investigation should begin. It seems that we have, then, three desiderata for an account of phenomenal acquaintance: an adequate account must explain why phenomenal thought underwritten by attention to experiences is direct, thick, and infallible in the senses specified above.

2.3. Three Puzzles

Above I argued that phenomenal acquaintance is direct, thick, and infallible. In this section, however, I change course and argue that it's actually quite puzzling how any cognitive relation could have these features understood in a non-deflationary manner. There are three related puzzles corresponding to each feature discussed above. In order to understand the first puzzle, we first need to think more about the nature of the directness of phenomenal acquaintance. In virtue of what is phenomenal acquaintance direct? As I see things, there are at least three aspects to its directness. First, a state of phenomenal acquaintance is directly referential in the sense that its content and referent are one and the same. Second, such a state is demonstrative in the sense that its referent, a particular phenomenal character, instead of being fixed by a descriptive condition, is determined by a mental demonstration (a state of attention). The object of attention in this case is your experiential state that has the phenomenal character in question. Third, if you're acquainted

with a phenomenal character *P* of a token experience *e*, it's not the case that the content of your state of acquaintance was inherited from some further intentional item – one distinct from your state of acquaintance and *e* – that represents *e* as having *P*.

Why think that these theses are relevant to the directness of phenomenal acquaintance? Well, imagine that states of phenomenal acquaintance weren't directly referential, and that their contents instead were descriptive conditions, conditions that determine their reference. In this case, a description mediates the relation between you and the object of your acquaintance. The same would be true if a "character" – a descriptive condition "off the record" with respect to content (i.e. not a feature of content) – fixed their reference (cf. Kaplan (1989)). In either case, when you have a state of phenomenal acquaintance, you have cognitive access to its object only "through" a description, so we have lost the sense in which its object is cognitively present. Hence, my first two claims about the directness of phenomenal acquaintance seem well motivated.

Turning to the third claim, suppose, for example, that you imagine a situation in which you're smoking, and you thereby form a desire to be in that situation; you think to yourself, "I want to do *that*", where the referent of 'that' is smoking a cigarette as depicted in your imagination. The content of the demonstrative element of this thought is borrowed from the content of a prior mental representation, the one involved in your imagining yourself smoking. Let's think more about the structure of this scenario. We're supposing that a token demonstrative mental representation *d* occurs, one that represents something *x* as being *P* (as being a scenario in which you're smoking). Now, *d* represents *x* as being *P* only because some other token mental representation *r* occurs – the one that occurs in imagination – that already represents *x* as being *P*, where *d*, *r*, and *x* are all distinct. This is the sense in which *d* borrows the content of a prior mental representation and *r* plays a

mediating role between *d* on the one hand and *x* and *P* on the other. So, from the perspective of your demonstrative thought, your cognitive access to the imagined situation is mediated by the mental representation you token in imagination. In other words, from the perspective of the occurrence of *d*, you make cognitive contact with *x* and *P* only through an intermediary – *r*.

So now imagine that in phenomenal acquaintance your cognitive contact with phenomenal character proceeds by way of such intermediaries. In this case, you're acquainted with a phenomenal character of your experience only because you have some further mental state – one distinct from both your state of acquaintance and your experience – that already represents your experience as having that phenomenal character. In this case you have cognitive access to the phenomenal character of your experience only "through" this further mental item. Again, recalling my discussion of directness above, it seems that we have lost the cognitive presence of phenomenal character.

Having gotten a better handle on the directness of phenomenal acquaintance, we can turn to the first puzzle. This puzzle concerns the compatibility of the directness of phenomenal acquaintance as understood above with the conception of attention I set out in the previous chapter. Recall that on my view to attend to an experience is to stand in the attention relation to the content of a mental representation, one that already represents that experience. The puzzle we face is that this conception of attention seems to conflict with the claim that phenomenal acquaintance is direct. For to be acquainted with a phenomenal character, you have to attend to an experience that has it. But in order to do this, you must already have a mental state that represents the experience as having a certain phenomenal character. Here it seems that we have postulated precisely the sort of mediating material ruled out by the directness of phenomenal acquaintance. Call this the 'directness puzzle'; the

puzzle of explaining how phenomenal acquaintance could be direct in our sense yet underwritten by attention, which exploits the content of prior mental representation.

Here is perhaps a more intuitive way to put the problem. The directness of phenomenal acquaintance recommends something like the following structure:

$$\begin{array}{ccc} & & P \\ & \nearrow & \uparrow \\ \text{You} & \rightarrow & e \end{array}$$

Figure 1: A First Pass

Let the horizontal arrow stand for the relation of attention, the arrow pointing up and to the right the relation of acquaintance, and the arrow pointing up that of exemplification. This figure represents the fact that your acquaintance with *P* (a phenomenal character) is a matter of you attending to *e* (a token experience) that instantiates *P*. But our conception of attention suggests that the figure above leaves out something important; it suggests that acquaintance is more perspicuously represented by the following structure:

$$\begin{array}{ccccc} & & & & P \\ & \nearrow & & & \uparrow \\ \text{You} & \succ & r & \Rightarrow & e \end{array}$$

Figure 2: A Second Pass

In the first figure we represented you attending to your experience as 'You → *e*', but having gotten clear on the sense in which attention is underwritten by prior mental representation, we can represent you attending to your experience as 'You ➢ *r* ⇨ *e*', where '*x* ➢ *y*' means (roughly) '*x* tokens some mental representation, where *y* is the mental representation token in question' and '*x* ⇨ *y*' means '*x* represents *y*'. Here it's understood that *r* plays the role characteristic of attention in your cognitive economy, i.e. that you stand in the attention relation to the content of *r*.

I think that having a pictorial representation of phenomenal acquaintance and our conception of attention makes vivid the sense in which a commitment to the latter seems to commit us to mental representation playing an unacceptable mediating role, one incompatible with the directness of phenomenal acquaintance. *r* mediates the connection between you and *e* and *P* in the sense that *r* is distinct from *e* and *P* and you make cognitive contact with *e* and *P* only to the extent that they figure in the content of this further mental item.

The second puzzle concerns the compatibility of the infallibility of phenomenal acquaintance with the conception of attention developed above, and it's much easier to state. Above I claimed that you're acquainted with a phenomenal character in virtue of attending to an experience with it. I also claimed that you attend to an experience in virtue of tokening a mental representation, one that represents your experience and plays a certain role. Now, if this mental representation misrepresents – if it represents your experience, say, as having a reddish character when it in fact isn't reddish – then, if you thereby judge that your current experience is reddish, you would be mistaken. But earlier we said that this is the sort of judgment that seems incapable of going wrong. What I'll call the 'infallibility puzzle', then, is this: since mental representations in general can misrepresent, why think that the mental representations in virtue of which we're acquainted with the phenomenal characters of our experiences can't misrepresent?

The third and final puzzle concerns the compatibility of the idea that phenomenal acquaintance is direct and underwritten by attention to experiences on the one hand and that it's thick on the other. As I've said above, when you attend to an experience, you thereby grasp a particular phenomenal character of it in a non-relational manner. This is what I'm calling the 'thickness' of phenomenal acquaintance. Levine (2006, unpublished ms), however,

points out that experiences fall under many types – both experiential and non-experiential – and claims that it's therefore unclear how attending to an experience eventuates in a grasp of a particular phenomenal character of that experience. Suppose again that you attend to your visual experience as of a white sheet with black letters on it, and you thereby come to have a (non-relational) grasp of the blackish character of your experience. The following question arises: in virtue of what did you come to grasp this experiential feature as opposed to another, for example the whitish character of your experience?

In setting out the first and second puzzles, we considered the idea that when you attend to an experience, thereby coming to be acquainted with a particular phenomenal character of it, you do so in virtue of standing in the attention relation to the content of a prior mental representation, one that represents the experience as having the phenomenal character in question. The postulation of such a mental representation would certainly explain how you grasp a particular phenomenal character in virtue of attending to an experience with it, but it would seem to be inconsistent with the directness of phenomenal acquaintance, as we have already seen. So this potential explanation, it would seem, is off limits. What is the explanation, then? This third puzzle, what I'll call the 'thickness puzzle', is the problem of explaining how attention to an experience eventuates in a grasp of a particular phenomenal feature of that experience, given that the former seems to underdetermine the latter.

Here is another way to get at the thickness puzzle. Suppose again that you attend to your visual experience and thereby grasp its blackish character. Further, suppose that your grasp of the blackish character is direct in that it meets the three conditions I set out above. It seems that the directness of your thought so understood rules out two natural places of looking for the source of its thickness. For if your thought had descriptive content it

wouldn't be surprising to learn that this descriptive material is that in virtue of which your grasp of the blackish character of your experience is thick. The same would seem to hold if there was a reference-fixing descriptive condition for your thought "off the record" with respect to its content. But the directness of your thought rules out these two potential explanations of its thickness.

2.4. Conclusion

In this chapter I suggested that phenomenal acquaintance has three key features – it's direct, thick, and infallible. But then I argued that it's actually quite puzzling how phenomenal acquaintance could have these features. How should we proceed from here? Well, let's distinguish between two general conceptions of phenomenal acquaintance, one that I call the 'representational conception' that I implicitly endorsed in our discussion above, and another that is at odds with our discussion so far, the 'subjective conception'. According to the representational conception, phenomenal acquaintance is underwritten by mental representation. We've already considered one way of developing this idea: you're acquainted with a phenomenal character in virtue of attending to an experience that has that phenomenal character, which is to stand in the attention relation to the content of a mental representation. Recall the representational theory of mind according to which intentional mental states are instantiated in virtue of token mental representations, and the intentional properties of the former are instantiated in virtue of the intentional properties of the latter. If you endorse the representational theory mind, this is the general conception of phenomenal acquaintance you want to work with, and our discussion so far has been framed in terms of it.

Perhaps you reject the representational theory of mind. In this case, you end up with the subjective conception of phenomenal acquaintance. According to the subjective conception, phenomenal acquaintance isn't couched in terms of mental representation; its relata are just a subject and a phenomenal character rather than a subject *cum* mental representation and a phenomenal character. Here you can retain talk of intentional mental states, as well as the idea that phenomenal acquaintance is such a state; it's just that you reject the idea that states of phenomenal acquaintance, *qua* intentional mental states, are to be understood in terms of mental representation.

I begin the next chapter by setting out (the early) Russell's conception of acquaintance. Here I say 'acquaintance' rather than 'phenomenal acquaintance' given Russell's claim that the objects of acquaintance include non-experiential items in addition to experiential ones. Then I return to the notion of phenomenal acquaintance and consider a version of the subjective conception inspired by Russell's discussion. But apart from the possibility that there are independent reasons to reject the representational theory of mind, why take the subjective conception seriously, especially given that our discussion of phenomenal acquaintance so far has been framed in terms of the representational conception? I think the answer is pretty clear – it seems that the representational conception of phenomenal acquaintance has gotten us into trouble! In the previous section I argued that it's puzzling how phenomenal acquaintance could have the features of directness, thickness, and infallibility. But the directness, infallibility, and thickness puzzles, at least as they are currently formulated, are really puzzles for the representational conception of phenomenal acquaintance. The directness and infallibility puzzles, for example, concern the compatibility of the claim that phenomenal acquaintance is underwritten by attention understood as an operation on the contents of mental representations with the directness and infallibility of

phenomenal acquaintance. If we reject the idea you're acquainted with a phenomenal character in virtue of standing in the attention relation to the content of a mental representation, these puzzles disappear, at least as they are currently formulated. So we should view what I do in the next chapter not only as an important historical discussion of our topic but perhaps as a way out of the problems I've raised in this chapter.

CHAPTER 3

RUSSELL ON ACQUAINTANCE

3.1. Introduction

In the previous chapter I argued that the representational conception of phenomenal acquaintance seems to get us into trouble. In particular, I proposed three puzzles concerning how phenomenal acquaintance could be direct, thick, and infallible on the representational conception. We'll do well, then, to consider the alternative to the representational conception that I described at the end of the previous chapter, the subjective conception. In the first part of this chapter I set out the details of Russell's conception of acquaintance, a version of the subjective conception. I say 'acquaintance' here rather than 'phenomenal acquaintance' given Russell's claim that the objects of acquaintance include non-experiential in addition to experiential items. First, I isolate what I take to be the two central features of Russellian acquaintance, its directness and infallibility, though these features are understood slightly differently from how I characterized them in the previous chapter. Second, I review Russell's conception of the objects of acquaintance. Third, I argue that by 'acquaintance' Russell actually has two different phenomena in mind, one quite specific in nature and one more general. Here it will become clear that the former is what is of interest to us. In the second part of the chapter, I evaluate Russell's (specific) account of acquaintance. I argue that, though his account may not face two of the three problems I raised for the representational conception, it's no better off than its rival with respect to the third. My ultimate conclusion is that it's a bad idea to try to resolve our three puzzles by excising mental representation from phenomenal acquaintance. Instead, we should return to the

representational conception and see if we can resolve them within the context of the representational theory of mind.

3.2. Two Features

Russell's conception of acquaintance, as well as discussions of Russellian acquaintance by commentators, tends to focus on its epistemic rather than metaphysical status. Russell (1914b) suggests, for example, that the special importance of the objects of acquaintance is largely epistemological rather than metaphysical. Commentators tend to focus in particular on his distinction between knowledge by acquaintance and knowledge by description. As a result, what Russellian acquaintance itself is supposed to be, considered independently from its relation to knowledge, hasn't been made satisfactorily clear. I will therefore focus primarily on the metaphysical status of Russellian acquaintance and put the distinction between knowledge by acquaintance and knowledge by description to the side. The idea that we should focus, at least to begin with, on its metaphysical rather than epistemic dimension is well motivated, recalling my discussion of this issue from the first chapter.

Russell writes, "I do not think that, when [I'm acquainted with an object], there is in my mind something which may be called an 'idea' of the object, the possession of which constitutes my [acquaintance with it]" (1913a, 22).[1] Here he claims that acquaintance isn't underwritten by ideas, and we can understand an idea as something like a mental representation. So for Russell you aren't acquainted with an item in virtue of standing in the attention relation to the content of a mental representation of that item, *contra* the representational conception of phenomenal acquaintance. But what more can we say about

[1] Cf. Russell (1911, 155-6).

Russellian acquaintance, other than that acquaintance in his estimation isn't underwritten by mental representation? Though Russell (1913a, 45) presents acquaintance as a primitive, we can extract two central features of it from his discussion. As I see it, these features – directness and infallibility – form the core of Russell's conception of acquaintance.[2]

Let's begin with the sense in which Russellian acquaintance is direct. Russell writes, "I say that I am *acquainted* with an object when I have a direct cognitive relation to that object, i.e. when I am directly aware of the object itself" (1911, 148). We are already familiar with the idea of cognitive immediacy from our discussion in the previous chapter of the directness of phenomenal acquaintance on the representational conception. As I read Russell, the directness of acquaintance for him is a function of the fact that acquaintance isn't couched in terms of mental representation. In other words, acquaintance is direct in the sense that it's a relation instantiated by subjects and certain objects rather than subjects *cum* mental representations and objects (cf. Russell (1913a, 35)).[3] According to Russell, if your acquaintance with an object were a matter of you possessing an idea of it (i.e. tokening a mental representation that represents the object in question), the idea would function as a cognitive intermediary between you and the object. Experiential introspection, Russell's claims, suggests that our cognitive access to ordinary physical objects like tables is so mediated, but not our cognitive access to the objects of acquaintance. For Russell, the objects of acquaintance are literally before the mind in thought, while ordinary physical

[2] One of Russell's interesting claims about acquaintance that I won't consider in the main text is this: he claims that there is a sense of 'naming' in which you can name only objects of acquaintance. Here he seems to have something like demonstrative reference in mind: "during the process of naming [the objects of acquaintance, they] appear merely as this, that, and the other" (1913a, 39). Cf. Russell (1913a, 48).

[3] Russell stays neutral on whether subjects – the things that experience things – persist through time. In Russell (1913b, 184 and 1918, 252) he claims that acquaintance is a relation between *acts* (rather than subjects) and certain objects. We'll ignore this complication.

objects are available to thought only by proxy, merely as whatever objects satisfy various descriptive conditions (i.e. whatever objects are picked out by our ideas).

As I presented Russell's position above, he thinks that phenomenological considerations support the idea that acquaintance is direct, and to say that acquaintance is direct is to say that it isn't underwritten by mental representation. Hence, he seems to think that phenomenological considerations support the idea that acquaintance isn't underwritten by mental representation, so in Russell's eyes arguments for the claim that acquaintance does involve mental representation are more theoretical in nature. Russell (1913a, 42) critiques two such arguments. The first is that acquaintance is underwritten by mental representation because we can be acquainted with objects that don't exist (e.g. when we think about unicorns). According to the second argument, the fact that one and the same object (e.g. an ordinary physical object) can present itself in experience in different ways gets us to the same conclusion. In response to the first argument, Russell claims that acquaintance is object involving in a sense I specify below. In response to the second, he seems to deny that an object of acquaintance can have different modes of presentation. In other words, if you're acquainted with *x*, there is no possible world in which *x* presents itself to you (or anyone else) in experience differently than it in fact does. Russell's claim here (understood as applying to sense data – more on sense data below) is reminiscent of Descartes' discussion of appearances in the *Meditations* and seems to prefigure Kripke's (1980) dictum that there is no appearance/reality distinction for phenomenal character, as well as Loar's (1997) claim that a phenomenal character serves as its own mode of presentation.

Another way of getting at the directness of Russellian acquaintance involves his discussion of the difference between the cognitive operations of presentation and judgment. Russell claims that acquaintance is "presentational" rather than judgmental in nature:

> When I speak of [acquaintance as a cognitive relation], I do not mean the sort of relation which constitutes judgment, but the sort which constitutes presentation. In fact, I think the relation of subject and object which I call acquaintance is simply the converse of the relation of object and subject which constitutes presentation. That is, to say that S has acquaintance with O is essentially the same thing as to say that O is presented to S (1911, 148).

When you judge that something is the case, there is cognitive distance between yourself and the object of your judgment; your cognitive contact with the object is mediated in the sense that it's cognitively present only as the conclusion of an inference. To put the point in terms of mental representation, if all thought is underwritten by mental representation, all thought is mediated in the sense that the only cognitive purchase we have on objects (be they experiences, ordinary physical objects, universals, or whatever) are as the contents of mental representations. We never, as it were, come into direct cognitive contact with the objects themselves. Presentation, for Russell, is different from judgment in this respect. When an object is presented in thought rather than introduced as the conclusion of an inference (in other words, when it's cognitively available in virtue of presentation rather than judgment), it's genuinely cognitively present. Acquaintance, according to Russell, is presentational rather than judgmental: "…we have acquaintance… without the intermediary of any process of inference of any knowledge of truths" (1912, 46).[4]

Now let's turn to the infallibility of Russellian acquaintance. To begin, Russell writes, "If an object is given in acquaintance, then that object has a certain relation to the subject which is acquainted with it. But this would be meaningless if there were no such object" (1913a, 48). Here the idea is that, if *x* is an object of acquaintance, *x* exists; there are no states of acquaintance with non-existent objects: "…What is called the unreality of an immediate object must always be the unreality of some other object inferred from the

[4] Cf. Russell (1913a, 39).

immediate object and described by reference to it" (1913a, 49). Call this the 'object involving' nature of acquaintance that I mentioned earlier. Compare acquaintance in this respect to belief: it's certainly possible for you believe, to use Russell's famous example, that the present king of France is bald, while there is no actual present king of France. Beliefs, we will suppose, are *de dicto* relations – they're cognitive relations to objects expressed by that-clauses, i.e. propositions – while Russellian acquaintance is a *de re* relation – it's a relation to objects that can be constituents of propositions rather than propositions themselves.

Recall Descartes' famous claim in the *Meditations* that, though it might not be true that he is seeing, hearing, and feeling what he thinks he is, he can be certain that he seems to see, to hear, and to be warmed. Russell is in broad agreement with this claim; in addition to claiming that acquaintance is object involving, he seems to endorse something like the following: if, in your acquaintance with an object, that object is presented as an *F*, then it is an *F*.[5] If, for example, an object is presented to you as a reddish sense datum, it is a reddish sense datum (cf. Russell 1913a, 49 and 1914a, 93). While judgment can lead us into error, presentation never leads us astray.[6]

For Russell, if you're acquainted with *x*, is it possible for you to fail to believe that *x* exists, or to believe that *x* doesn't exist? I think Russell would claim that both are possibilities. He claims that mental states are wholly distinct from their objects in the sense

[5] Notice that this principle isn't entailed by Russell's claim we reviewed earlier that the objects of acquaintance don't have multiple modes of presentation. It could be, for example, that an object of acquaintance *A* has a unique mode of presentation *M*, but *M* presents *A* as being an *F* when in fact *A* isn't an *F*.

[6] Russell also claims that the inferential process of *analysis* can't lead us into error, in particular the analysis of the structure of complex objects of acquaintance. See Russell and Whitehead (1910, 43).

that if *x* is an object of acquaintance, your state of acquaintance isn't a part of *x* or *vice versa*.[7] Presumably he would say the same thing about the objects of acquaintance and any beliefs you have about them. Russell (1914a, 81 and 1914b) claims that *x* is logically dependent on *y* just in case *x* is a part of *y*, so it seems that he would say that it's logically possible to be acquainted with *x* and either fail to believe that *x* exists or believe that *x* doesn't exist.[8]

3.3. The Objects of Acquaintance

Russell claims, "The object [of acquaintance] may be in the present, the past, or not in time at all; it may be a sensible particular, a universal, or an abstract logical fact" (1913a, 5). Here Russell describes the possible objects of acquaintance. Notice that Russell's core conception of acquaintance as characterized above doesn't appeal to the nature of its possible objects.[9] This underscores the idea that, for Russell, the nature of acquaintance as a cognitive relation and considerations regarding the nature of its objects are largely independent.[10] Though I'm ultimately more interested in the directness and infallibility of

---

[7] Later Russell (1921, 142) changes his mind about this. Having decided that there are no subjects of experience (or at least their postulation is unwarranted), he claims that the distinction between acquaintance and its objects is no longer tenable. In reporting Russell's views on acquaintance, I'm concerned with his pre-1921 conception. And in reporting his views on the relationship between the objects of acquaintance and ordinary physical objects like tables, I'm concerned with his pre-phenomenalist views on the matter. See Russell (1914a, 96ff and 1914b) for his first articulations of phenomenalism.

[8] For informative discussions of Russellian acquaintance, see Pears (1969, 1981), White (1981), and Savage (1989).

[9] As we saw above, Russell does claim that the objects of acquaintance have unique modes of presentation. But this is part of his response to an argument for the claim that acquaintance is underwritten by mental representation rather than part of his core conception of acquaintance.

[10] Sainsbury (1986), for example, claims that for Russell acquaintance can (and should) be considered separately from the nature of its objects.

Russellian acquaintance than Russell's views on the objects of acquaintance, in this section I discuss the latter. For one thing, Russell says much more about the latter than the former, so we have more material to work with. It will emerge from the foregoing discussion in the next section that Russell's use of the term 'acquaintance' is ambiguous in that it has something like a loose and strict sense. Sometimes he uses the term to mean something like 'cognitive relation' (the loose sense) while at other times he has in mind a cognitive relation marked by directness and infallibility as characterized above (the strict sense).

Russell claims that there are three sorts of sensible particular with which we're acquainted: sense data, and what I'll call 'introspection data' and 'memory data'. Let's begin with Russell's infamous sense data, the particulars with which you're acquainted via sensation. He claims that a sense datum isn't "the whole of what is given in sense at one time" but rather "a part of the whole as might be singled out by attention" (1914b, 6). This is reminiscent of the distinction between *a* phenomenal character and *the* phenomenal character of an experience I discussed at the beginning of the first chapter. There I claimed that, in attending to an experience, you thereby grasp particular phenomenal characters of your experience.

Russell claims, "…In the presence of my table I am acquainted with the sense-data that make up the appearance of my table – its colour, shape, hardness, smoothness, etc." (1912, 47). So there is an overall way, according to Russell, that an object appears in sensation, and there are specific and determinate ways the object appears that jointly comprise its overall appearance. These constituents of the object's overall appearance in sensation, so the idea goes, are sense data, those particulars with which we're acquainted via sensation. The claim that sense data are appearances of objects should sound familiar, recalling the phenomenal character as appearance thesis I appealed to in the first chapter in

our discussion of experiential transparency. According to this thesis, the phenomenal character *reddishness* is a non-experiential aspect of the world appearing red. On this conception of phenomenal character, it's natural to think of *reddishness* as a dyadic relation whose relata are a subject and a non-experiential aspect of the world. Like the phenomenal character as appearance model, Russell claims that a reddish sense datum is part of an appearance of a non-experiential aspect of the world, but he doesn't claim that it's relational in character; instead, a sense datum is an object in its own right. So sense data, themselves appearances of objects, are presented to the mind in sensation. As such, they are cognitively present in a manner that the objects they are appearances of are not.

If x is a sense datum, is being an object of acquaintance an identity condition for x? In other words, are there sense data with which we're not acquainted? Russell (1914b, 7) defines 'sensibilia' as those particulars that would be sense data were they presented to us in sensation. Sense data are trivially sensibilia, but "do sensibilia which are data [to the senses] at a certain time sometimes continue to exist at times when they are not data?" (1914b, 8) In other words, are there what I'll call 'mere sensibilia', sensibilia that aren't currently presented to us in sensation, in addition to sense data? Russell claims that sense data "…probably never persist unchanged after ceasing to be data [to the senses]" (1914b, 8).[11] He claims that their unchanged persistence is a logical possibility, but the denial of unchanged persistence is "a probable inference from empirically ascertained causal laws…" (1914b, 9). To persist unchanged is one thing, while persisting *simpliciter* is another. But given how Russell formulates the question quoted above, I presume he means to say that it's probable, though not logically necessary, that particulars that are presented to us in sensation fail to persist *simpliciter* after they no longer are so presented. I presume that Russell likewise thinks that it's

[11] Cf. Russell (1913b).

improbable that there are particulars that are mere sensibilia at one moment only later to be presented to us via sensation. Hence, Russell's view seems to be that it's probable that there are no mere sensibilia, and, if there are none, this is a contingent matter of fact (cf. Russell (1914a, 95)).[12]

This is to be contrasted with the objects of phenomenal acquaintance as I characterized them in the first chapter. There I claimed that they are particular phenomenal characters of our experiences, and we aren't acquainted with each phenomenal character of each of our experiences. On this conception of acquaintance, then, there are mere sensibilia in the sense that there are phenomenal properties of our experiences with which we aren't acquainted.

What else can we say about Russell's conception of sense data? First, Russell (1914a, 80) claims that sense data, when complex, have spatial relations, so in addition to being concrete they're spatial as well. Second, he (1914b, 1915) claims that they're physical and non-mental. Not mental? Russell anticipates suspicion of this claim. He suggests that you might think that sense data are mental because they're subjective in some sense or other. In response, Russell (1914b, 7) claims that they're subjective only in the "physiological" sense that they're causally dependent on our sense organs, nerves, and brains (cf. Russell (1914a, 71)). The causal dependence of sense data on the physical states of the brain presumably explains why for Russell there (probably) are no mere-sensibilia as a contingent matter of fact. He seems to suggest that, were sense data logically dependent on the physical properties

[12] Russell's phenomenalist program articulated in Russell (1914a, 96ff and 1914b), however, relies on the existence of mere sensibilia, for he defines ordinary physical objects in terms of sense data and mere sensibilia. He claims, however, that mere sensibilia ("appearance… in a place where no sense organs and nervous structure exist") should be viewed as "hypothetical scaffolding" (1914b, 13). Russell reasons that if phenomenalism is true, sense data are the ultimate subject matter of physics and are therefore physical in this sense. We will consider the matter of whether sense data are mental or physical shortly.

of minded creatures, we would have grounds to claim that they're mental. But given that mere sensibilia are logically possible, it would be a mistake to characterize sense data as mental.[13]

Before we continue, it's worth noting that Russell's views on the notions of subjectivity and the relation between necessity and the mental/physical distinction are at odds with how many of us think of these issues today. With respect to the former, Levine (2001), for example, argues that the *sine qua non* of the experiential is its subjectivity, and the subjectivity of experience – the sense in which an experience is *for* a subject – isn't to be captured by simple causal dependence. The movement of my blood causally depends on the operation of my heart, but the movement of my blood isn't for my heart in the way that, for example, my experience as of a coffee cup on the table is for me. With respect to the latter, the contemporary philosopher of mind would draw a different conclusion from the claim that sense data (assuming they exist) are logically dependent on the physical properties of minded creatures. She would have already agreed that sense data are mental, and conclude that their dependence on the physical renders them physical as well.

Getting back to Russell, he has less to say about the objects of acquaintance via introspection and memory. The objects with which we're acquainted via introspection are

[13] Additionally, Russell suggests that you might think that sense data are mental because they're private in the sense that, if you're acquainted with *x*, no one else is (1914a, 95). But he seems to claim that the privacy of sense data, like the coextension of sense data with certain physical properties of minded creatures, isn't logically necessary; that is, there are logically possible worlds in which sense data are public in the manner that, say, universals are in every world. He claims there is "no good reason" to think that a particular sense datum is ever sensated by more than one subject, but not that such a situation is logically impossible (1913b, 187). In line with what I said above about logical necessity and mentality, he seems to think that, were it logically necessary that sense data are private, we would have a good reason to think that sense data are mental in nature.

what I called 'introspection data' above.[14] For Russell, acquaintance via introspection involves second-order (and perhaps higher-orders of) acquaintance, and when you're acquainted with being acquainted with something, he claims that you're "self-conscious" (1912, 49).[15] Suppose, for example, that you're acquainted with a sense datum *x*. Russell's suggestion is that via sensation you came to be acquainted with *x*, and via introspection you can come to be acquainted with your state of being acquainted with *x*. Russell also claims that through introspection you become acquainted with "thoughts, feelings, desires, etc." (1912, 51).[16] Assuming that all acquaintance via introspection is second-order acquaintance, the idea here is that you can become acquainted with being acquainted with a thought, you can become acquainted with being acquainted with a feeling, and so on. Presumably the basis of your first-order acquaintance with thoughts, desires, and so on is sensation.

Turning to the objects of acquaintance via memory, Russell distinguishes between what he calls 'immediate memory' and what I'll call 'mediate memory' (1913a, 57). In the case of mediate memory, you're acquainted with a mental image (presumably via sensation) and judge that it resembles a past sense datum or introspection datum. In this case, you aren't acquainted with the past sense datum or introspection datum itself. He claims that in the case of immediate memory you're acquainted with a past sense datum or introspection datum, claiming that "…remembered objects sometimes – at least in the case of the

---

[14] Russell claims, "…A place must be found for what appears as an experience of our own experiencing, since it is hard to see how otherwise we should have arrived at the notion that we have experiences" (1913a, 33). Cf. Russell (1912, 49).

[15] In Russell 1911 he stays neutral on whether if you're acquainted with being acquainted with *x*, you thereby are acquainted with yourself. Later, in Russell (1912, 51), however, he claims that in this case you "probably" are acquainted with yourself. But then later Russell (1913a, 36ff), he's skeptical again. And finally in Russell (1921) he rejects the existence of the subject!

[16] Cf. Russell (1940, 205).

immediate past – are still experiences, so that the objects experienced are not necessarily contemporaneous with the experiencing" (1913a, 33)[17] and, more poetically, "…there is our awareness of the immediate past, the short period during which the warmth of sensation gradually dies out of receding objects, as if we saw them under a fading light" (1913a, 72).[18] A past sense datum or introspection datum with which you're acquainted via immediate memory is what I called a 'memory datum' above. I take it that Russell above is describing the kind of memory of past views of an object that allow you, for example, to integrate these views into an experience of a persisting object over time.[19]

Considering Russell's view on acquaintance with logical facts would take us too far afield (his view, very roughly, is that we can be acquainted with the logical form of propositions), but here is Russell's take on acquaintance with universals in his own words:

> …In addition to [our acquaintance with sensible particulars] we have also (though not in quite the same way) what may be called awareness of *universals*. Awareness of universals is called *conceiving*, and a universal of which we are

---

[17] Russell states that "…We may define [immediate memory] as a two-term relation of subject and object, involving acquaintance, and such as to give rise to the knowledge that the object is in the past" (1913a, 70); cf. Russell (1912, 48ff).

[18] If you're acquainted with an object via immediate memory, you were recently acquainted with that object via sensation or introspection, but it is now "felt as past" (1913a, 73); it is "given in a different way from that in which it was given when it was a sense-datum… which makes us call it *past*…" (1913a, 72). Here it sounds like Russell is claiming that one and the same object of acquaintance can be given under different temporal modes of presentation. This is in conflict with the claim I attributed to Russell earlier that the objects of acquaintance have unique modes of presentation. This is an interesting issue, but I'll put it to the side for now.

[19] He might also have in mind after images, images that continue to appear in your visual field well after your exposure to the original stimulus has ceased. One of the most common afterimages is the bright glow that seems to float before your eyes after staring at a light bulb or a headlight for a few seconds. If Russell indeed has after images in mind here, it may be misleading to describe a memory datum as a past sense datum or introspection datum, for in this case you and the object of your acquaintance via immediate memory are concurrent. It's just that the object of your acquaintance serves as an appearance of an object situated like thus-and-so when that object is no longer so situated.

> aware is called a *concept*. Not only are we aware of particular yellows [which are sensible particulars], but if we have seen a sufficient number of yellows and have sufficient intelligence, we are aware of the universal *yellow*; this universal is the subject in such judgments as "yellow differs from blue"… And the universal yellow is the predicate in such judgments as "this is yellow," where "this" is a particular sense-datum (1911, 150).

The idea that we stand in a special cognitive relation to universals is an old one, going back at least as far as Plato's *Republic*. I'll say more about Russell's conception of acquaintance with universals in the next section.

3.4. Evaluating Russellian Acquaintance

Russell claims above that you can be acquainted with universals in addition to sensible particulars but "not in quite the same way". What is the difference here supposed to be? An obvious difference concerns the nature of their respective objects: in the case of sensation, introspection, and immediate memory, the objects of acquaintance are *concreta*, while in the case of conception the objects of acquaintance are *abstracta*.[20] For our purposes, however, a more important difference is that of the nature of your cognitive access to objects of acquaintance *vis-à-vis* sensation, introspection, and immediate memory on the one hand and conception on the other. Let me explain.

Recall Russell's claim that when you're acquainted with a sensible particular either via sensation, introspection, or immediate memory, you don't grasp it as the object of an inference; its cognitive availability is a matter of presentation rather than judgment. In this sense your acquaintance with a sensible particular for Russell, though cognitive in a nature, isn't properly viewed as a cognitive *achievement*, for it's the sensible particular, as it were, that is doing the work; instead of you seeking it out by inference, it somehow avails itself to you,

[20] Moreover, a universal, unlike a sensible particular, is public in that more than one subject can be acquainted with it. See note 13.

as if by its own accord. In capturing this idea, Russell (1914a, 35) claims that neither intellect nor intuition are involved in your acquaintance with sensible particulars.

In this respect Russell seems to treat universals *qua* objects of acquaintance differently. In the passage quoted above, Russell seems to suggest that you're acquainted with *F*-ness in virtue of *F*-ness figuring in the result of an inference. Here the idea is that you look at some *F*'s, judge that they're similar in some respect, and then judge that there must be something, *F*-ness, that accounts for the similarity. The proposal, then, seems to be that you grasp *F*-ness as that which is common between various objects. This, of course, is the familiar story we get from Plato, and it seems that Russell's treatment of acquaintance with universals is merely a recapitulation of it. Russell's conception of acquaintance with universals so characterized obviously doesn't fit with the claim that acquaintance is direct in the sense specified earlier. What is going on here?

To address this issue, we need to turn to Russell's motivation for positing the acquaintance relation in the first place, something I touched on earlier. As I see it, Russell has at least four largely independent reasons for doing so. The first consideration, nicely articulated in the *Problems of Philosophy*, is phenomenological in nature. Russell points out that when you attend to, say, a patch of red in our visual field, it seems that the object of your thought enjoys a special cognitive immediacy; it's "given" or cognitively present in a way that other sorts of objects, for example, tables, aren't. This cognitive presence is what I've been calling the 'directness' of Russellian acquaintance. We can look at this cognitive immediacy as a datum to be examined and explained, and Russell's conception of acquaintance (of sensible particulars) as a characterization of it. This should all be familiar from our earlier discussion of Russell on the phenomenology of the directness of acquaintance.

The second motivation concerns understanding. Here Russell claims that in order to understand a proposition, you must have a cognitive grasp of the constituents of the proposition; a cognitive grasp of this sort is a necessary condition for thought. He notes that we, of course, do understand various propositions, and he calls the cognitive grasp of their constituents, both of particulars and universals, 'acquaintance'. Russell claims, for example, "All cognitive relations – attention, sensation, memory, imagination, believing, disbelieving, etc. – presuppose acquaintance" (1913a, 5). These ideas are codified in his infamous acquaintance principle: to understand a proposition you must be acquainted with its constituents. This consideration, unlike the first one, is intended to show that we're acquainted with both sensible particulars and universals.

The third concerns perception. Here Russell is interested in providing an account of perception and thinks that something like acquaintance must be centrally featured in such an account. For Russell, we, strictly speaking, perceive only the objects of acquaintance rather than their distal causes.

The fourth motivation is epistemological in nature. Russell is part of the Cartesian tradition of internalist foundationalism in epistemology, and claims that only something like acquaintance is up to the task of playing the role of that which underwrites the self-evidence of our basic beliefs.

When we focus on phenomenological considerations, we get the conception of acquaintance that I developed earlier in the chapter. When we focus on the other considerations mentioned above – issues regarding understanding, perception, and epistemology – we end up with a more general conception of acquaintance, one that seems to leave the directness of acquaintance behind. Putting phenomenological considerations to the side, Russellian acquaintance seems to end up meaning something like 'being cognitively

related to *x*', where '*x*' ranges over particulars, universals, and logical facts. Given my discussion of phenomenal acquaintance in the last two chapters, it's clear that Russellian acquaintance in this general sense isn't my concern.

So let's return to Russell's discussion of acquaintance with universals. The general conception of acquaintance I outlined in previous chapters is this: to be acquainted with a phenomenal character is to attend to an experience that has it and thereby grasp that phenomenal character in thought. On this conception, the objects of acquaintance are phenomenal properties. Assuming, for the moment, that phenomenal properties are universals, it's natural to think that Russell's discussion of acquaintance with universals will have some bearing on this conception of phenomenal acquaintance. But not only does the feature of directness drop out of the picture in Russell's discussion of acquaintance with universals, but there is no sense in which acquaintance with universals is thick either, recalling our discussion of thickness from the previous chapter. For Russell seems to claim that, in your acquaintance with a universal, you grasp it as the commonality between various objects. This runs contrary to the determinate aspect of the thickness of phenomenal acquaintance, the fact that we grasp the objects of acquaintance as a specific qualities, identifiable in their own right, not merely by their relation to other qualities and objects.

How should we proceed from here? Here is what I propose we do. In what follows, let's focus our discussion on Russellian acquaintance with particulars, specifically sense data, putting to the side memory data and introspection data. So from now on we'll understand Russellian acquaintance as follows. The objects of acquaintance are experiential particulars (sense data) rather than phenomenal characters *qua* types as I suggested previously. Here acquaintance with a sense datum is direct in the sense that it's not a matter of your having a mental representation of that sense datum. We shall continue to maintain that acquaintance

is achieved by attention, but we'll have to drop the idea (for now, anyway) that attention is underwritten by mental representation. Acquaintance with a sense datum is infallible in the sense that if it's presented in experience as being a sense datum of a certain type (e.g. as a reddish sense datum), then it is in fact a sense datum of that type. And while Russell doesn't discuss the feature of thickness, an account of phenomenal acquaintance should include something like this feature. Now we're claiming that the objects of acquaintance are sense data rather than phenomenal properties themselves, so Russellian acquaintance isn't thick in the exact sense of the term developed in the previous chapter. Let's understand the thickness of Russellian acquaintance instead as follows: when you attend to an experience and thereby grasp a particular sense datum associated with it (e.g. you attend to your current visual experience and thereby grasp a reddish sense datum), you don't grasp it as a "know-not-what", and your grasp of it isn't relational in character. So Russellian acquaintance, we will suppose, has the features of directness, infallibility, and thickness, but these features are understood slightly differently from how we understood them in the previous chapter.

At the end of the last chapter I concluded that it seems that the representational conception of phenomenal acquaintance seems to get us into trouble. We should see, then, if we have any better luck with our rival version of the subjective conception, Russellian acquaintance, as set out above.[21]

---

[21] If you're uncomfortable with Russell's notion of a sense datum (as I think you should be!), you can replace it with the notion of an instance of a phenomenal property. Instances of phenomenal properties, like sense data, are particulars, so in switching one notion out for the other you would retain the idea that the objects of Russellian acquaintance are particulars rather than phenomenal characters *qua* types. (I have more to say about the nature of phenomenal property instances in the next chapter.) The objections that I'm going to make shortly against Russellian acquaintance apply to this modified view as well. My objections to Russell, therefore, don't proceed upon the details of his conception of sense data. This, in my estimation, is how things should be, given that Russell's core conception of acquaintance can be separated from his conception of the objects of acquaintance.

Below I consider how Russellian acquaintance sits with respect to the directness, infallibility, and thickness puzzles. If this account has the means to resolve these puzzles, I think we have a good reason to accept it. But before we turn to these issues, I would like to say something about the status of Russellian acquaintance as a cognitive relation. You might object to the Russell-inspired account as follows. Russell describes acquaintance as a cognitive relation, but, without recourse to mental representation, its status as a cognitive relation is unclear. For what is it about Russellian acquaintance that *is* cognitive; why does being acquainted with a sense datum in this case count as thinking about it? Acquaintance for Russell is a two-term relation, one that takes a subject and a sense datum as its relata. But it's easy to see, so the objection goes, that the fact that a subject and a sense datum are the relata of the relation alone isn't sufficient for the relation to count as a cognitive relation, as a form of thought. Consider, for example, the relation *x* co-exists-with *y*. I'm currently having an experience associated with a reddish sense datum, so I, *qua* subject, and the sense datum jointly instantiate the relation. This relation, however, clearly isn't cognitive in nature. So, by excising mental representation from acquaintance, it seems that we've lost our grip on its cognitive nature.

I have some sympathy with this objection, but I don't think that it's ultimately persuasive. As you will recall from previous chapters, in my view the representational theory of mind is the correct theory of cognition, so I think that cognition is in fact underwritten by mental representation. But there is certainly nothing incoherent about supposing that cognition isn't so underwritten, and the objection above seems to suggest otherwise. Remember that Russell presents acquaintance as a primitive cognitive relation. As such, acquaintance in his estimation has intentionality and other cognitive features built into it, and that's that. As I read McDowell (1984, 1986), he makes a similar claim. He appeals to the

notion of a *de re* sense in an attempt to understand acquaintance as a cognitive relation without recourse to mental representation, and he seems to take the notion of a *de re* sense as primitive. While such accounts may be unsatisfying, there is nothing incoherent about them. Moreover, there are non-primitivist alternative theories of cognition that excise mental representation from cognition altogether, not just from acquaintance. Advocates of connectionism about cognition can be read as excising mental representation from cognition (cf. Van Gelder (1995)). And proponents of the enactive approach to cognition argue that perception and the grasping of experiential features in introspection are cognitive activities, yet they don't (or at least needn't) involve mental representation (cf. Noë (2004, 2009)).

So putting this objection to the side, let's turn to the three puzzles I set out in the previous chapter. The first two puzzles ostensibly concern the compatibility of the directness and infallibility of phenomenal acquaintance with the claim that it's underwritten by mental representation. How can phenomenal acquaintance be direct when it's mediated by mental representation (the directness puzzle)? Why is phenomenal acquaintance infallible when it seems that we have no reason to believe that the mental representations that underwrite it can't misrepresent (the infallibility puzzle)?

As I see it, the chief virtue of Russellian acquaintance is that it gives us a straightforward resolution of the directness puzzle. Russell, as we have seen, claims that acquaintance with particulars is direct in the sense that it isn't underwritten by mental representation. It is therefore obvious that the proponent of Russellian acquaintance doesn't face the problem of having to reconcile the directness of acquaintance with it being mediated by mental representation. The infallibility puzzle may also lose its bite. If it's not the case that to be acquainted with a sense datum is to stand in the attention relation to the content of a mental representation that represents it as falling under a certain type, we obviously

don't have to worry about why it should be that the mental representation in question couldn't misrepresent. With respect to these puzzles, then, Russellian acquaintance is attractive.

What about the third puzzle? As I set out the thickness puzzle in the previous chapter, the problem is this: How can phenomenal acquaintance be thick when its directness seems to rule out the possibility that to be acquainted with a particular phenomenal character is to stand in the attention relation to the content of a mental representation, one that represents an experience as having that phenomenal character? You might think that this puzzle doesn't apply to Russellian acquaintance because it isn't underwritten by mental representation and it isn't thick in the sense specified in the previous chapter. Things aren't so simple, however.

The problem for the representational conception, you will recall, is this. Suppose you attend to your visual experience as of black letters on a white page, and you come to grasp the blackish character of your experience. But how is it that in attending to this experience you came to grasp one particular phenomenal character rather than another? If, in attending to your experience, you stand in the attention relation to the content of a mental representation, one that represents your experience as having a blackish character, we would have an explanation. But the directness of phenomenal acquaintance seems to rule this explanation out, for to postulate such a mental representation seems to introduce objectionable mediating material. So now suppose that phenomenal acquaintance isn't underwritten by mental representation, its objects are sense data rather than phenomenal properties, and phenomenal acquaintance is achieved by attending to experiences. Unfortunately, we're left with the same kind of problem: how is that in attending to your experience you grasp one sense datum rather than another, since any experience is associated

with various sense data? Here we face a similar underdetermination problem. The only way I can think of to answer this question is to appeal precisely to what is ruled out by the account, *viz.* mental representation. Here we would say that, in attending to your experience, you token a mental representation, one that represents your experience as being associated with a particular sense datum. Without recourse to mental representation, we can say that the sense datum in question presents itself to you, but this, rather than an explanation, is just a restatement of the problem.

The advocate of Russellian acquaintance can, of course, say that this too is a primitive matter – there just is no illuminating answer to the question of how it is that we come to be acquainted with sense data in response to the underdetermination worry; we just manage to do so, and that's that. And, assuming that acquaintance is a primitive cognitive relation, I suppose that we shouldn't be surprised that there might not be an illuminating answer to this problem.

In considering Russellian acquaintance I'm willing to grant for the sake of argument that the status of acquaintance as a cognitive relation is brute matter. I'm not willing to grant, however, that how we come to stand in the acquaintance relation is a brute matter. As I've set things up in this dissertation, an adequate account of phenomenal acquaintance must explain how it is that phenomenal acquaintance could be direct, thick, and infallible, and, as should be clear from above, I've left open the possibility that different accounts of phenomenal acquaintance might understand these features slightly differently. For my purposes here, resolving the puzzles that correspond to these features would discharge this explanatory burden. Central to the thickness puzzle is the concern about underdetermination: how is it that attention to experience results in acquaintance, be the objects of acquaintance sense data or phenomenal properties? To fail to resolve this concern

is to fail to resolve the thickness puzzle, which in turn is to fail to explain how phenomenal acquaintance is thick. If I'm right in thinking that Russellian acquaintance doesn't have the resources to resolve the underdetermination concern, it fails as an account of phenomenal acquaintance, at least on my terms.

3.5. Conclusion

To sum things up, we can think of the Russell-inspired account of phenomenal acquaintance as facing the following dilemma. In order to capture the directness and infallibility of phenomenal acquaintance, we need to think of Russellian acquaintance as applying just to particulars, specifically sense data. In doing so, however, it seems that we fail to come to terms with the underdetermination problem at issue in the thickness puzzle, so we lose our grip on the thickness of phenomenal acquaintance. We might try to resolve the underdetermination problem by focusing on Russell's discussion of acquaintance with universals. Perhaps you come to grasp a particular sense datum partly in virtue of your acquaintance with a universal the sense datum instantiates. But if we go in this direction, however, it seems that we lose our grip on the directness and infallibility of phenomenal acquaintance. So either way we go here, it seems that we fail to capture all of the key features of phenomenal acquaintance.

The moral I'm inclined to draw from the apparent failure of the Russell-inspired account of phenomenal acquaintance is that we were too quick to abandon the representational conception. In the second chapter I argued that our appeal to mental representation got is into trouble, but in this chapter we've seen, paradoxically, that apparently we need to appeal to it in giving an account of phenomenal acquaintance (specifically in coming to terms with the thickness of phenomenal acquaintance). So though

we didn't succeed in coming upon a viable account of phenomenal acquaintance in this chapter, our discussion has been instructive, at least in that we've come to see that mental representation seems to play an essential role in phenomenal acquaintance. In the next chapter I return to the representational conception and see if we can resolve the puzzles of directness, infallibility, and thickness. I argue that we can.

CHAPTER 4

AN ACCOUNT OF PHENOMENAL ACQUAINTANCE

4.1. Introduction

In the second chapter I focused on the representational conception according to which phenomenal acquaintance is underwritten by mental representation, and I posed three problems for it, the directness, infallibility, and thickness puzzles. Having concluded that the representational conception seems to get us into trouble, in the third chapter I considered a version of an alternative to the representational conception – the subjective conception – inspired by Russell's discussion of acquaintance. There I argued, however, that, Russellian acquaintance fails to resolve the thickness puzzle. I concluded that the lesson to take away from our discussion of Russell is that we were too quick to abandon the representational conception of phenomenal acquaintance. In this chapter I develop a version of the representational conception that resolves our three puzzles, and I defend it against various objections. In the course of defending it against these objections I further develop the proposal.

4.2. The Account

My account of phenomenal acquaintance consists of three theses. The first thesis appeals to a distinction between token states and instances of the properties those states instantiate, so let me begin by saying something about this distinction. To begin, consider the type/token distinction applied to phenomenal properties (characters). Phenomenal properties are types, and their instances are tokens. So there is a distinction, for example, between the property *painfulness* (*qua* type) and an instance of *painfulness* (*qua* token). I take the

notion of a property instance as a primitive, although a property instance can be modeled as an ordered triple with slots for objects, properties, and times. So, supposing that you're in pain at a time *t*, we can model the instance of *painfulness* that occurs at *t* as <You, *painfulness*, *t*>.[1]

Though I take the notion of a property instance as primitive, I do wish, however, to endorse a substantive thesis about the nature of token states. In my view, token states are complexes of property instances and themselves property instances. Let me explain. Suppose that a token state *s* occurs at *t*. On my view, this means that some property *P* is instantiated at *t*, and the *P*-instance occurring at *t* is an improper part of *s*. I claim that *s*, *qua* token state, is complex in the sense that it has further property instances as constituents, as proper parts. But which instances? My proposal is this: There is some property *Q* such that an instance of *Q* is a proper part of a property instance *p* (*qua* token state) just in case *p* instantiates *Q*. In other words, the constituents of *p* are instances of the properties *p* instantiates.[2] Let's consider some examples. Suppose that you're having an experience as of a brown table at *t*. My claim is that in this case your token perceptual state is an instance of something like the property *having a perception as of a brown table*, and any property instantiated by this token state at *t*, e.g. *being brownish*, is such that an instance of it partially composes the perceptual state

[1] Kim (1976) claims that events are such ordered triples. Here Kim offers an *analysis* of what it is to be an event, while I merely claim that property instances may be usefully described as such. See Lewis (1986) for objections to Kim's claim *qua* analysis but not description of events. Shoemaker (2007) treats property instances roughly as I do above.

[2] You might think of this proposal as amounting to an account of what it is for a token state to instantiate a property. The thrust of such an account would be that the instantiation of properties by token states is to be ultimately understood in mereological terms. I wish to officially stay neutral, however, on whether we should understand the above proposal as providing such an account. I also wish to stay neutral on whether property instances and objects stand in the part-whole relation. Nothing that I go on to do in this chapter requires taking a stand on either issue.

token at that time. So, if *P* is the former property and *Q* the latter, your token perceptual state is a *P*-instance that has a *Q*-instance as a proper part. Now suppose that you're happily digesting your dinner at *t*. In this case, a token state of digestion *d* occurs at *t*, and *d* is an instance of something like the property *digesting your dinner*. *d* instantiates the property *being so as to release gastric acid* at *t*, and an instance of this property partially composes the digestive state token at that time.[3] So, just as, for example, a chair is an object and has further objects as proper parts (legs, arms, etc.), a token state is a property instance that has further property instances as proper parts.

Having made clear how I view the relationship between property instances and token states, here is the first thesis in my account of phenomenal acquaintance:

> *Thesis 1*: Token experiential states themselves are property instances, and they have instances of the phenomenal characters they instantiate as proper parts.

Let's now move on to the second thesis. It has two parts, and each concerns instances of phenomenal properties. Here is the first part:

> *Thesis 2a*: Instances of phenomenal properties are token mental representations.

By endorsing the representational theory of mind, I'm obviously committed to the existence of mental representations. My claim is that, among the token mental representations that constitute your cognitive economy when you're having an experience, some of these mental representations are instances of those phenomenal properties instantiated by your experience. So consider, for example, the phenomenal property *blackishness*. The claim is that if *b* is an instance of *blackishness*, *b* is a token of a particular mental representation.

---

[3] Intuitively, a token physical state is one that has only instances of physical properties as proper parts, while a token mental state is one that has at least one instance of a mental property as a proper part.

What do these special mental representations represent? Enter the second part of the second thesis in my account of phenomenal acquaintance:

Thesis 2b: Instances of phenomenal properties represent themselves.

I claim, in other words, that instances of phenomenal properties are self-representational. Instances of phenomenal properties are self-representational, I claim, in the following sense: if p is an instance of a phenomenal property, then p is a (token) mental representation, one that represents something x as being a particular property instance q, where $x = p = q$. In short, instances of phenomenal properties, *qua* token mental representations, are identical to what they represent. In this case there is no distinction between representational vehicle – the thing that does the representing – and that which is represented.

To get a better handle on the second part of the second thesis in my account of phenomenal acquaintance, let's begin by considering a general feature of the representation of property instances. To represent something x as being property instance p is (at least) to represent x as an instance of the property P that p is an instance of. If this weren't the case, it would be unclear what it would be to represent one property instance rather than another. So to represent x as being property instance p – where p, say, is an instance of the property *scarletness* – is to represent x as an instance of *scarletness*. Regarding the individuation of property instances, I take it that, if p is an instance of P, then, for all Q, if p is an instance of Q, $P = Q$. In other words, a property instance is an instance of only one property. So to represent x as being p where p is an instance of *scarletness*, you don't have to represent x as an instance of any property other than *scarletness*. Though it's true, for example, that anything that is scarlet is red, since p isn't an instance of *redness*, x needn't be represented as an instance of this property in order to be represented as being p. In fact, if x is p and x is represented as an instance of *redness*, we have a case of misrepresentation.

Putting together the condition of property instance representation described above and my characterization of the self-representational status of phenomenal property instances, we get the following schema: if *p* is an instance of a phenomenal property *P*, then *p* is a mental representation, one that represents itself as an instance of *P*. So if *p* is an instance of the phenomenal property *reddishness*, *p* is a mental representation, one that represents itself as an instance of *reddishness*. Returning to the idea that phenomenal characters are appearances, we can say that if *p* is an instance of *reddishness*, *p* represents itself as a particular occasion of a non-experiential aspect of the world appearing red.

I'll argue shortly that the claim that instances of phenomenal properties represent themselves is crucial to resolving our puzzles. But do we have independent reasons to think that this claim is true? I think so. Phenomenal acquaintance is a matter of attention or focal awareness, so, given that we aren't focally aware of each of our experiences, we aren't acquainted with each of their phenomenal characters. Contrast focal awareness of the experiential with peripheral awareness of it, recalling our discussion in the first chapter. Many think that we're peripherally aware of each of our experiences. And, again recalling our previous discussion, higher-order and same-order monitoring theorists (cf. Rosenthal (2004) and Kriegel (2006), respectively) make a stronger claim – they claim that awareness is at the very nature of experience, that an experience at minimum just is a mental state of which its subject is aware. On the assumption that awareness is underwritten by mental representation, they claim that experiences by their nature are accompanied by mental representations, ones that represent them. If instances of phenomenal properties represent themselves, it's clear why each experience is accompanied by a mental representation. So the first thesis of my

account of phenomenal acquaintance, in other words, is a natural way of explaining why experience and mental representation come as a package.[4]

We can make short work of the remaining thesis in my account of phenomenal acquaintance. Throughout the previous chapters I have claimed that acquaintance is a matter of attending to an experience and grasping a phenomenal character of it. The third thesis involves a qualification of this claim. I wish to claim now not that you attend to an experience *per se*; instead, you attend to a component of it, an instance of the phenomenal property you come to be acquainted with. Here's how the proposal works. Suppose you're acquainted with the blackish character of your current visual experience. How did this come to be? What you did was attend to a component of your experience, an instance of the very property you're now acquainted with.

Above I claimed that to attend to something is to stand in a certain relation to the content of a mental representation, one that plays the role characteristic of attention and represents the object in question. So what is the mental representation in virtue of which you attend to a component of your visual experience in our example above? Answer: the object of attention – an instance of the blackish character of your experience – and the mental representation in virtue of which you attend to it are one and the same. So in this case an instance of the blackish character constitutive of your experience plays "double duty" – it's both the item you're attending to and the mental representation that underwrites your state of attention. This, of course, is to be contrasted with how attention normally works. When you attend, say, to a horse, the horse and the mental representation of it in virtue of which

[4] Moreover, Sosa (2003), Levine (2006a), Horgan and Kriegel (2007) and others claim that in introspection the peripheral awareness of an experience isn't registered as an extra mental act; instead, it's registered as a feature internal to the experience. Claiming that experiences involve self-representation comports with this putative deliverance of experiential introspection.

you attend to it are certainly distinct; one is a mental item while the other is an ordinary physical object. So the third thesis in my account of phenomenal acquaintance, more fully stated is this:

> *Thesis 3*: The state of attention in virtue of which you're acquainted with a phenomenal character is implemented by you standing in the attention relation to the content of a certain mental representation, an instance of a phenomenal property, *qua* self-representation, which is the very object of attention.

My account of phenomenal acquaintance, then, as three components, one about the structure of experience (token experiences have instances of their phenomenal characters as components), one about self-representation (instances of phenomenal properties are self-representational), and one about the relation between self-representation and the attention relevant to phenomenal acquaintance (to become acquainted with a phenomenal character of your experience is to attend to instance of that phenomenal character constitutive of your experience, and the mental representation in virtue of which you attend to it and the object of attention are one and the same). With my account of phenomenal acquaintance on the table, I can now resolve the three puzzles set out in the second chapter, the directness, infallibility, and thickness puzzles. After resolving these problems, I consider various objections to my account of phenomenal acquaintance.

4.3. Resolving Our Puzzles

Let's begin with the directness puzzle. It should be clear now how attention can be underwritten by prior mental representation and phenomenal acquaintance can be direct. For, suppose that you attend to an instance of a phenomenal property (a component of a token experience that instantiates the property in question), and you thereby become acquainted with a phenomenal character, the phenomenal property the object of your

attention is an instance of. It's true that attending to this property instance requires that it already be the content of a mental representation. The reason why this is compatible with the directness of phenomenal acquaintance is because this mental representation and the object of your attention are one and the same, given that instances of phenomenal properties represent themselves as instances of the properties they're instances of. This is perfectly compatible with the directness of phenomenal acquaintance, for here it's not the case that your state of acquaintance inherits its content from some additional mental item, one distinct from the object of your attention. So, returning to our graphic representation of phenomenal acquaintance, our figure now looks like this, where the arrow pointing down means 'is an instance of':

$$\begin{array}{ccc} & & P \\ & \nearrow & \downarrow \\ \text{You} & \rightarrow & r \text{ (= a component of } e\text{)} \end{array}$$

Figure 3: Phenomenal Acquaintance

In this case, *r* doesn't mediate your connection to the object of attention because *r* just is the object of attention.

Let's now return to the infallibility puzzle. I claim that you're acquainted with a phenomenal character in virtue of standing in the attention relation to the content of an instance of that phenomenal character *qua* self-representation. Now, it's certainly possible to represent something x as being a property instance p when in fact x is a distinct property q. If p is an instance of property P and q is an instance of property Q, in this case x is represented as being an instance of P when in fact it's an instance of Q. But what *can't* happen, however, recalling our earlier discussion of the representation of property instances, is this: x is represented as being p, x is thereby represented as being an instance of P, but p isn't an instance of P. Such misrepresentation is impossible, for in this case you simply

wouldn't count as representing *x* as being *p* in the first place. It follows, then, that when *p* represents itself in the sense that its representational vehicle and what it represents are identical, it correctly represents itself as being an instance of *P*. Hence, instances of phenomenal properties, *qua* self-representations, not only represent themselves as instances of phenomenal properties, but they always do so correctly. It follows that the mental representations that underwrite the states of attention at issue with phenomenal acquaintance can't misrepresent, and this explains why the experiential judgments we make in virtue of our acquaintance with phenomenal character can't go wrong.

Finally, we return to the thickness puzzle. Above I argued that it's unclear how attending to a particular experience would confer a (non-relational) grasp of one phenomenal character rather than another, given that experiences instantiate many phenomenal properties and phenomenal acquaintance is direct. I should have said, however, that it's unclear how attending to an experience *in toto* would confer such a grasp. But suppose instead that the object of attention is an instance of a phenomenal property, a component of the experience that instantiates the property. In this case, it's no mystery how your attention confers a grasp of a particular phenomenal character, for an instance of a phenomenal property represents itself as an instance of the particular phenomenal property it's an instance of. In this case, there is only one candidate for the phenomenal property you grasp through attention. So if the objects of attention that underwrite states of phenomenal acquaintance, rather than experiences *per se*, are instances of phenomenal properties – self-representational components of these experiential states – the thickness puzzle is resolved.

I conclude that my three claims set out in the previous section, if true, resolve our three puzzles concerning phenomenal acquaintance on the representational conception. As I see it, the job of an account of phenomenal acquaintance is to resolve these puzzles, so I

conclude that my account of phenomenal acquaintance consisting of these claims is promising, if not plausible.

4.4. Objections

In this section I set out and respond to some potential objections to my account of phenomenal acquaintance. First, I consider a batch of objections concerning my claim that instances of phenomenal properties are self-representational. Considering these objections will take up the bulk of this section. Second, I respond to an objection concerning the directness of phenomenal acquaintance and my claim that attention to experience is underwritten by mental representation. Third, I consider an objection to my resolution of the thickness puzzle. In responding to these objections I further develop my account of phenomenal acquaintance.

I begin by considering some objections concerning my claim that instances of phenomenal properties are self-representational. I have argued that this claim helps resolve our three puzzles and that we have independent reasons to believe that it's true, so I think that I've made a reasonable case for it. You might, however, be suspicious of the very notion of self-representation. Such a concern, however, would be misguided. Consider, for example, the sentence "The font of this very sentence is Garamond". This sentence represents itself, and it's a veridical representation.[5] It would seem, then, that the notion of self-representation itself is perfectly respectable. So if there is a problem with the idea that instances of

[5] Notice that this sort of self-representation can go wrong. This is because in such cases there is a descriptive element to the content over and above the representational vehicle.

phenomenal properties represent themselves, the problem isn't with the notion of self-representation *per se*.[6]

As Kriegel (2005) points out, however, there is an important difference between self-representation in natural language and mental self-representation, self-representation in the language of thought. A sentence in a natural language represents itself only thanks to our interpretation of it; it is, following Grice (1957) and Searle (1980), a case of derivative intentionality. Our intentions, rather than the nature of the symbols and their concatenation *per se*, are that in virtue of which lexical items represent themselves. An instance of a phenomenal property, on the other hand, if self-representational, is self-representational in a non-derivative sense. It is, we can say, a case of 'original' self-representation. You might object, therefore, to the notion of original self-representation while admitting that derived self-representation is perhaps unproblematic.

What we need is a way to naturalize original self-representation. Following Kriegel (2003, 2005), it would seem that causal accounts of mental representation are of no help here, given that representational vehicles don't cause themselves.[7] I think the best option available here might be conceptual role semantics (CRS). According to CRS, we should characterize a system of representational vehicles *en masse* akin to the way we inter-define

---

[6] There are, of course, notorious paradoxes associated with self-representation. Considering this issue in any detail would take us too far off course. I will say, however, that I don't think that these problems throw into doubt the very idea that there are self-representational bits of natural language.

[7] Kriegel's (2003) view on this matter is in fact more nuanced than I presented it above. Let *e* be a token experience and *m* the token state of peripheral awareness that accompanies *e*. Kriegel points out that if *e* and *m* are identical, then causal accounts of mental representation fail to capture the self-representational status of *e* (*m*), since *e* (*m*) doesn't cause itself. He defends the claim that experiences are self-representational and is sympathetic with causal accounts of mental representation, so he considers the view that *e* and *m*, though distinct are proper parts of a common state *e**. In this case, *e* and *m* are intimately related, but their relation isn't so intimate that they can't stand in causal relations.

theoretical terms. Here's roughly how it works. Step 1: specify what the system's representational vehicles do and what they're responsible for. Step 2: from this description of the entire system, extract an individuating profile of each representational vehicle, what we'll call a 'conceptual role'. The conceptual role of a representational vehicle is typically thought to include the inferential relations it stands in to other representational vehicles, so the conceptual role of a representational vehicle is viewed as a semantic feature of it. (As will become clear below, I think that inferential role is just one aspect of the conceptual role of a representational vehicle. The idea is that there are things representational vehicles do and are responsible for in addition to standing in inferential relations to other representational vehicles.) Step 3: for each representational vehicle, specify those causal relations in virtue of which it has the conceptual role it does. Call the set of causal relations associated with the conceptual role of each representational vehicle its 'functional role'. According to CRS, the functional role of a representational vehicle determines its content in virtue of determining its conceptual role. In other words, functional role determines conceptual role, and conceptual role in turn determines content. The notion of determination at issue here is transitive, so functional role itself determines content. The overall proposal, then, is that we should understand content first in terms of a further semantic notion – conceptual role – and then understand the latter in terms of a syntactic notion – functional role.[8] So I'm interested in the claim that the functional role of the mental representations tokens of which are instances of phenomenal properties determine their self-representational nature in virtue of determining their conceptual role. It's because of the functional role of an instance of a phenomenal property that it represents itself.

[8] For influential defenses CRS see, for example, Field (1977), Loar (1981), Lycan (1984), and Block (1986).

What are the central features of those conceptual roles in virtue of which instances of phenomenal properties represent themselves? To address this matter I need to introduce the distinction between what I call 'constitutive self-representation' and 'non-constitutive self-representation'. Let's begin by doing a bit of reverse engineering. Suppose we have completed Step 2 in the recipe for content I outlined above in describing CRS. Suppose, in other words, that we have isolated the conceptual roles of each representational vehicle in a representational system. Imagine that at this stage we realize that *R* – a particular representational vehicle (type) in our system – plays a particular conceptual role *C*, one that determines that its tokens represent themselves. In other words, we learn that *R* is self-representational in virtue of its conceptual role. Now that we have *C* on the table, let's forget for the moment about *R* *qua* representational vehicle and ask ourselves the following question: What representational vehicles individuated syntactically *could* play *C*? Returning to *R*, we can put the question this way: is there anything about *C* that suggests that only *R* could play *C*, a conceptual role such that, if a representational vehicle plays it, its tokens represent what the tokens of *R* in fact represent? If there is a many-one relationship between possible representational vehicles and *C*, then the answer is 'no' – in this case there is, as it were, no road back to *R* *qua* representational vehicle from *C*. If there is no road back to *R* from *C* in this sense, I'll say that *C* is a 'promiscuous' conceptual role; it's the sort of conceptual role such that it's possible for different types of representational vehicle to play it. So suppose that *R*'s conceptual role is promiscuous. It follows that there is a possible world w_1 (the actual world) in which *R* plays *C* and *R*-tokens thereby represent themselves, and there is a possible world w_2 in which some representational vehicle *Q* distinct from *R* plays *C* and *Q*-tokens thereby represent *R*-tokens. The former is a case of self-representation, while the latter is a case of other-representation. There is a clear sense, then,

in which promiscuous conceptual roles aren't essentially tied to the phenomenon of self-representation, given that such conceptual roles are compatible with worlds like w_2 from above. If a representational vehicle is self-representational in virtue of a promiscuous conceptual role, I'll say that it's 'non-constitutively self-representational'. There is nothing about the conceptual role *C* of a representational vehicle *R* that is non-constitutively self-representational that suggests that only *R* can play *C* and thereby be such that its tokens represent *R*-tokens.[9]

Above I set out the notion of non-constitutive self-representation in terms of token representational vehicles representing tokens rather than types. There is a related notion, however, concerning the representation of types. Consider a system *S* of representational vehicles that includes, among other representational vehicles, MENTAL REPRESENTATION and HORSE. Let *M* be the conceptual role of MENTAL REPRESENTATION that determines its content (the property *being a mental representation*) and *H* be the conceptual role of HORSE that determines its content (*horseness*). So tokens of MENTAL REPRESENTATION represent *being a mental representation* in virtue of MENTAL REPRESENTATION playing *M*, and tokens of HORSE represent *horseness* in virtue of HORSE playing *H*. The extension of MENTAL REPRESENTATION consists of all the elements of *S*, so this representational vehicle is self-representational in the sense that its tokens are included in its own extension. We can say, then, that MENTAL REPRESENTATION is self-representational in virtue of playing *M*. But no

[9] Though I've introduced the notion of non-constitutive self-representation in the context of CRS, we can separate the former from the latter. Any account of content has something like the following form: a representational vehicle *R* is such that its tokens have the content they do in virtue of the fact that *R* has property *F*. *F* might concern conceptual roles, asymmetric dependence, proper functions, and so on. Here, then, is a rough characterization of non-constitutive self-representation independent from CRS: a representational vehicle *R* is non-constitutively self-representational just in case *R* is self-representational in virtue of having *F*, but *F* isn't essentially tied to the phenomenon of self-representation; *F* can be instantiated by representational vehicles distinct from *R*.

matter how we fill in the details with respect to *M*, there is nothing about it that suggests that only MENTAL REPRESENTATION could have played *M*; HORSE, for example, could have played *M* instead of *H* and would thereby have represented what MENTAL REPRESENTATION in fact represents instead of *horseness*. In this case, the extension of HORSE would have consisted of all the elements of *S* instead of all the horses. If all of this is right, then *M* is a promiscuous conceptual role. While it's true that whatever plays *M* is self-representational in the sense that its tokens fall within its own extension, there is no road back from *M* to MENTAL REPRESENTATION. MENTAL REPRESENTATION is non-constitutively self-representational in this sense.

The mental representations tokens of which are instances of phenomenal properties, on the other hand, are constitutively self-representational. Here's why. Consider REDDISH_1, that representational vehicle tokens of which are instances of the phenomenal property *reddishness*. (It will become clear later why the name for this concept is subscripted – you can ignore it for now.) Suppose that $C_{1\text{-}A}$ is the conceptual role of REDDISH_1 in virtue of which REDDISH_1-tokens represent themselves as instances of *reddishness*. (Again, ignore the subscript for now.) Why think that it's not the case that any representational vehicle distinct from REDDISH_1 could play $C_{1\text{-}A}$ and thereby be such that its tokens represent REDDISH_1-tokens, instances of *reddishness*? (This isn't to deny that tokens of concepts distinct from REDDISH_1 can represent instances of *reddishness*, it's just that they can't do so in virtue of playing $C_{1\text{-}A}$.) In other words, why think that $C_{1\text{-}A}$ is non-promiscuous? The answer to this question is implicit in my discussion in the previous section. There I claimed that the way to explain why phenomenal acquaintance is direct, thick, and infallible is to claim that instances of phenomenal properties represent themselves. This, I claim, is the *only* way to reconcile the claim that phenomenal acquaintance has these features with the conception of attention I

endorse suggested by the representational theory of mind. I can now put this point in another way. When you're acquainted with the reddish character of your experience, there is some representational vehicle Q that plays a certain conceptual role $C_{1\text{-}A}$. Remember that the conceptual role of a representational vehicle, very roughly, is a specification of what the representational vehicle does and what it's responsible for. So what can we say about Q's job description in your cognitive economy? Well, drawing on my earlier discussion of the three key features of phenomenal acquaintance, we can say this much: Q is (at least in part) responsible for the directness, thickness, and infallibility of phenomenal acquaintance. For Q to play $C_{1\text{-}A}$, then, is at minimum for Q to play a role in your cognitive economy in virtue of which your grasp of the reddish character of your experience is direct, thick, and infallible. My claim is that only REDDISH_1 can play $C_{1\text{-}A}$, so $\text{REDDISH}_1 = Q$. So now we have a handle on the nature of the conceptual roles in virtue of which instances of phenomenal properties represent themselves, and we can see why the phenomenal self-representation at issue in my account of phenomenal acquaintance is constitutive self-representation.[10]

Let's now consider some other objections that target my appeal to phenomenal self-representation. In the first chapter I appealed to the idea that phenomenal characters are about non-experiential aspects of the world in the sense that the former are ways the latter

[10] Kriegel (2005) claims that experiential states are self-representational, but he rejects the idea that the fact that they represent themselves is to be explained by their conceptual role. He claims that CRS would require us to say that the *categorical* property of an experience representing itself is *identical* to the *dispositional* property of playing such-and-such conceptual role, but categorical properties aren't dispositional properties. But why think that an experience representing itself is a categorical property of it? As far as I can tell, Kriegel thinks that this is the case because he endorses the same-order monitoring account of experience. An experience representing itself is a categorical property of it on his view because he thinks that experiences are experiences in virtue of representing themselves, and since being an experience is a categorical matter, so too must a mental state representing itself. I remain neutral, however, on what familiar account of experience, if any, is correct. Since I don't endorse the same-order monitoring theory, I see no reason to insist that original self-representation is a categorical matter.

appear. With this claim in mind, you might think that if we're going to claim that instances of phenomenal properties are token mental representations, we should claim that they're other-representational instead of self-representational. For to say that an instance of a phenomenal property represents itself seems at odds with the notion that phenomenal characters are about non-experiential aspects of the world.

Here is a simpler way to get at roughly the same idea, one that doesn't appeal to the phenomenal character as appearance thesis. Suppose you're having an experience as of a brown table. I claim that your token experiential state in this case is itself a property instance and is composed of various property instances, some of which are token mental representations that represent themselves, e.g. an instance of *brownishness*. But on this account, in virtue of what is your experience about, say, the brownness of the table? The concern is that, in claiming that there is a self-representational aspect to experience, we've lost the sense in which our experiences involve other-representation, representation of non-experiential aspects of the world. And surely, so the objection goes, the other-representational status of experience is a surer thing than its putative self-representational status.

How should we respond to this objection? One possibility that we haven't considered is this: instances of phenomenal properties are *both* self-representational and other-representational; they fall under two representational types. Here the claim isn't that there is a single type of representational vehicle such that its tokens in some contexts represent themselves and in others represent distinct items. For to claim, for example, that tokens of the concept REDDISH_1 (= instances of *reddishness*) in certain contexts represent themselves as instances of *reddishness* but in other contexts represent non-experiential aspects of the world as being red would be to introduce ambiguity in the language of thought.

Instead, the claim is that instances of phenomenal properties do both simultaneously in the sense that we can model their content as a conjunction, as having two aspects. On this proposal the content of a token of REDDISH$_1$ has two aspects, one corresponding to its representational vehicle and the other corresponding to a non-experiential aspect of the world. So the moral is that we don't have to choose between self-representation and other-representation, contrary to the objection under consideration.

How does the dual content proposal jibe with my claim above that instances of phenomenal properties represent themselves in virtue of conceptual role? Well, consider again REDDISH$_1$, that representational vehicle tokens of which are instances of the phenomenal property *reddishness*. And suppose again that $C_{1\text{-}A}$ is the non-promiscuous conceptual role of REDDISH$_1$ in virtue of which its tokens represent themselves as instances of *reddishness*. $C_{1\text{-}A}$, you will recall, is that in virtue of which your grasp of the reddish character of your experience is direct, thick, and infallible. Now the proposal is that REDDISH$_1$-tokens not only represent themselves as instances of *reddishness*, but they represent non-experiential aspects of the world as being red as well. If REDDISH$_1$-tokens are self-representational in virtue of the conceptual role of REDDISH$_1$, then, if they're other-representational as well, we should say that this too is a matter of conceptual role. Here we can claim something like this. $C_{1\text{-}A}$, it turns out, is only a partial specification of the complete conceptual role of REDDISH$_1$. The complete conceptual role of REDDISH$_1$ is C_1, and let $C_{1\text{-}B}$ be the aspect of C_1 not included in $C_{1\text{-}A}$ (this is where the subscripts become important). $C_{1\text{-}B}$ includes relations to judgments of the form "*x* is red", desires of the form "I desire that red *x*", and so on. As such, $C_{1\text{-}B}$, unlike $C_{1\text{-}A}$, is promiscuous. So the proposal is that the $C_{1\text{-}A}$ aspect of C_1 is that in virtue of which REDDISH$_1$-tokens are self-representational, and the $C_{1\text{-}B}$ aspect of C_1 is that in virtue of which REDDISH$_1$-tokens are other-representational. If this is

right, then we have an explanation of how our experiences can be self-representational and still be about non-experiential aspects of the world.

Notice that on the view outlined above your beliefs about red things in part determine the other-representational content of REDDISH$_1$-tokens. We don't want to claim, however, that your beliefs about the phenomenal character *reddishness* are in any way involved in determining their self-representational status. For to do so it seems would be to introduce the sort of mediating material ruled out by the directness of phenomenal acquaintance. On the proposal under consideration, the self-representational aspect of the content of REDDISH$_1$ isn't tied to its inferential role, while its other-representational aspect is. We will return to these issues shortly.

There is a potential problem, however, for the dual content proposal. The problem concerns the infallibility of phenomenal acquaintance. In my resolution of the infallibility puzzle, I claim that the mental representations in virtue of which we attend to instances of phenomenal properties can't misrepresent due to their self-representational status. But now we're claiming that instances of phenomenal properties are other-representational as well as self-representational. So consider, for example, an instance of *reddishness*, a token of the concept REDDISH$_1$. According to the proposal under consideration, this property instance has content that can be expressed as a conjunction in the sense that it represents itself as an instance of *reddishness*, and it represents a non-experiential aspect of the world as being red. But suppose that the non-experiential aspect of the world it represents as being red isn't red. In this case the other-representational aspect of the content of the instance of *reddishness* is non-veridical. But if one aspect of the content of a mental representation is false, so too is its content as a whole. Here the instance of *reddishness*, a token of REDDISH$_1$, is a non-veridical mental representation. But if all of this is right, we've lost our grip on the infallibility of

phenomenal acquaintance. Now we must admit that the mental representations in virtue of which we attend to instances of phenomenal properties can go wrong, so the beliefs about our experiences constituted by these mental representations can go wrong as well.

One way to respond to this objection would be to claim that the infallibility of phenomenal acquaintance only requires that the self-representational aspect of the token mental representations in virtue of which we're acquainted with the phenomenal characters of our experiences can't go wrong. And nothing in the discussion above suggests that such mental representations can go wrong in this sense. Let's suppose that the content of an instance of REDDISH_1 is not only expressible as a conjunction but really is conjunctive in nature, so its content is $A\&B$ were 'A' means 'x is an instance of *reddishness*' and 'B' means 'y is red'. Here we can say that, so long as A must be true, we don't have a problem.

I think that this response to the objection from infallibility posed above, however, points to a deeper problem for the dual content proposal. For suppose that the content of REDDISH_1 is $A\&B$ from above because REDDISH_1 plays C_1, were $C_{1\text{-}A}$ and $C_{1\text{-}B}$ are aspects of C_1 such that $C_{1\text{-}A}$ determines that A figures in the content of REDDISH_1 and $C_{1\text{-}B}$ determines that B figures in its content. If this is right, then it seems that the compositionality of content requires that there are some representational vehicles R_1 and R_2 such that (i) R_1, R_2, and REDDISH_1 are distinct; (ii) the content of R_1 is A (not $A\&B$) in virtue of it playing $C_{1\text{-}A}$; and (iii) the content of R_2 is B (not $A\&B$) in virtue of it playing $C_{1\text{-}B}$. But if there is a mental representation like R_1, shouldn't we identify REDDISH_1 – a phenomenal concept whose tokens represent themselves as instances of *reddishness* – with it rather than with a molecular mental representation, one of whose constituents is such that its tokens represent non-experiential aspects of the world as being red? I think we should.

So how should we proceed from here? Happily, I think there is another way to resolve the concern about how our experiences are about non-experiential aspects of the world, one that doesn't appeal to the idea that their contents are conjunctive in nature. This response is implicit in my earlier discussion of the phenomenal character as appearance thesis, and it's far simpler than the proposal we considered and rejected above. Suppose, for the moment, that the phenomenal character as appearance thesis is true. So to have an experience with a reddish character is to have an experience an aspect of which is a non-experiential aspect of the world appearing red. In the section before last, I noted that this thesis is compatible with my claim that instances of phenomenal properties are self-representational. I suggested that that if *p* is an instance of *reddishness*, we can say that *p* is self-representational in the sense that it represents itself as a particular occasion of a non-experiential aspect of the world appearing red. If this is how we are to understand the self-representational status of instances of phenomenal properties, our experiences are other-representational *in virtue of* being self-representational. In other words, they get to be about the world because they're about themselves! Now, to properly motivate this response, I would have to do more by way of defending the phenomenal character as appearance thesis. Part of a defense would appeal, for example, to the fact that it's well suited to explain the transparency of experience as I noted in the first chapter. Considering these issues in further detail, however, would take us too far afield, so for now I'll settle on the following conditional claim: provided that the phenomenal character as appearance these is coherent and well-motivated, we have a plausible response to objection under consideration.

Let's now consider another objection targeting my claim that instances of phenomenal properties are self-representational. Certain non-human animals obviously have experiences. Suppose, for example, that a dog has a painful experience. I'm committed to the

claim that a component of this token experience is an instance of the phenomenal property *painfulness*. On my view, this property instance is a token of a concept, what I'll call 'PAIN_1'. (Again, you can ignore the subscript for now.) But do we want to say that tokens of this concept for dogs are self-representational, that they represent themselves as instances of the property *painfulness*? Well, while there might be no problem with this claim *per se*, there is a problem if its self-representational status is a matter of playing a conceptual role like $C_{1\text{-}A}$ described above, one in virtue of which cognitive access to phenomenal character is direct, thick, and infallible. For dogs presumably don't have mental states about their own mental states, so obviously they're incapable of being acquainted with the phenomenal characters of their experiences. So we don't want to say that in the cognitive economy of dogs PAIN_1 is self-representational in virtue of playing a conceptual role that explains how dogs can be acquainted with the phenomenal character of their experiences.

I think the thing to say in response to this objection is this: though dogs have experiences, for dogs instances of phenomenal properties aren't self-representational because it's not the case that they're tokens of mental representations that play the requisite conceptual role in their cognitive economy. I claim that instances of phenomenal properties are tokens of mental representations that play a certain role in *our* (*qua* humans) cognitive economy – a role in virtue of which their instances represent themselves that explains the directness, infallibility, and thickness of phenomenal acquaintance – not that they play such a role in *any* cognitive agent with experiences. But the experiences of dogs, like those of humans, are about non-experiential aspects of the world, so should we say that instances of phenomenal properties are tokens of mental representations in their cognitive economy, mental representations with other-representational content? I don't think that we need to take a stand on this issue. For recall how I initially put the objection to phenomenal self-

representation discussed above – since I'm committed to the claim that instances of phenomenal properties are tokens of mental representations (in the cognitive economy of humans), shouldn't I say that they're other-representational rather then self-representational, given that our experiences are about non-experiential aspects of the world? By contrast, notice that in the present context (*viz.* the cognitive economy of dogs), we're not assuming that they're tokens of mental representations. So we're free to claim that a dog's experiences are about non-experiential aspects of the world in virtue of the dog tokening mental representations whose tokens aren't instances of phenomenal properties. And if we wish to claim that instances of phenomenal properties are token mental representations in their cognitive economy – ones that represent non-experiential aspects of the world – we can claim that they do so because they're instances of mental representations that play conceptual roles like $C_{1\text{-}B}$ described above, conceptual roles tied to beliefs about pain. If all of this is right, we can admit that dogs and the like have experiences (as we certainly should), yet not be committed to the implausible claim that they're the sorts of creatures that can be acquainted with the phenomenal characters of their experiences.

Here's a final objection to my appeal to phenomenal self-representation. Recall my observation in the second chapter that there are ways of thinking about phenomenal characters such that our thoughts about them don't seem to be direct, thick, and infallible. A related point is that you can think about having an experience with a particular phenomenal character without actually having an experience with it. This consideration can be turned into an objection to my claim that instances of phenomenal properties are self-representational. For suppose that you're reading a philosophy paper and working through someone's account of phenomenal character. Though you're presently not having an experience with a reddish character, you suppose that you are so that you can trace out various consequences of the

account.[11] In supposing that you're having such an experience in this case, it seems that you token the concept REDDISH$_1$. But if I'm right that tokens of this concept are instances of *reddishness*, it follows that you're having an experience with a reddish character, contrary to our assumption. So it turns out that on my view you can't suppose that you're having an experience with a reddish character without actually having one. But surely such a claim is false.[12]

My response to this objection is this: in the case above, you don't actually token the concept REDDISH$_1$, contrary to what the objection above suggests. Let's distinguish between two concepts REDDISH$_1$ and REDDISH$_2$ (here the subscripts become important). Tokens of REDDISH$_1$ represent items as being instances of the phenomenal property *reddishness* in virtue of the fact that REDDISH$_1$ plays $C_{1\text{-}A}$ (as described above), and tokens of REDDISH$_1$ are identical what they represent. (Here the idea is that $C_{1\text{-}A}$ is the conceptual role of REDDISH$_1$ in its entirety, not merely an aspect of it.) Tokens of REDDISH$_2$ also represent items as being instances of *reddishness*, but they aren't identical to what they represent. My claim is that in the case described above in which you suppose that you're having a reddish experience when you're not actually having one, you token REDDISH$_2$ rather than REDDISH$_1$. Tokening REDDISH$_2$ doesn't require that you have a reddish experience, given that its tokens aren't identical to instances of *reddishness*.

---

[11] Here I'm not thinking of a case in which you represent yourself as having a reddish experience in virtue of imaginatively recreating a reddish experience which itself has a reddish character (or something close to it). The case I have in mind above is one in which you aren't having an experience with a reddish character period, though you represent yourself as having one.

[12] An objection along these lines lead Papineau (2007) to reject the quotational account of phenomenal concepts he developed and defended in Papineau (2002).

In virtue of what do tokens of REDDISH$_2$ represent instances of *reddishness*? Here the answer can't be that they do so because REDDISH$_2$ plays a conceptual role like $C_{1\text{-}A}$. For remember that $C_{1\text{-}A}$ is the conceptual role of REDDISH$_1$ in virtue of which your cognitive access to the reddish character of your experience is direct, infallible, and thick. In the scenario described above, you token REDDISH$_2$ in the absence of any reddish experience, so here it's obviously not the case that you have cognitive access to the reddish character of an experience that has these features. I propose that the conceptual role of REDDISH$_2$, unlike $C_{1\text{-}A}$, is promiscuous, and REDDISH$_2$, unlike REDDISH$_1$, is such that your beliefs about the phenomenal character *reddishness* are involved in its content determination. The general idea is that the conceptual role of REDDISH$_2$, like the conceptual role of those mental representations in virtue of which a dog's experiences are about non-experiential aspects of the world, is purely a matter of inferential role. Since your beliefs about *reddishness* are involved in the content determination of REDDISH$_2$, your cognitive relation to *reddishness* isn't direct when you token this concept. This is as it should be, given that we're considering a scenario in which you don't have a reddish experience to begin with. (We can say roughly the same things about PAIN$_1$ from above, distinguishing it from PAIN$_2$.)

Let's now put potential problems with phenomenal self-representation to the side and consider a potential problem with my claim that phenomenal acquaintance is direct. Above I argued that if being acquainted with a phenomenal character is a matter of standing in a relation to the content of an instance of a phenomenal property *qua* self-representational semantic entity, it's clear how phenomenal acquaintance can be both direct and underwritten by mental representation. You might object, however, claiming that my proposal doesn't fully capture the sense in which phenomenal acquaintance is direct. The concern is that the presence of a self-representation in this case would still play an unacceptable mediating role;

it doesn't matter whether its representational vehicle and content are one and the same. Even if the mental representation in question represents itself, it's still the case that you have cognitive access to phenomenal character only "through" the medium of mental representation. This, so the objection goes, doesn't comport with the directness of phenomenal acquaintance.

I have some sympathy with this objection, and the general point here is one we're already familiar with from our discussion of Russellian acquaintance. The important thing to realize, however, is that not all mental representations are equal when it comes to cognitive immediacy. When you token a mental representation that's identical to what it represents, there is just you and the object represented. When you token a mental representation that is distinct from what it represents, there are three things: you, the object represented, and a distinct token representational vehicle that mediates the connection between the first two. In going self-representational, we eliminate this sense of mediation.

Our situation here is reminiscent of Dennett (1984)'s discussion of freewill: there are only certain senses of 'freewill' in which freewill is worth having, and these are precisely the senses, or so he argues, in which we are free. Analogously, there are only certain senses of 'directness' in which it's worth having. I articulated one such sense above and argued that my account of phenomenal acquaintance captures it. Perhaps there are other senses of directness in which it's worth having, and these need to be secured. I don't wish to rule this out. I do think, however, that Russell's conception of directness in the end is directness not worth having.

Finally, let's consider a potential problem with my resolution of the thickness puzzle. The thickness puzzle, as you will remember, is the problem of explaining how attention to an experience confers a grasp of one phenomenal character rather than another, given that

experiences have more than one phenomenal character. In response, I proposed that, instead of attending to the experience *per se*, you attend to an instance of the phenomenal property you come to be acquainted with, one that is both constitutive of the experience in question and represents itself. You might think, however, that my response to the thickness puzzle generates a distinct but similar problem, this time at the level of experiential components (instances of phenomenal properties) rather than experiences themselves. My resolution of the puzzle, so the objection goes, presupposes that instances of phenomenal properties can be the objects of thought. But how is it that in thought a subject gloms onto one phenomenal property instance rather than another? On my account, a given experience consists of many phenomenal property instances, so in virtue of what does the subject make cognitive contact with one as opposed to another? The worry, then, is that a puzzle resembling the thickness puzzle arises at the level of instances of phenomenal properties.

Is this new problem really a problem? I don't think so. We started out with the question of how attention to an experience confers a grasp of a particular phenomenal property. As I noted in the second chapter, the directness of phenomenal acquaintance rules out a natural explanation, that you attend to the experience in virtue of standing in a relation to the content of a mental representation that both represents the experience as having the phenomenal character in question and is distinct from what it represents. But notice that we didn't reject this potential explanation by claiming that it's unclear how you can stand in a relation to the content of a mental representation in the first place. To do so would have been misguided, given our commitment to the representational theory of mind. The thrust of the objection under consideration, however, is that we should have done so, for the current objection is really a demand for an explanation of how it is that a subject can stand in a relation to the content of a mental representation, given my claim that instances of

phenomenal properties themselves are mental representations. My response to the objection, then, is this: if there is an interesting question here (although I'm not sure that there is), it's one for the representational theory of mind and not about attending to instances of phenomenal properties *per se*. The status of the representational theory of mind here is that of a plausible background assumption, and to defend it in any detail falls outside the scope of my project.

4.6. Conclusion

I conclude that none of the objections to my account of phenomenal acquaintance discussed above are persuasive. Now that we have an account of phenomenal acquaintance that not only resolves the directness, infallibility, and thickness puzzles, but survives (at least some degree of) critical scrutiny, in the next chapter I return to an issue discussed in the first chapter, that of the phenomenal concept strategy. I wish to appeal to my account of acquaintance in order to explain why there is an explanatory gap between the physical and the phenomenal in a way amenable to physicalism.

CHAPTER 5

PHENOMENAL ACQUAINTANCE AND PHYSICALISM

5.1. Introduction

In this final chapter, I turn to the phenomenal concept strategy I mentioned in the first chapter. There I objected to one way of implementing the strategy – simply imputing to phenomenal thought whatever features might explain away our dualist intuitions. I proposed instead that we put the phenomenal concept strategy to the side and consider phenomenal acquaintance on its own terms. But now that we have an account, we can return to the issues of what implications, if any, phenomenal acquaintance has for physicalism. In this final chapter I argue that my account of phenomenal acquaintance, in addition to being compatible with physicalism, explains why there is an explanatory gap between the mental and physical, thereby undercutting various familiar objections to physicalism. In other words, my account of phenomenal acquaintance is what advocates of the phenomenal concept strategy have been looking for in responding to objections to physicalism, or so I claim.

5.2. The Explanatory Gap and Physicalism

Before we turn to the consequences of my account of phenomenal acquaintance for physicalism *vis-à-vis* the explanatory gap, we should get clearer on the nature of the explanatory gap and physicalism, and how the two are related. Following Levine (2001, 2007), compare the following identity statements: "Water is H_20" and "Pain is C-fiber firing". Suppose that we learn all there is to know about the molecular structure of H_20, including how it's responsible for the superficial properties of what we identify as water. Then suppose that someone asks, "How could water be *that*?" where 'that' in this context

refers to H_20. It seems that there is no cognitive significance to this question. Here a demand for further explanation seems misplaced; what more needs to be said? Now suppose that we come to know all about the nature of C-fiber firing and someone then asks, "How could pain be *that*?" where 'that' in this context refers to the property of C-fiber firing. It seems that there is indeed cognitive significance to this question, for a distinctive arbitrariness attends the connection between pain *qua* phenomenal property and C-fiber stimulation (or any physical property, for that matter), while this isn't the case for the connection between water and H_20.

We encounter the same contrast if we shift our focus to the claim that phenomenal properties are instead identical to functional properties. Consider the following claim: "Pain is identical to a certain second-order property, the property of having some property or other, instances of which play causal role *R* (the causal role of pain). In the actual world, some physical property, say C-fiber stimulation, is such that its instances play *R*, so the property of being in pain is instantiated in virtue of physical properties *cum* causal roles in the actual world." Suppose that we come to know all about the nature of C-fiber firing, including its causal role, as well as the causal role of pain (and we find that they are the same), and then someone asks, "How could pain be *that*?" where 'that' in this context refers to the functional property described above. A distinctive arbitrariness also attends the connection between pain and the functional property in question. The connection between them isn't *completely* arbitrary granting their sameness in causal role, but why, from the first-person perspective, should it be like so-and-so to instantiate a physical property, an instance of which plays a certain causal role? Indeed, why should it be like anything at all?

With what I've said above in the background, I can now give a perspicuous formulation of the explanatory gap. There is an explanatory gap between the physical and

the phenomenal in the sense that it seems that, no matter what we learn about the physical, any identification of a phenomenal property with a physical or functional property will seem arbitrary; it seems that it will always make sense to ask, "How could this phenomenal property be that physical/functional property?" The cognitive significance of this question indicates that we lack an explanation of the phenomenal in terms of the physical. Were we to come upon such an explanation, we would come to have a general understanding of how it is that pain is identical to some physical or functional property, just as we have a general understanding of how it is that water is identical to H_20. The cognitive significance of the question posed above would disappear.

Now we can turn to physicalism. Though I've made certain claims about physicalism in the course of this dissertation, I haven't offered anything like a statement of the view. In the end I don't think that type identity theory, so-called 'reductive physicalism', is tenable (for the standard reasons), so let's put it to the side for now. We needn't address, then, our lack of a handle on how it could be that the property of being in pain is identical to a particular physical property because we're assuming that they're distinct.[1] Let's understand physicalism, then, as the claim that mental properties are identical to functional properties, and these functional properties, if instantiated in the actual world, are instantiated ultimately in virtue of physical properties. In other words, in the actual world, mental properties are instantiated in virtue of physical properties *cum* causal roles. This thesis is often called 'non-reductive physicalism' or 'functionalism', and is understood as a particular version of

[1] See Trogdon (2009a), however, for some considerations sympathetic to reductive physicalism. There I argue that the debate in sparse ontology between the priority monist and priority pluralist is relevant to the thesis. I argue in particular that a familiar, important objection to reductive physicalism works in the pluralist's framework but not the monist's. See Schaffer (forthcoming) and Trogdon (2009b) for further discussions of monism and pluralism.

physicalism. Since this is the only version of physicalism that I think has a chance of being right, I will refer to it simply as 'physicalism'.

Physicalism so understood is a thesis about the actual world as well as other possible worlds. With respect to the actual world, it says that mental properties, if instantiated, are instantiated ultimately in virtue of physical properties (*cum* causal roles), and these properties alone. It rules out the possibility that mental properties are instantiated in the actual world either in a basic way (in virtue of no properties at all) or ultimately in virtue of further mental properties. With respect to other possible worlds, there are two points to note. First, the thesis doesn't say that a mental property, if instantiated in another possible world, is instantiated in virtue of a physical property there – perhaps it's instantiated in virtue of a non-physical property.[2] Second, the in-virtue-of relation (= realization, or perhaps the latter is a species of the former) is a relation of asymmetric metaphysical necessitation. Hence, it says that if a mental property is instantiated in virtue of a physical property in the actual world, then, if that physical property is instantiated in any other possible world and plays the causal role it plays in the actual world, so too is the mental property, and the latter is instantiated in virtue of the former.

Now that we have working conceptions of the explanatory gap and physicalism, we can turn to the connection between the two. Since physicalism is a thesis about ontological dependence relations between mental and physical properties while the explanatory gap concerns the intelligibility or explanation of the former in terms of the latter, you might

[2] In Levine and Trogdon (forthcoming), we argue that, contrary to what many think, it doesn't make sense to say both that physicalism so understood is true in the actual world, yet in other possible worlds mental properties are instantiated in a basic way, not in virtue of any other properties. We conclude from this that there is an important sense in which physicalism is a necessary rather than contingent thesis, again contrary to the (more-or-less) received view on this matter.

think that the explanatory gap is orthogonal to physicalism. Stoljar, for example, claims that "…the fact that non-experiential truths fail to render experiential truths intelligible does not by itself give us reason to doubt supervenience; absence of intelligibility does not mean absence of entailment" (2006, 98). On this view the existence of the explanatory gap isn't thought to directly tell against physicalism.

I maintain, however, that entailment or ontological dependence and intelligibility or explanation cannot be so clearly separated. In Levine and Trogdon (forthcoming), we argue that there are no brute necessities: metaphysically necessary truths that are neither conceptual nor logical truths. Why the prohibition against brute metaphysical necessity so understood? A truth is brute just in case there is no illuminating explanation for why it's true. Some truths about our world are, or at least might be, brute in this way. That certain fundamental physical magnitudes are what they are – say the speed of light, or the gravitational constant – might be just brute facts. If you ask why, you get the answer, "because that's how the world is". The actual, we argue, is the proper place for the brute. "That's just the way it is" seems appropriate when it's contrastable with "of course it could have been otherwise". And brute nomological necessity is okay, since this is just part of the bruteness of the actual world. But brute metaphysical necessity – being told not only that this just happens to be how things are, but this happens to be how they have to be – this we really don't understand. What "just happens to be" can't be "how it has to be".

But how do we get from the claim, on the one hand, that there are no metaphysically necessary truths such that there is no illuminating explanation concerning why they're true, to the claim, on the other, that all metaphysically necessary truths are either conceptual or logical truths? The basic idea is that conceptual truths and logical truths are intelligible in the sense that, when you're considering whether a certain proposition is true and come to realize

that it's a conceptual or logical truth, you realize that there is nothing further to understand with respect to why it's true in the actual world and all possible worlds. The only way for metaphysical truths to be intelligible, so the idea goes, is for them to be conceptual or logical truths. But what about true identity statements – aren't they brute metaphysical necessities? We don't think so. While identity is brute in the sense that it's not the case that something is identical to itself in virtue of some further fact obtaining, it isn't metaphysically brute given that the identity relation, in our estimation, anyway, is a logical relation.

Suppose, then, that all metaphysical truths are either conceptual truths or logical truths and are, therefore, intelligible. Hence, if, for the life of us, we can't understand even in principle why it should be that the instantiation of one property ensures the instantiation of another property across the space of metaphysically possible worlds, we have a fairly good reason to believe that there is no conceptual or logical truth in the vicinity. In this case we therefore have a fairly good reason to believe that the instantiation of the former doesn't metaphysically necessitate the instantiation of the latter, that the latter isn't instantiated in virtue of the former. And this, given the explanatory gap, is precisely the situation we face with the phenomenal and the physical/functional. We just don't see why the instantiation of any particular physical property playing a particular causal role should ensure the instantiation of a particular phenomenal property across metaphysically possible worlds, let alone in the actual world. Hence it seems that the explanatory gap poses a formidable, though, of course, not decisive, problem for physicalism.

The physicalist, therefore, needs to find some way to accommodate the explanatory gap. I follow Levine (2001) in viewing other prominent challenges to physicalism, e.g. the conceivability argument (cf. Chalmers (1996, forthcoming)), the knowledge argument (cf. Jackson (1982, 1986)), and so on, as something like special cases of the explanatory gap. So if

the physicalist can come to terms with the explanatory gap, she will thereby be able to deal with these challenges to physicalism as well. The idea is that, were the physicalist to come to terms with the explanatory gap, it would become clear why it seems to us that there are metaphysically possible zombie worlds, but in such a way that doesn't require their possibility, *contra* the conceivability argument. It would also become clear why it seems to us that Mary gains knowledge in such a way that is incompatible with physicalism after she leaves the confines of her room, but in such a way that doesn't require that this is the case, *contra* the knowledge argument. But short of providing the very explanation we lack – short of bridging the gap, as it were, between the physical/functional and the phenomenal – how might the physicalist respond to the explanatory gap?

5.3. The Phenomenal Concept Strategy

This is where what Stoljar (2005) has dubbed the 'phenomenal concept strategy' comes into play. Instead of trying to produce the missing explanation, the idea is to explain in a way compatible with physicalism why it seems to us that the connection between the phenomenal and the physical/functional is arbitrary. The strategy is to first look to the way we *think* about our experiences when we have them. Then the idea is to show that there are various features of these thoughts that can, on the one hand, be explicated within the physicalist's framework and, on the other, explain why there is an explanatory gap, why it seems to us that, whatever the relevant physical/functional truths turn out to be, they would fail to explain the phenomenal truths.[3]

[3] Some (e.g. Stoljar (2005) and Chalmers (2007)), present the phenomenal concept strategy as primarily an effort to resolve the conceivability and knowledge arguments against physicalism, while others (e.g. Tye (1999) and Levine (2007)) in this context focus on the explanatory gap as I do above. Some (e.g. Loar (1997)) focus on both.

The goal, in other words, is to explain in a way amenable to physicalism the 'appearance of contingency' between the physical/functional and the phenomenal, to use Kripke's (1980) term. Now, we know that if physicalism is true, the relation between the physical/functional and the phenomenal is one of metaphysical necessity, and, with our admonition against brute metaphysical necessity, we also know in this case that the relation is either conceptual or logical. But the relation seems contingent and thereby seems neither conceptual nor logical. An explanation of the appearance of contingency, taken together with independent arguments for physicalism (think causal exclusion arguments), clinches the case for the thesis, or so the idea goes. Remember, however, that such an explanation doesn't close the explanatory gap – instead of providing the missing explanation, it explains why we lack such an explanation. Hence, though successfully deploying the phenomenal concept strategy would explain why the connection between the physical/functional and the phenomenal seems neither conceptual nor logical, it wouldn't tell us how the phenomenal truths actually follow from the physical/functional truths as a matter of conceptual or logical necessity. In other words, it would tell us why the connection seems metaphysically brute, but it wouldn't tell us why it isn't brute. So in focusing on explaining why there is an explanatory gap rather than bridging the gap, we've traded in a very ambitious explanatory project for a less ambitious one.

What, however, would count as an explanation of the explanatory gap? Levine claims that the substantive and determinate nature of our conception of phenomenal character is the basis for the explanatory gap:

> [The idea] that modes of presentation by which we come into cognitive contact with qualia are substantive and determinate… explains why there is an explanatory gap between qualia and their material bases… The epistemic puzzle arises precisely because we have the kind of cognitive grasp of qualitative character that we do. Put simply, the substantive nature of our

> conception provides the material for the substantive nature of our explanatory demand (2001, 8-9).

I agree with this diagnosis of the explanatory gap as far as it goes, but I would add that, in addition to being substantive and determinate, the directness and infallibility of our grasp of phenomenal character that attention affords us contribute to the explanatory gap.[4] It is the fact that we can grasp a property as a phenomenal character in this distinctive way – in a way marked by directness, thickness, and infallibility – and the fact that we can't grasp a property as a physical or functional property in this way, that results in our impression that the phenomenal isn't to be explained in terms of the physical. Henceforth, we'll assume that to explain why there is an explanatory gap in a way compatible with physicalism is to explain how thought about our experiences is direct, infallible, and thick in a way compatible with physicalism.[5]

Levine claims elsewhere, however, that the above diagnosis of the explanatory gap creates a new explanatory problem:

> One might say that there now is a second explanatory gap: between implementations of cognitive architecture and whatever it is about phenomenal concepts – in my terms, that [they afford a substantive and determinate grasp of phenomenal character] – that is responsible for the original explanatory gap. If one thought the original explanatory gap was a

---

[4] Levine (in conversation) doesn't object to this addition, though he is less sure about the feature of infallibility. And his (2001) conception of the substantive and determinate nature of our grasp of phenomenal character includes some of what I've described as its directness.

[5] You might claim that explaining these features is only part of what is necessary to successfully implement the phenomenal concept strategy. The directness and thickness of your acquaintance with, say, the reddish character of your experience (putting its infallibility to the side for the moment) provide you with a special conception of *reddishness*. But what we need, so you might argue, is an explanation of this special conception itself, not just of that which provides you with it (the features of directness and thickness). While I think that there may be something to this suggestion, I don't have a firm grip on the nature of the explanatory demand here. For it seems that to explain how your acquaintance with *reddishness* is direct and thick in this case just is to explain what's special about your conception of it. What else needs to be explained?

> problem and needed to be explained away, then one ought to be bothered by this one as well (2007, 165).

Levine (2006a, 2007) argues that we don't really have a clue how to discharge this new explanatory burden. So he claims, in other words, that in deploying the phenomenal concept strategy, we're trading one explanatory gap for another, so we haven't really made any progress in defending physicalism. He claims that, just as we have no idea how an explanation of the phenomenal in terms of the physical might go, we are more or less equally at a loss when it comes to explaining the substantive and determinate nature of our grasp of phenomenal character in terms that are compatible with physicalism. Recalling my discussion of the three puzzles concerning phenomenal acquaintance, I certainly agree that, on the face of it, it's quite puzzling how phenomenal thought could have the features of directness, thickness, and infallibility. The job of the proponent of the phenomenal concept strategy is to give a physicalist friendly explanation of how such thoughts have these features.

Proponents of the phenomenal concept strategy differ on what the relevant features of phenomenal thoughts are that are supposed to explain why there is an explanatory gap between the physical/functional and the phenomenal. I would like to mark off two conceptions of the structure of phenomenal concepts in order to discuss two corresponding approaches to the phenomenal concept strategy. At various points in previous chapters we've considered the distinction between the higher-order and same-order monitoring accounts of experience. I pointed out that proponents of both views agree that experience is to be understood in terms of (peripheral) awareness, but they disagree about how experiences and their corresponding states of awareness are related. The proponent of the higher-order monitoring account claims that experiences and their corresponding states of awareness are wholly distinct (they share no parts; neither is constitutive of the other), while, according to one version of the same-order monitoring account, the former are proper parts

or constitutive of the latter.[6] There is a corresponding disagreement about the structure of phenomenal concepts: the more-or-less standard view is that phenomenal concepts are directly referential and wholly distinct from their referents (cf. Loar (1997), Perry (2001), Tye (2003), Aydede and Güzeldere (2005), and Levin (2007)), while a competing view is that phenomenal concepts are in some sense constituted by their referents (cf. Papinaeu (2002, 2007), Chalmers (2003), and Balog (unpublished ms)). Those who endorse versions of the constitutive account and are sympathetic with the phenomenal concept strategy tend to claim that it's the fact that phenomenal concepts are constituted by their referents but physical/functional concepts aren't that explains why there is an explanatory gap, while those who endorse versions of the standard account claim that the relevant difference is a cognitive one independent from the metaphysical notions of constitution, parthood, and so on. Here the cognitive difference is normally taken to be a function of a difference in possession, acquisition, or application conditions (or some combination thereof) between phenomenal and physical/functional concepts. Let's consider the phenomenal concept strategy cast in terms of versions of both the standard and constitutive accounts of phenomenal concepts and see how it fares.

We begin with an example of the standard account of phenomenal concepts. Loar's (1997) discussion of phenomenal concepts is the *locus classicus* of the phenomenal concept strategy, and he defends a version of the standard account. What is Loar's account, then? Central to it is the claim that phenomenal concepts are demonstrative in nature, that their

[6] In the first chapter we considered a different version of the same-order monitoring account, one according to which the experience-making state of awareness is a proper part of the experience rather than the other way around. See Kriegel (2006) for formulations and discussions of a variety of versions of the same-order monitoring account.

contents have the form 'that *P*' (where *P* is a phenomenal property).[7] Without getting into the (much discussed) further details of Loar's proposal, why think that it might explain why there is an explanatory gap between the physical and the phenomenal in a way compatible with physicalism?

To begin, it doesn't seem that we have any reason to think that the claim that phenomenal concepts are demonstrative in nature is incompatible with physicalism. There are demonstrative expressions in natural language and Mentalese is a language *par excellence*, so we have every reason to expect that there are demonstrative expressions in the language of thought.[8] Putting the issue of the compatibility of his account of phenomenal concepts with physicalism to the side, Loar claims that on his proposal phenomenal concepts and physical/functional concepts are "cognitively isolated", and this is relevant to explaining why there is an explanatory gap. As far as I can tell, for Loar phenomenal concepts and physical/functional concepts are cognitively isolated in the sense that, given the demonstrative nature of phenomenal concepts, you can possess a phenomenal concept without being able to discern *a priori* any link between the referent of that concept and what you conceive of in a non-demonstrative fashion. Let's temporarily bracket my claim that to give a physicalist friendly explanation of the explanatory gap is to explain in a way

[7] See Perry (2001), Levin (2007), and Schroer (unpublished ms) for other accounts of phenomenal concepts according to which they're demonstrative or indexical in nature. Loar also claims that phenomenal concepts are "recognitional" in the sense that a subject posses a phenomenal concept whose referent is phenomenal property *P* just in case the subject has undergone an experience with *P* and thereby has certain dispositions to recognize, discriminate, and identity experiences with *P*. (Stoljar (2005) interprets Loar differently on this count, claiming that for Loar having had an experience with *P* is a contingent feature of phenomenal concept acquisition, not an essential feature of phenomenal concept possession.) I think that the demonstrative nature rather than the recognitional nature of Loarian phenomenal concepts is relevant to the phenomenal concept strategy, so I won't discuss recognitional concepts further. For a contrary view, see note 12.

[8] See Levine (1988) for a discussion of demonstratives in Mentalese.

compatible with physicalism how phenomenal acquaintance is direct, thick, and infallible. If instead all we need to do is explain why there is no *a priori* connection between the phenomenal and the physical/functional, perhaps we have a successful implementation of the phenomenal concept strategy on our hands.

Now, many of us are comfortable with the notion of *a posteriori* necessity, so why think that the connection between the phenomenal and the physical/functional should be *a priori* in the first place? In considering this question, let's suppose with Loar that phenomenal concepts are demonstrative in nature and therefore directly referential. According to Loar, many of us endorse the following set of claims. Since a phenomenal concept *P* is directly referential, it affords a direct grasp of its referent. From this we think it follows that, if *P* and a physical/functional concept were co-referential, we would see that this is so *a priori*; in particular, we would see *a priori* that there is a physical mode of presentation of the referent of *P*.[9] Loar claims that we come to think that there is a special problem concerning the idea that phenomenal properties are physical/functional properties when we realize that we lack such *a priori* knowledge. He maintains, however, that it's misguided to think that the co-reference of phenomenal concepts (*qua* directly referential semantic entities) and physical/functional concepts requires that you could see *a priori* that there is a physical mode of presentation of the referent of *P*. To insist that this is so, Loar claims, would be to think that the directly referential status of phenomenal concepts affords a grasp of the *essence* of

[9] Loar extrapolates this thesis from Kripke's (1980) discussion of the mind/body problem. Though Kripke is indeed suspicious of the claim that the connection between the phenomenal and the physical/functional is *a posteriori*, claiming that the standard way of explaining away the appearance of contingency doesn't work in the mind/body case, the thesis goes beyond anything Kripke actually claims. This is because Kripke doesn't explicitly address how the modes of presentation associated with physical/functional concepts function in his discussion of the mind/body problem. His discussion instead is restricted to the modes of presentation associated with phenomenal concepts.

phenomenal properties. Loar calls this view 'Platonism' about mental essences. He claims that we have no good reason to accept Platonism, and once we reject it our feeling that as a matter of principle the phenomenal and the physical/functional must be distinct vanishes.

Whether or not Loar's diagnosis of why some think that physicalism is an *a priori* matter is correct, I suspect that if cognitive isolation is understood as conceptual independence we don't have an explanation of the explanatory gap.[10] Following Block and Stalnaker (1999) and Levine (2001), it seems that the concepts of H_20 and water, for example, are independent from another, yet there is nothing like an explanatory gap between the H_20 truths and the water truths (as I argued above).[11] Hence, if there is an explanatory gap between the *A*-truths and the *B*-truths, it isn't a function of the independence of the *A*-concepts from the *B*-concepts and *vice versa*. Hence, to give a (true) theory of phenomenal concepts according to which they are independent from physical/functional concepts isn't to explain why there is an explanatory gap between the phenomenal and the physical.[12]

So now the question is this: putting to the side the notion of cognitive isolation, does Loar's account of phenomenal concepts explain why phenomenal acquaintance is direct, infallible, and thick? The answer, I think, is 'no'. Although in the first chapter I argued that

---

[10] It may be that Loar's conception of cognitive isolation is richer than what I have attributed to him above. He seems to suggest, for example, that the (putative) fact that phenomenal concepts and physical/functional concepts are realized in different areas of the brain is relevant to their cognitive isolation. But how could this be relevant, given that our concern is cognitive rather than neurobiological in nature? Perhaps Loar's suggestion is, given that these concepts are realized in type distinct neurological structures, their functional roles have important differences. At any rate, it's not terribly clear what Loar has in mind here.

[11] Chalmers and Jackson (2001) disagree, arguing that such concepts aren't independent.

[12] Stoljar (2005) has a different take on the phenomenal concept strategy cast in terms of Loar's proposal. He accepts that an explanation of why there is no *a priori* connection between the phenomenal and the physical/functional would suffice, but he argues that in the end Loar hasn't provided one. He also focuses on the recognitional rather than demonstrative aspect of Loar's proposal (see note 6).

the directness of phenomenal acquaintance is in part captured by the fact that phenomenal acquaintance is demonstrative in nature, I pointed out that we are still left with a puzzle about directness (not to mention the infallibility and thickness puzzles).[13] Extrapolating from this case study, the moral I'm inclined to draw is that any account of phenomenal concepts whose defining feature is that phenomenal concepts are directly referential doesn't seem to have the resources to explain phenomenal acquaintance. Hence, they too will fail to explain why there is an explanatory gap.

Let us then turn to the constitutive account of phenomenal concepts. Before we get into the details of a specific version of this account, why think that we should work with it in deploying the phenomenal concept strategy in the first place? In other words, what does it have that the standard account lacks? The idea, so far as I can tell, is that by building a phenomenal property into a phenomenal concept, as it were, phenomenal concepts and phenomenal properties will have the sort of intimate relationship required by the directness, infallibility, and thickness of phenomenal acquaintance. We will see if this idea is borne out.

To begin, what does it mean to say that phenomenal properties are constitutive of phenomenal concepts? Different versions of the constitutive account answer this question differently. Let's consider Balog's (unpublished ms) account. According to Balog, a phenomenal property is constitutive of a phenomenal concept just in case it's metaphysically necessary that if you token the phenomenal concept then the phenomenal property is instantiated.[14] There are at least two issues we need to pursue here. First, in virtue of what is

[13] It seems clear that adding in the claim phenomenal concepts are recognitional in nature (see note 6) doesn't help resolve the directness puzzle.

[14] Papineau's (2007) conception of the relevant sense of 'constitution' is much weaker. He claims that the relevant sense of 'constitution' is contingent in that there are nomologically

there a metaphysically necessary connection between phenomenal concepts and phenomenal properties? Recalling my discussion of brute metaphysical necessity earlier in this chapter, the connection must ultimately be either conceptual or logical. Since we're dealing with the relation between a concept and its referent, it's hard to see how the relation could itself be conceptual, so it seems that it's logical in nature, whatever it is. Second, granting that it's metaphysically necessary that if you token a phenomenal concept a certain phenomenal property is instantiated, in virtue of what does this property figure in the content of the concept in question? Metaphysical necessitation alone, of course, isn't sufficient for mental representation; otherwise, the color scarlet would represent the color red.[15]

In response to the second question, Balog appeals to the conceptual role of phenomenal concepts. According to Balog, there is an operation on instances of phenomenal properties that introduces phenomenal concepts in a manner similar to how quotation introduces lexical items in natural language. In the case of natural language, when we put quotation marks around a symbol, we get a new symbol that represents the old symbol. For Balog, an instance of a phenomenal property is itself a symbol (it's a non-

---

possible worlds in which the tokening of a phenomenal concept doesn't require the instantiation of phenomenal properties.

[15] Balog (unpublished ms) claims that causal/nomic accounts of content fixation require an external relation between a phenomenal concept and its referent. She claims that, since phenomenal properties are constituted by their referents and constitution is an internal relation, causal/nomic accounts of content fixation are therefore out. This isn't quite right, however. Recall from the fourth chapter (note 9) Kriegel's (2003, 2005) view on the relationship between causal/nomic accounts of mental representation and the same-order monitoring account of experience. Suppose that *e* is a token experience and *m* is the token state of peripheral awareness that accompanies *e*. Kriegel, as we saw, considers the view that *e* and *m*, though wholly distinct (they share no parts) are proper parts of a common state *e**. In this case, *e* and *m* are distinct, so there is no principled objection to appealing to a causal/nomic account of mental representation in this context. Assuming that constitution in Balog's sense isn't identity (as we will soon see, it isn't), phenomenal properties and their referents are distinct, so the same consideration applies to her discussion. For more on these issues, see note 16.

conceptual mental representation in the language of thought, one that perhaps represents some feature of an object of perception), and when we "quote" that symbol, we get a new symbol (a phenomenal concept) that represents the phenomenal property instance that is the old symbol. Just as (we will assume) the content of a quotation in natural language is determined by a disposition to accept all instances of a certain disquotational schema, the same goes for phenomenal concepts. We can understand these dispositions as the conceptual role of the term in natural language and the phenomenal concept, respectively.[16]

Notice that if Balog's answer to the second question is correct, it seems that we also have an answer to our first question about metaphysical necessity. Again, in virtue of what does the tokening of a phenomenal concept metaphysically necessitate the instantiation of a certain phenomenal property? Well, returning to natural language, it's obvious that any metaphysically possible world in which there is an occurrence of the lexical item "apple" (here we have a quotation of a quotation) is a world in which there is an occurrence of the lexical item 'apple'. The same goes for phenomenal concepts. Any metaphysically possible world in which you token the Mentalese symbol ***P*** is a world in which you token the Mentalese symbol **P**. On Balog's proposal, **P** is an instance of a phenomenal property, and ***P*** is a phenomenal concept, one generated by "quoting" the phenomenal property instance. Not only is it metaphysically necessary that if you token ***P*** you also token **P**, but, following our discussion above, the latter is the content of the former. Notice that the metaphysical necessitation in this case is a matter of syntax, and, as such, is a matter of logic.

[16] Though a phenomenal concept being partially constituted by its referent is compatible with a causal/nomic account (recalling our discussion from the previous note), it seems incompatible with the proposal concerning quotation. For it isn't clear how the quotation mechanism Balog has in mind could be a matter of causal-nomic connection.

Hence, we don't have a case of brute necessity here. It would seem, then, that Balog has answers to both of our questions posed above.

Having gotten a handle on the essential details of Balog's proposal, we can return to the phenomenal concept strategy. It would seem that her version of the constitutive account of phenomenal concepts is compatible with physicalism, but does it explain why there is an explanatory gap? In particular, does it explain why phenomenal acquaintance is direct, infallible, and thick, resolving our corresponding puzzles? Let's begin with the infallibility puzzle. In this case, we want to answer the following question: if phenomenal acquaintance is underwritten by attention and attention is underwritten by mental representation, why think that the mental representations in question can't misrepresent? If the mental representations in question are phenomenal concepts and Balog's version of the constitutive account is correct, it seems that we have an answer to our question. Given the quotational nature of phenomenal concepts on Balog's proposal, it's metaphysically necessary that if a phenomenal concept is instantiated, the property an instance of which it represents is also instantiated. Hence, there is no room for misrepresentation on her account.

Now let's turn to the thickness puzzle. Suppose, for example, that you're having a reddish experience and, in virtue of attending to your experience, you grasp the phenomenal property *reddishness*. Our question now is this: given that your experience instantiates multiple phenomenal properties, how is it that via attention you glommed onto *reddishness* rather than some other phenomenal property? What should Balog say? First, let's suppose (as I do in my account of phenomenal acquaintance) that token experiential states have instances of their phenomenal characters as components, and acquaintance is a matter of attending to instances of phenomenal properties *qua* components of token experiential states rather than attending to experiences *per se*. Second, suppose that the states of attention in virtue of which

we're acquainted with our experiences are underwritten by Balogian phenomenal concepts. So the Balog-inspired proposal is something like this. If you've come to be acquainted with the reddish character of your experience, here's how you did so. You attended to a component of your experience, an instance of the phenomenal property you're now acquainted with. Your state of attention was underwritten by a Balogian phenomenal concept, an instance of which represents the object of attention – an instance of *reddishness* – in virtue of "quoting" it. Now, recall my discussion of the representation of property instances from the previous chapter. There I argued that to represent something x as being property instance p is (at least) to represent x as an instance of the property P that p is an instance of. Remember that, if this weren't the case, it would be unclear what it would be to represent one property instance rather than another. So in representing an instance of *reddishness*, the token Balogian phenomenal concept represents it *as* an instance of *reddishness*. If this is how the attention relevant to phenomenal acquaintance works, it's clear how it is that you came to grasp *reddishness* in virtue attending to (a component of) your experience. Thus it seems that we have resolution of the thickness puzzle.

Finally, let's consider the directness puzzle. The question we want to answer in this case is this: if phenomenal acquaintance is underwritten by attention, and the mental representations in virtue of which you attend to (components of) your experiences are distinct from your experiences (and their components), how could it be that phenomenal acquaintance is direct? Perhaps if the relevant mental representations are Balogian phenomenal concepts, we have an answer to our question. For in this case, phenomenal properties are constitutive of their corresponding phenomenal concepts in virtue of the fact that tokens of the former are "quotations" of instances of the latter. It's this relation akin to

quotation, so the idea goes, in virtue of which our cognitive access to phenomenal character is direct.

Though I think the Balog-inspired proposal may be able to handle the infallibility and thickness puzzles, I think it fails to resolve the directness puzzle. Let me explain why. Let's begin by recounting Levine's (2007) observation that the mere physical presence of an instance of a phenomenal property when you token a phenomenal concept that represents it doesn't explain the sense in which our cognitive access to phenomenal character when we attend to our experiences is direct. After all, in the case of the veridical perception of a tomato, the tomato is physically present but not cognitively present in the relevant sense. So it's clear that mere physical presence doesn't amount to cognitive presence. Hence, it would be a mistake to say that on the Balog-inspired proposal our cognitive relation to phenomenal character is direct in the appropriate sense in virtue of the fact that phenomenal concept tokens are always accompanied by the phenomenal property instances they represent. And the fact that an instance of a phenomenal property is "in your head" while the tomato isn't doesn't help either. Physical proximity and cognitive proximity are orthogonal.

You might respond by reminding us that on the Balog-inspired account the relation between a token phenomenal concept and its referent – an instance of a phenomenal property – isn't mere physical co-presence, for their relation is semantic in nature. This was the whole point of the discussion of the quotation-like mechanism in virtue of which token phenomenal concepts represent what they do. So not only do tokens of phenomenal concepts and instances of phenomenal properties enjoy physical co-presence, but the former represent the latter. The idea is that physical co-presence and this semantic relation are enough to explain their cognitive proximity.

I think it's clear, however, that this won't work either. Returning to the example of your veridical perception of the tomato, not only is the tomato physically co-present with you, but it obviously figures in the content of your perception; your perception, after all, is one of the tomato. But again, your cognitive access to the tomato in this case isn't direct in the appropriate sense. So pointing out that token phenomenal concepts and phenomenal property instances are not only physically co-present but are such that the former represent the latter doesn't secure the directness of phenomenal acquaintance.

Perhaps it's not just the physical presence of the phenomenal property instance and the fact that it's the content of a mental representation, but also the modal connection between Balogian phenomenal concepts and phenomenal properties that is supposed to give us directness. So let's grant for the sake of argument that, though tokens of Balogian phenomenal concepts and instances of phenomenal properties they represent are distinct, it's metaphysically necessary that if you token such a concept, a phenomenal property is instantiated, an instance of which is represented by the concept. Does this secure this directness of phenomenal acquaintance? I don't think so. For we can represent the proposal as follows:

$$\text{You} \nearrow b \blacktriangleright p \overset{P}{\downarrow} \text{(= component of } e\text{)}$$

Figure 4: Balogian Phenomenal Acquaintance

Here we represent you attending to p – an instance of phenomenal property P and a component of your experience e – as 'You ➢ b ▸ p' where 'x ▸ y' means 'x is a quotation of y', and the other symbols have the same values as before. Here it's understood that b, a token Balogian phenomenal concept, plays the role characteristic of attention in your cognitive economy. Let b be a token of the mental representation B. Though there is a

necessary connection between *b* and *p* in the sense that it's metaphysically necessary that if you token *B* you have an experience that instantiates *P*, *b* and *p*, *ex hypothesi*, are distinct. As such, *b* mediates your cognitive relation to *p*, and therefore to *P*. So it seems that adding in the modal fact doesn't secure the directness of phenomenal acquaintance.

It would seem, then, that though Balog's proposal may be able to explain why phenomenal acquaintance is infallible and thick, it fails to explain its directness. As such, it fails to explain why there is an explanatory gap. Extrapolating from this case study, the moral I'm inclined to draw is that any account of phenomenal concepts whose central claim is that there is a metaphysically necessary connection between phenomenal concepts and their referents (which, of course, entails the sort of physical co-presence discussed above) doesn't have the resources to explain phenomenal acquaintance. Hence, they also fail to explain why there is an explanatory gap.

Now, in the second chapter I claimed that to give an account of phenomenal acquaintance is to explain why it is direct, infallible, and thick, and in this chapter I argued that to explain why there is an explanatory gap is to explain why phenomenal acquaintance has these features. Hence, if my account of phenomenal acquaintance is successful and compatible with physicalism, I have successfully implemented the phenomenal concept strategy.

Is my account of phenomenal acquaintance compatible with physicalism? The only reason I can see that you might not think so involves its appeal to self-representation. If the self-representational status of instances of phenomenal properties can't be accommodated within the physicalist's framework, we certainly have a problem. But in the previous chapter I proposed that we can understand original self-representation in terms of conceptual role,

and CRS, *qua* naturalization of mental representation, is compatible with physicalism.[17] I conclude, then, that my account is compatible with physicalism, or at least I see no reason to think that it isn't.

Let me briefly rehearse, then, my account of phenomenal acquaintance. When you attend to an instance of a phenomenal property, *qua* self-representational semantic entity, you thereby grasp the phenomenal property in question and are acquainted with it. Since the mental representation that underwrites your state of attention is identical to what it represents, the directness puzzle is resolved. Since the mental representation in question is a self-representation, it can't misrepresent, so the infallibility puzzle is resolved. Since the object of your attention is an instance of a phenomenal property – a constituent of your experience – it's no mystery how you come to grasp a particular phenomenal character, so the thickness puzzle is also resolved. If all of this is correct, we have an explanation of how phenomenal acquaintance could be direct, infallible, and thick, and we thereby have an explanation of the explanatory gap, one compatible with physicalism. I conclude, therefore, that we have successfully implemented the phenomenal concept strategy. It would seem that the explanatory gap, therefore, doesn't undermine physicalism. If Levine (2001) is correct that the conceivability and knowledge arguments are merely different manifestations of the explanatory gap, I have also shown that they too fail to undermine physicalism.

5.4. An Objection

Chalmers (2007) proposes an argument against all versions of the phenomenal concept strategy. The conclusion of his argument is that any account of phenomenal

[17] I haven't explicitly addressed the matter of how we're to understand the conceptual role of those mental representations tokens of which are instances of phenomenal properties in terms of their functional role, but I don't see any principled difficulties here.

concepts either isn't tame enough to be explained in physical terms or isn't powerful enough to explain our epistemic situation with regard to experience.[18] The argument applied to my account is as follows. Let '*P*' stand for the conjunction of all and only the microphysical truths in the actual world, and '*Q*' stand for the conjunction of all and only the truths about phenomenal acquaintance according to my account. *Q* includes, for example, the fact that instances of phenomenal properties are self-representational, that they represent themselves in virtue of their conceptual role, that to be acquainted with a phenomenal character is to stand in a relation to the content of a self-representation, and so on. *Q*, in short, includes all aspects of my account of phenomenal acquaintance. Chalmers' argument takes the form of a dilemma, and the argument applied to my account begins with this premise:

(1) Either we can conceive of worlds in which *P* is the case (and there are no additional microphysical truths) but *Q* isn't, or we can't conceive of such worlds.

This apparently is logical truth, so there is no problem here. Let's begin with the first horn of the dilemma:

(2) Suppose that we can conceive of worlds in which *P* is the case but *Q* isn't.

Chalmers claims that there is a tight connection between conceivability and explanation. His idea, very roughly, is that $A\&\sim B$ is conceivable just in case it's not the case that the *A*-truths explain the *B*-truths. Equivalently, the *A*-truths explain the *B*-truths just in case it's not the case that $A\&\sim B$ is conceivable. Let's call this the 'conceivability/explanation principle' for short. Hence, Chalmers endorses the following:

[18] As Chalmers notes, this is similar to Levine's (2007) claim that no account of phenomenal concepts both captures the substantive and determinate nature of our cognitive relation to our experiences and meets what he calls the 'Materialist Constraint', the admonition against taking relations like phenomenal acquaintance as primitive, *a la* Russell, and the insistence that such relations are otherwise compatible with physicalism.

(3) If we can conceive of worlds in which *P* is the case but *Q* isn't, the microphysical truths don't explain the truths about phenomenal acquaintance on my account.

From (2) and (3), it follows that:

(4) My account of phenomenal acquaintance isn't tame enough to be explained in microphysical terms.

Assuming that something is explainable in physical terms only if it's ultimately explainable in microphysical terms, it follows that:

(5) My account of phenomenal acquaintance isn't tame enough to be explained in physical terms (period).

In this case, though my account may explain why there is an explanatory gap, it isn't physically explicable, so we have no business appealing to it in deploying the phenomenal concept strategy.

Now let's turn to the second horn:

(6) Suppose that it's not the case that we can conceive of worlds in which *P* is the case but *Q* isn't.

Let's assume with Chalmers that we can conceive of zombie worlds, worlds in which both *P* is the case and there are no experiences. Let '*E*' stand for the truths about our epistemic situation with respect to experience in the actual world. Two individuals share an epistemic situation with respect to experience just in case they have corresponding beliefs about the experiential, sharing truth-value and epistemic status. Assuming that we can conceive of zombie worlds – worlds in which our physical/functional duplicates fail to have experiences – we get the following:

(7) We can conceive of worlds in which *P* is the case but *E* isn't.

Zombies fail to have any experiences whatsoever, so obviously they fail to share our epistemic situation with respect to experience. Since we can conceive of *P* in the absence of *E* but we can't conceive of *P* in the absence of *Q* (premise 6), it follows that:

(8) We can conceive of a world in which *P* and *Q* are the case but *E* isn't.

From (8) it trivially follows that:

(9) We can conceive of a world in which *Q* is the case but *E* isn't.

By the conceivability/explanation principle, it follows that:

(10) *Q* doesn't explain *E*.

Given (10):

(11) My account of phenomenal concepts isn't powerful enough to explain our epistemic situation with regard to experience.

By disjunctive syllogism, we reach the conclusion:

(12) My account of phenomenal acquaintance either isn't tame enough to be explained in physical terms or powerful enough to explain our epistemic situation with regard to experience.

What should we make of this argument? Well, as I noted above, many claim that the lack of an *a priori* connection doesn't preclude the sort of explanation we're interested in (explanation involving asymmetric metaphysical necessitation), so if we understand the conceivability of *A*&~*B* simply as the *A*-truths failing to *a priori* entail the *B*-truths, then we have no reason to think that the *A*-truths fail to explain the *B*-truths in the relevant sense. Hence, if we reject the conceivability/explanation principle, Chalmers' argument as outlined above fails.

Chalmers is sensitive to this concern, however, and claims that we can recast the argument with a more robust notion of conceivability than that described above. In fact, he reports that he is happy to understand the relevant sense of 'conceivability' as what Levine (2001) call's 'thick' conceivability: *A*&~*B* is thickly conceivable just in case there is an explanatory gap between the *A*-truths and the *B*-truths. The revised conceivability/explanation principle, then, is something like this: if *A*&~*B* is thickly conceivable, the *A*-truths don't explain the *B*-truths. In this case, there are two senses of 'the

A-truths don't explain the *B*-truths' that we need to distinguish, however, for on one sense the revised principle is clearly plausible while on the other it's questionable. As we have seen, to say that *A*&~*B* is thickly conceivable is to say that there is an explanatory gap between the *A*-truths and the *B*-truths. This is to say that it strikes us that, no matter what we learn with respect to the *A*-truths, we will never be able to explain the *B*-truths in terms of the *A*-truths. In other words, it strikes us that the *A*-truths "in principle" fail to explain the *B*-truths. Notice that this doesn't mean, however, that the *A*-truths in fact fail to explain the *B*-truths. It could be that somehow they do, it's just that for some reason or other it strikes us that this couldn't be the case. Let's distinguish, then, between two versions of the revised conceivability/explanation principle:

> If *A*&~*B* is thickly conceivable, the *B*-truths aren't explainable in terms of the *A*-truths.
>
> If *A*&~*B* is thickly conceivable, it strikes us that the *A*-truths, no matter what they turn out to be, fail to explain the *B*-truths.

Given our discussion above, the second version is clearly warranted (it's a definitional matter), while it's unclear that we should endorse the first version. Though I argued earlier in this chapter that the fact that there is an explanatory gap between the physical and the phenomenal is a serious challenge to physicalism (despite what some think), I didn't say that it outright falsifies the position, that it shows that the phenomenal isn't explainable in terms of the physical. If it did, there would be no further work for us to do; we could quit and go home.

Having made this clarification, we can turn to the argument. Chalmers leaves it to us to recast his argument in terms of the notion of an explanatory gap, and this, as best I can tell, is how the argument is supposed to go. Our first premise:

(13) Either there is an explanatory gap between the *P*-truths and the *Q*-truths, or there isn't.

This, like the premise of the original argument, is apparently a logical truth, so again there is no problem here. Let's begin with the first horn:

(14) Suppose that there is an explanatory gap between the *P*-truths and the *Q*-truths.

This is essentially what Levine (2007) suspects is the case when we deploy the phenomenal concept strategy with an account of phenomenal concepts that is compatible with physicalism. He argues that if we restrict ourselves to the resources available to the physicalist and don't accept acquaintance as a brute cognitive relation, any account of phenomenal concepts will fail to explain the thickness of our grasp of phenomenal character. Given (14), it strikes us that the physical/functional truths in principle fail to explain the truths about phenomenal acquaintance on my account. Therefore, (14) suggests (but doesn't entail) that:

(15) My account of phenomenal acquaintance isn't tame enough to be explained in microphysical terms.

As we did above, we'll assume that something is explainable in physical terms only it is ultimately explainable in microphysical terms. In this case, given (15), it follows that:

(16) My account of phenomenal acquaintance isn't tame enough to be explained in physical terms (period).

In this case, though my account may explain why there is an explanatory gap between the physical and the phenomenal, our suspicion is that it isn't physically explicable. Hence, we have no business appealing to it in deploying the phenomenal concept strategy, for in this case we are at best trading one explanatory gap for another.

Now let's turn to the second horn:

(17) Suppose that it's not the case that there is an explanatory gap between the *P*-truths and the *Q*-truths.

Assuming that there is an explanatory gap between the microphysical truths and the phenomenal truths, we can thickly conceive of worlds in which *P* is the case but there are no experiences. Given (17), we can't thickly conceive of worlds in which *P* is the case but *Q* isn't. It follows that we can thickly conceive of worlds in which *P* and *Q* are the case but *E* isn't. It trivially follows that we can thickly conceive of *Q* in the absence of *E*. Hence:

(18) There is an explanatory gap between the *Q*-truths and the *E*-truths.

(18) suggests (but doesn't entail) the following:

(19) My account of phenomenal concepts isn't powerful enough to explain our epistemic situation with regard to experience.

It would seem, then, that we have reached the same conclusion:

(20) My account of phenomenal acquaintance either isn't tame enough to be explained in physical terms or powerful enough to explain our epistemic situation with regard to experience.

I have taken the care to present the new argument in detail because I think it fails. The problem, I claim, is the rationale for (18) – the claim that there is an explanatory gap between the *Q*-truths and the *E*-truths – from above. Let's consider more carefully our sub-argument for (18). It included the following inference:

(1*) We can thickly conceive of a world in which *P* is the case but there are no experiences.

(2*) Hence, we can thickly conceive of a world in which *P* is the case but *E* isn't.

This inference *seems* okay. Zombie worlds are worlds in which *P* is the case but there are no experiences. Worlds in which there are no experiences are obviously worlds in which *E* isn't the case. Hence, if we can conceive of zombie worlds, we can conceive of worlds in which *P* is the case but *E* isn't.

But now let's consider the inference put explicitly in terms of the notion of an explanatory gap:

(1**) There is an explanatory gap between the *P*-truths and the phenomenal truths.

(2**) Hence, there is an explanatory gap between the *P*-truths and the *E*-truths.

Why think that (2**) follows from (1**)? The inference is permissible only if the following scenario is incoherent:

> It strikes us that *A* in principle fails to explain *B*, but it's not the case that it strikes us that *A* in principle fails to explain our epistemic situation with respect to *B*.

It's obvious that there is nothing incoherent about this scenario. So perhaps Chalmers would merely claim that if we have reason to believe that *A* in principle fails to explain *B*, we thereby have reason to believe that we aren't going to be able to explain our epistemic situation with respect to *B* in terms of *A*. But why should this be the case? Well, let's return to our epistemic situation with respect to phenomenal character. However we attempt to explain it, presumably we will have to say something about the *nature* of phenomenal character. Hence, if we suspect that phenomenal character isn't itself to be explained ultimately in terms of the microphysical, we shouldn't expect that our epistemic situation with respect to it is explainable ultimately in terms of the microphysical either.

You will recall, however, that all I say about the nature of phenomenal character in my account of phenomenal acquaintance is that instances of phenomenal properties are self-representational, and I argued that their self-representational status is determined by the conceptual role of the representational vehicles they're tokens of. It would seem that the conceptual role of these mental representations are in principle explainable in terms of the microphysical, so we have no problem here.

The transition between (1**) and (2**), then, isn't warranted without further argument. You can't conclude from that fact that there is an explanatory gap between the microphysical truths and the phenomenal truths that there is an explanatory gap between the

microphysical truths and the truths about our epistemic situation with respect to experience. Here's my diagnosis of what has gone wrong here. When we put the inference I discussed above in terms of conceivability (as, I take it, Chalmers would have us do), it's hard to notice the mistake. With the thin notion of conceivability, it's permissible to move from the claim that we can conceive of a world in which *P* is the case but there are no experiences to the claim that we can conceive of a world in which *P* is the case but *E* isn't without further argument. But with the thick notion of conceivability it turns out that we can't make the corresponding transition without further argument. From the claim that we can thickly conceive of a world in which *P* is the case but there are no experiences, it doesn't follow that we can thickly conceive of a world in which *P* is the case but *E* isn't. The term 'thick conceivability' is misleading to the extent that when we hear 'conceivability' we're inclined to think that such a transition is permissible.

I conclude, therefore, that Chalmers' argument fails. I embrace the second horn of the dilemma in the revised version of the argument, but deny that to do so is to show that my account of phenomenal acquaintance fails to explain why there is an explanatory gap.

5.5. Conclusion

In this chapter I argued that my account of phenomenal acquaintance is not only compatible with physicalism, but it explains why there is an explanatory gap between the physical truths and the phenomenal truths. I also responded to Chalmers' (2007) master argument against the phenomenal concept strategy. Though at the beginning of the dissertation I set aside the phenomenal concept strategy and decided to address phenomenal acquaintance on its own terms, we've seen in this chapter that my account of phenomenal acquaintance can be used to successfully implement the strategy.

BIBLIOGRAPHY

Aydede, M. and G. Güzeldere (2005). Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness, *Nous*, *39*, 2, 197– 255.

Balog, K. (unpublished ms). Acquaintance and the Mind Body Problem.

Block, N. (1986). Advisement for a Semantics in Psychology. In *Midwest Studies in Philosophy*, *10*, 257–274. (Reprinted in S. Stich and T. Warfield (Eds.) (1994). *Mental Representation: A Reader* (pp. 81–141). Oxford: Blackwell.).

Block, N. (1995). On a Confusion about a Function of Consciousness, *Behavioral and Brian Sciences*, 18, 227–247. (Reprinted in N. Block, O. Flanagan, and G. Guzeldere (Eds.) (1997). *The Nature of Consciousness* (pp. 375–416). Cambridge, MA: MIT Press.).

Block, N. (1996). Mental Paint and Mental Latex. In *Philosophical Issues*, 7, *Perception*, 19–49.

Block, N. (2003). Mental Paint. In M. Hahn and B. Ramberg (Eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (pp. 165–200). Cambridge, MA: MIT Press.

Block, N. and R. Stalnaker (1999). Conceptual Analysis, Dualism, and the Explanatory Gap, *The Philosophical Review*, *108*, 1, 1–46.

BonJour, L. and E. Sosa. (2003). *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Oxford: Blackwell Publishers.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Chalmers, D. (2003). The Content and Epistemology of Phenomenal Belief. In Q. Smith and A. Jokic (Eds.), *Consciousness: New Philosophical Essays* (pp. 220–272). Oxford: Clarendon.

Chalmers, D. (2007). Phenomenal Concepts and the Explanatory Gap. In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge* (pp. 167–194). Oxford: Oxford University Press.

Chalmers, D. (forthcoming). The Two-Dimensional Argument Against Materialism. In B. McLaughlin (Ed.), *The Oxford Handbook of the Philosophy of Mind*. Oxford: Oxford University Press.

Chalmers, D. and F. Jackson. (2001). Conceptual Analysis and Reductive Explanation, *Philosophical Review*, *110*, 3, 315–360.

Clark, A. (2004). Feature-Placing and Proto-Objects, *Philosophical Psychology*, *17*, 4, 443–471.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown, and Co.

Dennett, D. (1984). *Elbow Room*. Cambridge, MA: MIT Press.

Dennett, D. (2001). Surprise, Surprise, *Behavioral and Brain Sciences*, *24*, 982.

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

Field, H. (1978). Mental Representation, *Erkenntnis*, *13*, 1, 9–61. (Reprinted in S. Stich and T. Warfield (Eds.) (1994). *Mental Representation: A Reader*. (pp. 34–77). Oxford: Basil Blackwell.).

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. (1994). *The Elm and The Expert*. Cambridge, MA: MIT Press.

Fodor, J. (1990). A Theory of Content I: The Problem. In Fodor, J. (1990). *A Theory of Content and Other Essays* (pp. 51–88). Cambridge, MA: MIT Press.

Fumerton, R. (1995). *Metaepistemology and Skepticism*. London: Roman and Littlefield Publishers.

Grice, P. (1957). Meaning, *Philosophical Review*, 66, 377–388.

Hellie, B. (2007). Factive Phenomenal Characters. In *Philosophical Perspectives*, *21*, *Philosophy of Mind* (pp. 259–306). Oxford: Blackwell Publishing.

Horgan. T. and U. Kriegel. (2007). Phenomenal Epistemology: What is Consciousness that We May Know It So Well? In *Philosophical Issues*, *17*, *The Metaphysics of Epistemology*, 123–144.

Jackson, F. (1982). Epiphenomenal Qualia, *Philosophical Quarterly*, *32*, 127–136.

Jackson, F. (1986). What Mary Didn't Know, *Journal of Philosophy*, *83*, 291–295.

Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In J. Almog, J. Perry, and H. Wettstein (Eds.), *Themes From Kaplan* (pp. 481–564). Oxford: Oxford University Press.

Kim, J. (1976). Events as Property Exemplifications. In M. Brand and D. Walton (Eds.) *Action Theory* (pp. 159–77). Dortrecht, Holland, D. Reidel Publishing Co. (Reprinted in E. Sosa (Ed.) (1993). *Supervenience and Mind: Selected Philosophical Essays* (pp. 33–52). Cambridge, MA: Cambridge University Press.).

Kind, A. (2003). What's So Transparent About Transparency? *Philosophical Studies*, *115*, 225–244.

Kriegel, U. (2003). Consciousness, Higher-Order Content, and the Individuation of Vehicles, *Synthese*, *134*, 477–507.

Kriegel, U. (2005). Naturalizing Subjective Character, *Philosophy and Phenomenological Research*, *71*, 1, 23–57.

Kriegel, U. (2006). The Same-Order Monitoring Theory of Consciousness. In U. Kriegel and K. Williford (Eds.), *Self-Representational Theories of Consciousness* (pp. 143–170). Cambridge, MA: MIT Press.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Langsam, H. (1997). The Theory of Appearing Defended, *Philosophical Studies*, *87*, 33–59.

Levin, J. (2007). What Is a Phenomenal Concept? In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge* (pp. 87–111). Oxford: Oxford University Press.

Levine, J. (1983). Materialism and Qualia: The Explanatory Gap, *Pacific Philosophical Quarterly*, *64*, 354–361.

Levine, J. (1988). Demonstrating in Mentalese, *Pacific Philosophical Quarterly*, *69*, 33, 222–240.

Levine, J. (2001). *Purple Haze*. Oxford: Oxford University Press.

Levine, J. (2003). Experience and Representation. In Q. Smith and A. Jokic (Eds.), *Consciousness: New Essays* (pp. 57–76). Oxford: Oxford University Press.

Levine, J. (2006a). Conscious Awareness and (Self-)Representation. In U. Kriegel and K. Williford (Eds.), *Self-Representational Theories of Consciousness* (pp. 173–198). Cambridge, MA: MIT Press.

Levine, J. (2006b). Color and Color Experience: Colors as Ways of Appearing, *Dialectica*, *60*, 3, 269–282.

Levine, J. (2007). Phenomenal Concepts and Materialist Constraint. In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge* (pp. 145–166). Oxford: Oxford University Press.

Levine, J. (2008). Secondary Qualities: Where Consciousness and Intentionality Meet, *The Monist*, *91*, 2, 215–236.

Levine, J. (unpublished ms). Demonstrative Thought.

Levine, J. and K. Trogdon. (forthcoming). The Modal Status of Materialism, *Philosophical Studies*.

Lewis, D. (1986). Events. In D. Lewis, *Philosophical Papers*, Vol. II. Oxford: Oxford University Press.

Loar, B. (1981). *Mind and Meaning*. Cambridge: Cambridge University Press.

Loar, B. (1997). Phenomenal States. (Revised Version). In N. Block, O. Flanagan, and G. Guzeldere (Eds.), *The Nature of Consciousness* (pp. 597–616). Cambridge, MA: MIT Press.

Loar, B. (2003). Transparent Experience and the Availability of Qualia. In Q. Smith and A. Jokic (Eds.), *Consciousness: New Essays* (pp. 77–96). Oxford: Oxford University Press.

Lycan, W. (1984). *Logical Form in Natural Language*. Cambridge, MA: MIT Press.

Lycan, W. (1995). Why We've Only Just Begun, *Behavioral and Brain Sciences*, *18*, 262–263.

Lycan, W. (2004). The Superiority of HOP to HOT. In R. J. Gennaro (Ed.), *Higher-Order Theories of Consciousness: An Anthology* (pp. 93–114). John Benjamins Publishing Co. Amsterdam/Philadelphia.

Martin, M. (2002). The Transparency of Experience, *Mind and Language*, *4*, 376–425.

McDowell, J. (1984). De Re Senses, *Philosophical Quarterly*, *34*, 283–94. (Reprinted J. McDowell, (1998). *Meaning, Knowledge, and Reality* (pp. 214–227). Cambridge, MA: Harvard University Press.).

McDowell, J. (1986). Singular Thought and the Extent of Inner Space. In P. Pettit and J. McDowell (Eds.), *Subject, Thought, and Context* (pp. 137-168). Oxford: Clarendon Press. (Reprinted in J. McDowell, (1998). *Meaning, Knowledge, and Reality* (pp. 228–259). Cambridge, MA: Harvard University Press.).

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Noë, A. (2009). *Out of Our Heads*. New York: Hill and Wang.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. (2007). Phenomenal Concepts and Perceptual Concepts. In T. Alter and S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge* (pp. 111–144). Oxford: Oxford University Press.

Pears, D. (1969). *Bertrand Russell and the British Tradition in Philosophy*. New York: Random House.

Pears, D. (1981). The Function of Acquaintance in Russell's Philosophy, *Synthese*, *46*, 2, 149–66. (Reprinted in A. D. Irvine (Ed.) (1999). *Bertrand Russell: Critical Assessments*, *3*, *Language, Knowledge, and World* (pp. 221–235). New York: Routledge.).

Perry, J. (2001). *Knowledge, Possibility and Consciousness*. Cambridge, MA.: MIT Press.

Rey, G. (1998). A Narrow Representational Account of Qualitative Experience. In J. Tomberlin (Ed.), *Philosophical Perspectives*, *12*, *Language, Mind, and Ontology* (pp. 435–457). Atascadero: Ridgeview Press.

Rosenthal, D. (2004). Varieties of Higher-Order Theory. In R. J. Gennaro (Ed.), *Higher-Order Theories of Consciousness: An Anthology* (pp. 17–44). John Benjamins Publishing Co. Amsterdam/Philadelphia.

Russell, B. (1911). "Knowledge by Acquaintance and Knowledge by Description." Reprinted in J. C. Slater (Ed.) (1992). *The Collected Papers of Bertrand Russell*, Vol. 6 (pp. 148–161). New York: Routledge.

Russell, B. (1912). *The Problems of Philosophy*. Oxford: Oxford University Press.

Russell, B. (1913a). *Theory of Knowledge: The 1913 Manuscript*. In E. R. Eames (Ed.) (1984). *The Collected Papers of Bertrand Russell*, Vol. 7 (pp. 5–178). New York: Routledge.

Russell, B. (1913b). The Nature of Sense-Data: A Reply to Dr. Dawes Hicks. Reprinted in J. C. Slater (Ed.) (1992). *The Collected Papers of Bertrand Russell*, Vol. 6 (pp. 184–189). New York: Routledge.

Russell, B. (1914a). *Our Knowledge of the External World*. London: The Open Court Publishing Company.

Russell, B. (1914b). The Relation of Sense-Data to Physics. Reprinted in J. C. Slater (Ed.) (1996). *The Collected Papers of Bertrand Russell*, Vol. 8 (pp. 3–26). New York: Routledge.

Russell, B. (1915). Letter on Sense-Data. Reprinted in J. C. Slater (Ed.) (1986). *The Collected Papers of Bertrand Russell*, Vol. 8 (pp. 87). New York: Routledge.

Russell, B. (1918). On Sensations and Ideas. Reprinted in J. C. Slater (Ed.) (1986). *The Collected Papers of Bertrand Russell*, Vol. 8 (pp. 252–255). New York: Routledge.

Russell, B. (1921). *The Analysis of Mind*. London: George Allen & Unwin LTD.

Russell, B. (1940). *An Inquiry Into Meaning and Truth*. London: George Allen and Unwin.

Russell, B. and A. N. Whitehead (1910). *Principia Mathematica*, Vol. 1. Cambridge: Cambridge University Press.

Sainsbury, R.M. (1986). Russell on Acquaintance. In G. Vesey (Ed.), *Philosophers Ancient and Modern*, Royal Institute of Philosophy Lecture Series, 20, Supplement to *Philosophy* (pp. 219–44). Cambridge: Cambridge University Press.

Savage, C. W. (1989). Sense-data in Russell's Theories of Knowledge. In C. W. Wade and C. A. Anderson (Eds.), *Rereading Russell: Essays on Bertrand Russell's Metaphysics and Epistemology*. Minnesota Studies in the Philosophy of Science, Vol. 12 (pp. 138-68). Minneapolis: University of Minnesota Press.

Schaffer, J. (forthcoming). Monism: The Priority of the Whole, *Philosophical Review*.

Schroer, R. (2007). The Reticence of Visual Phenomenal Character: A Spatial Interpretation of Transparency, *Australasian Journal of Philosophy*, *85*, 3, 393–414.

Schroer, R. (unpublished ms). Where's the Beef? An Account Of How The Contents of Phenomenal Concepts Can Be Substantial While Also Containing Demonstrative Elements.

Schwitzgebel, E. (2008). The Unreliability of Naïve Introspection, *Philosophical Review*, *117*, 245–273.

Searle, J. (1980). Minds, Brains and Programs, *Behavioral and Brain Sciences*, *3*, 3, 417–457

Sellars, W. (1997). *Empiricism and Philosophy of Mind*. Cambridge, MA: Harvard University Press.

Shoemaker, S. (2007). *Physical Realization*. Oxford: Oxford University Press.

Siewart, C. (2004). Is Experience Transparent? *Philosophical Studies*, 117, 15–41.

Siewart, C. (2007). Who's Afraid of Phenomenological Disputes? *The Southern Journal of Philosophy*, *45*, 1–21.

Smith, D. W. (1989). *The Circle of Acquaintance: Perception, Consciousness, and Empathy*. Synthese Library, 205. Dordrecht, Holland: Kluwer.

Sosa, E. (2003). Privileged Access. In Q. Smith and A. Jokic (Eds.), *Consciousness: New Essays* (pp. 273–294). Oxford: Oxford University Press.

Spener, M. (unpublished ms). Phenomenal Adequacy and Introspective Evidence.

Stoljar, D. (2005). Physicalism and Phenomenal Concepts, *Mind and Language*, *20*, 5, 469–94.

Stoljar, D. (2006). *Ignorance and Imagination*. Oxford: Oxford University Press.

Trogdon, K. (2009a). Physicalism and Sparse Ontology, *Philosophical Studies*, *143*, 2, 147–165.

Trogdon, K. (2009b). Monism and Intrinsicality, *Australasian Journal of Philosophy*, *87*, 1, 127–148.

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge: Cambridge University Press.

Tye, M. (1999). Phenomenal Consciousness: the Explanatory Gap as Cognitive Illusion, *Mind*, *108*, 705–725.

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

Tye, M. (2003). A Theory of Phenomenal Concepts. In A. O'Hear (Ed.), *Minds and Persons* (pp. 91–105). Cambridge: Cambridge University Press.

Van Gelder, T. (1995). What Might Cognition Be, if not Computation? *Journal of Philosophy*, *91*, 345–381.

White, A. R. (1981). Knowledge, Acquaintance, and Awareness. In P. A. French, T. E. Uehling, Jr., and H. K. Wettstein (Ed.), *Midwest Studies in Philosophy*, Vol. 6: *The Foundations of Analytic Philosophy* (pp. 159–72). Minneapolis: University of Minnesota Press.

Witmer, G., B. Butchard, and K. Trogdon. (2005). Intrinsicality without Naturalness, *Philosophy and Phenomenological Research*, 70, 326–350.